Childbearing and labour force participation in South Africa: sibling composition as an identification strategy?

Gabrielle van der Stoep

University of KwaZulu-Natal

1. Background

In general, empirical studies find a negative relationship between childbearing and labour force participation (Boushey 2008; Connelly *et al* 2006; Chapman *et al* 2001; Angrist and Evans 1998; Gormick *et al* 1996; Bronars and Grogger 1994; Browning 1992)

The key problem in identifying the strength of a causal relationship is the **endogeneity of childbearing** status:

i) Decision to have children may be a function of labour force participation

ii) Unobservable characteristics correlated with both childbearing decisions and labour force participation.

Instrumental variable (IV) estimation used to address endogeneity.

Two properties of an IV:

i) Must be uncorrelated with omitted or unobserved variables in the error term.

ii) Must be strongly correlated with the endogenous explanatory variable (Wooldridge 2003; Bound *et al* 1995)

If two properties not met then:

- Large standard errors on IV estimates
- Asymptotic bias
- Inconsistent IV estimates

Suggested instruments:

Twin births:	Bronars & Grogger (1994); Rosenzwieg & Wolpin (1980)
Religious affiliation:	Ryder & Westoff (1972)
Infertility status:	Aguero & Marks (2008)
Chinese Lunar Calendar:	Vere (2008)

Sibling sex composition:

Angrist & Evans (1998) - USA
Iacovou (2001) – UK
Cruces & Galiani (2007) - Argentina, Mexico

Sibling Composition:

Assumptions and findings:

i) Parents have a **preference for a mixed sibling sex composition** or a "taste for balance" (Ben-Porath & Welch 1976).

Compared to parents whose first two children are of mixed sex, parents whose first two children are of the same sex are **more likely** to have a third birth.

ii) The sex of any child born is **randomly** assigned.

2. Objective

To investigate whether sibling composition is a **suitable identification strategy** to estimate the effects of childbearing on labour force participation among Africans in South Africa.

A strong correlation between sibling composition and further childbearing **may not hold**

➔ Arnold (1992)
➔ Ben-Porath and Welch (1976)
➔Gangadharan and Maitra (2003)

Possible reason? **Larger desired family sizes** among Africans in South Africa compared to countries in which similar studies have been conducted.

3. Data and definitions

General Household Survey (GHS) 2002

Same sex sibling composition can only be used to instrument for **further** childbearing among women with **at least two children.**

Sample = 7,477 African women aged 20 to 49 having had at least **two live births**, at least one child still alive and 18 years or younger

3. Data and definitions

Distinguish sample by

1) Co-residency status:

   Co-resident mother = mother co-resident with at least one biological child

   Not co-resident mother = mother not co-resident with at least one child (see Posel and van der Stoep 2008)

2) Marital status:

   Married = "married or living together as husband and wife"

4. Methodology

Structural model:

$Y_i = \alpha + \beta T_i + \delta X_i + e_i$ (1)

Two stage least squares estimation:

First-stage equation using <u>one</u> instrument:

$T_i = \gamma(S_i) + \lambda X_i + \eta_i$ (2a)

S_i = 1 if first two children born are of same sex and 0 if mixed sex.

First-stage equation using <u>two</u> instruments:

$T_i = \gamma_o(G_i) + \gamma_1(B_i) + \lambda X_i + \eta_i$ (2b)

G_i, (B_i) = 1 if first two children born are both girls (boys), and 0 otherwise.

5.1 Property one

In developing countries with strong son preferences and sex-selection, property one may be violated.

- But no evidence of sex selection in among Africans in South Africa:
  - No strong son preferences (Ganghadaran & Maitra 2003)
  - Infant sex ratios not skewed in favour of boys (Garenne 2004)
- Also no systematic differences in demographic characteristics across women with mixed vs. same sex sibling composition.

Conclusion: sibling composition does not violate property one

Table 1: Differences in mean characteristics of African women by the sex composition of their first two births, 2002

Mean characteristic difference= mean (same sex) – mean (mixed sex)	Co-resident mothers having had at least two births		Not co-resident mothers having had at least two births	
Age	-0.251 (0.171)	t = -1.4664	0.693 (0.457)	t = 1.516
Age at first birth	-0.066 (0.098)	t = -0.6736	0.385 (0.251)	t = 1.537
Married	0.003 (0.012)	t = 0.2182	0.049 (0.032)	t = 1.557
Rural	-0.004 (0.012)	t = -0.2979	-0.015 (0.032)	t = -0.471
Total monthly household earnings[a] (2002 prices)	5.538 (76.196)	t = 0.0727	-36.383 (146.400)	t = -0.2485
Years of schooling	0.195* (0.100)	t = 1.9373	-0.120 (0.255)	t = -0.469
No schooling	-0.017** (0.008)	t = -2.2030	0.001 (0.020)	t = 0.067
Incomplete Primary	-0.013 (0.010)	t = -1.308	-0.026 (0.027)	t = -0.982
Completed primary	0.014** (0.007)	t = 2.006	-0.010 (0.020)	t = -0.523
Incomplete Secondary	0.005 (0.012)	t = 0.405	0.035 (0.031)	t = 1.149
Matric	0.005 (0.009)	t = 0.588	-0.007 (0.022)	t = -0.328
Post-matric	0.006 (0.006)	t = 0.883	0.007 (0.014)	t = 0.496
Labour force participant	0.015 (0.012)	t = 1.254	-0.030 (0.026)	t = -1.174

Source: GHS 2002. **Notes**: Standard deviations are in parentheses. Means differences are calculated by subtracting the mean characteristic for women with mixed sex siblings from those with same sex siblings. T-statistic is abbreviated by 't'. The null hypothesis of the t-statistic is that the mean difference is zero i.e. no difference exists in the average characteristics of women with a same and mixed sibling sex composition. *** 1 % significance level; ** 5% significance level; * 10% significance level.

