

# Preference for Boys, Family Size and Educational Attainment in India

Santosh Kumar<sup>1</sup>   Adriana Kugler<sup>2</sup>

<sup>1</sup>Sam Houston State University

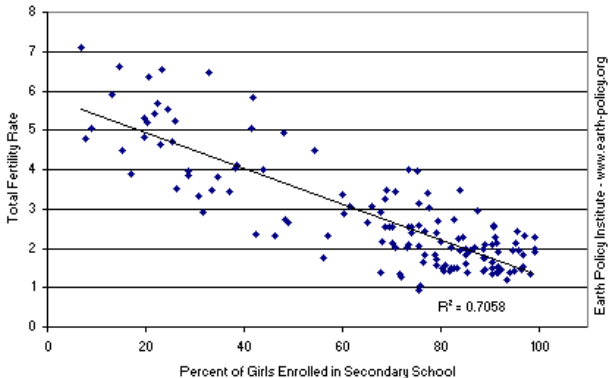
<sup>2</sup>Georgetown University

UNU-WIDER Development Conference, Helsinki, Finland  
6-7 June 2016

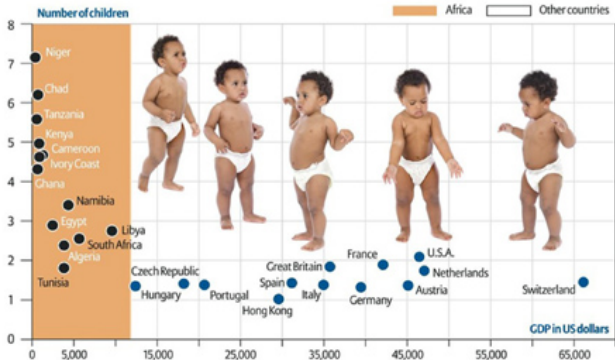
# Motivation

- Poverty is widespread; about 21% of world's population live on less than \$1.25 a day
- Human capital is critical input for economic growth and development
- However, human capital accumulation rate has been slow in several developing countries
- Can high fertility explain the low level of human capital accumulation?

## Female Secondary Education and Total Fertility Rates



Source: EPI from UNESCO



Source: Allianz/UN Population Division, World Population Prospects. 2008 revision, IMF

# Motivation

- At the microeconomic level, family size and human capital move in opposite direction
- A smaller family will have more resources to spend on each child
- Becker's fertility model: trade-off between child quantity and child quality (Q-Q Trade-off)

# Research question

- What is the causal impact of family size on children's education and health in India?
  - Is there any evidence of Q-Q trade-off in India?
- Does the impact vary household's characteristics?

## Preview of the main results

- We find strong evidence of a quantity-quality trade off in educational outcomes
- Increasing the household size by one child reduces the literacy rate by 3.4% and the years of schooling by 2.6%
- Trade-off is more pronounced rural areas, in low caste, and low-wealth households with an extra children reducing the years of schooling by as high as 10.6% (0.3) years
- In contrast, no significant effect of family size on health outcomes are visible.

# Identification strategy

$$\text{Education/Health}_{ihd} = \beta_0 + \beta_1 * \text{FamilySize}_{hd} + \beta_2 * X_{ihd} + \mu_d + \epsilon_{ihd} \quad (1)$$

- FamilySize is # of children under 21 years of age
- X is a vector that includes child and parents' characteristics (age, gender, caste, birth order, rurality, parents education & age)
- $\mu_d$  is district fixed effect
- $\beta_1 < 0$  implies Q-Q trade-off



# Econometric challenge

- The OLS regression above is unlikely to provide a causal estimate of family size on child quality
- FamilySize is likely endogenous as child quality and quantity are jointly determined
  - OLS estimates may be biased
  - Upward biased if wealthier households have fewer children and invest more in education
  - Downward biased if highly committed parents have more children and invest more in education

# Econometric challenge

- We use Instrumental Variable (IV) method
- Gender of first-born as an instrument
- For the instrument to be valid:
  - It has to be highly correlated with family size
  - It has no relation to quality other than through family size

## Instrument (Gender for first-born)

- Son preference in India
- Payment of dowries and large gender pay gap in India imply that it is more expensive to support a girl
- Boys tend to take care of parents at old age, so in a society with limited safety nets parents prefer sons

## Poor Indian Family with 7 children

First born is a girl



# Threats to the validity of the instrument

- Presence of sex-selective abortion may invalidate the instrument
- However, no evidence of selective abortion:
  - Fetal sex determination became illegal in India in 1996 after passing of PNDT Act
  - Many studies have found no evidence of sex-selective abortion for first-born but for second-born (Bhalotra and Cochrane, 2010; Ebenstein, 2007; Jha et al., 2011; Portner, 2010; Rosenblum, 2010)
  - Regression of IV on exogeneous variables

# Two Stage Least Square (2SLS) model

First stage:

$$FamilySize_{hd} = \beta_0 + \beta_1 * Z_{hd}(FirstGirl) + \beta_2 * X_{ihd} + \mu_d + \epsilon_{ihd} \quad (2)$$

Second stage:

$$Education/Health_{ihd} = \beta_3 + \beta_4 * \widehat{FamilySize}_{hd} + \beta_5 * X_{ihd} + \mu_d + \epsilon_{ihd} \quad (3)$$

# Data

- District Level Household and Village Survey (DLHS-3, 2007-08)
  - Nationally representative household survey; N=600,000 households
  - Interviewed 1000 households in each district
  - National Family Health Survey (NFHS-3)
  
- Child quality measures
  - Probability of being literate
  - Probability of ever attending school
  - Years of schooling and Current enrolment
  
  - Weight, Height, weight-for-age z-score, height-for-age z-score, weight-for-height z-score
  - Underweight, stunting, wasting

# RESULTS

- RESULTS



## Descriptive Statistics of the Education Sample

	All	First-born girl	First-born boy
Child Age (5-20 years old)	9.60 (3.45)	9.41 (3.34)	9.79 (3.55)
<b>Gender of first child (female=1)</b>	<b>0.49</b> <b>(0.49)</b>		
Literate	0.82 (0.38)	0.81 (0.39)	0.83 (0.37)
Ever attended school	0.9 (0.30)	0.89 (0.31)	0.91 (0.29)
Still enrolled	0.95 (0.21)	0.95 (0.21)	0.95 (0.22)
Years of schooling	3.08 (2.92)	2.94 (2.85