5.2 Property two

Table 2: First-stage results - dependent variable is *more than two births*, T_i

	All mothers (not distinguished by co-residency with children)		
	Married and Unmarried	Married	Unmarried
Instrument: Same sex siblings			
Coefficient: same sex	0.020** (0.009)	0.011 (0.012)	0.031** (0.015)
Partial R^2	0.0006	0.0002	0.0014
F-statistic	4.541	0.822	4.216
Instrument: Two girls, two boys			
Coefficient: Two Girls	0.023** (0.011)	0.010 (0.014)	0.043** (0.018)
Coefficient: Two Boys	0.016 (0.011)	0.011 (0.014)	0.019 (0.019)
Partial R^2	0.0006	0.0002	0.0018
F-statistic	2.406	0.384	2.751
N	7,477	4,389	3,088

Source: GHS 2002

Notes: Standard deviations are in parentheses. Other exogenous variables include the woman's age, age squared, age at first birth, her educational status and marital status. Controls are also included for household compositional variables, dummy variables for province of residence and whether the woman lives in a rural or urban area. *** 1 % significance level; ** 5% significance level; * 10% significance level.

Table 2 continued…

	Co-resident mothers		
	Married & unmarried	**Married**	**Unmarried**
Instrument: Same sex siblings			
Coefficient: same sex	0.018* (0.010)	0.011 (0.012)	0.026* (0.017)
Partial R^2	0.0005	0.0002	0.0010
F-statistic	3.24202	0.840	2.499
Instrument: Two girls, two boys			
Coefficient: Two Girls	0.026** (0.012)	0.013 (0.015)	0.045** (0.020)
Coefficient: Two Boys	0.009 (0.012)	0.008 (0.015)	0.008 (0.021)
Partial R^2	0.0007	0.0002	0.0020
F-statistic	2.303	0.419	2.508
N	6,487	3,949	2,538
	Not co-resident mothers		
	Married & unmarried	**Married**	**Unmarried**
Instrument: Same sex siblings			
Coefficient: same sex	0.039 (0.028)	0.015 (0.041)	0.056 (0.037)
Partial R^2	0.0021	0.0003	0.0044
F-statistic	2.019	0.128	2.30815
Instrument: Two girls, two boys			
Coefficient: Two Girls	0.017 (0.034)	-0.012 (0.051)	0.048 (0.045)
Coefficient: Two Boys	0.062* (0.034)	0.041 (0.051)	0.065 (0.045)
Partial R^2	0.0035	0.0022	0.0046
F-statistic	1.674	0.452	1.203
N	990	440	550

5.2 Property two

R^2 value very small (< 0.005)

F-statistics well below value of 10 for a strong instrument (Stock *et al* 2002)

Weak correlation between sibling composition and further childbearing; and only observed for select sample of mothers.

Conclusion: Sibling composition fails property two

6. OLS & second-stage results

Table 3: Second-stage and OLS results - dependent variable is *labour force participation*, Y_i

	All mothers (not distinguished by co-residency with children)		
	Married and unmarried	**Married**	**Unmarried**
OLS	-0.013 (0.013)	-0.051*** (0.018)	0.037** (0.018)
IV: same sex	-0.230 (0.525)	0.599 (1.481)	-0.546 (0.558)
IV: two girls & two boys	-0.038 (0.501)	0.346 (1.415)	-0.083 (0.424)
N	7,477	4,389	3,088
	Co-resident mothers		
	Married and unmarried	**Married**	**Unmarried**
OLS	-0.007 (0.014)	- 0.048** (0.019)	0.048** (0.021)
IV: same sex	-0.565 (0.700)	0.099 (1.316)	-0.928 (0.895)
IV: two girls & two boys	0.032 (0.527)	0.145 (1.324)	-0.023 (0.459)
N	6,487	3,949	2,538
	Not co-resident mothers		
	Married and unmarried	**Married**	**Unmarried**
OLS	0.016 (0.028)	-0.020 (0.050)	0.030 (0.034)
IV: same sex	0.631 (0.757)	3.601 (10.233)	-0.003 (0.500)
IV: two girls & two boys	0.153 (0.487)	0.306 (1.120)	-0.080 (0.493)
N	990	440	550

Source: GHS 2002 **Notes:** Standard errors are in parentheses. Other exogenous variables include the woman's age, age squared, age at first birth, her educational status and marital status. Controls are also included for household compositional variables, dummy variables for province of residence and whether the woman lives in a rural or urban area. *** 1 % significance level; ** 5% significance level; * 10% significance level.

6. Second-stage results

Imprecise IV estimates result of weak explanatory power of sibling composition instruments in the first stage.

➔ indicated by *very* small partial R^2

But asymptotic variance =

$\sigma^2 / (STT_T * R^2)$

Resulting standard errors on IV estimates 20 to 200 times larger than on OLS estimates.

Concluding remarks

- Sibling composition as an instrument for childbearing among Africans in South Africa, **does not violate property one** (uncorrelated with the error term)
- however, it is **fails property two** (whether a strong correlation exists between the instrument and endogenous explanatory variable)
- No strong preference exists for a mixed sibling sex composition - possible reason is large desired family sizes among the sample of African women.
- Therefore IV has weak explanatory power in first-stage estimations.

Implications:

i) large standard errors on IV estimates

ii) IV estimates are noisy, imprecise, have no meaningful interpretation.

Questions?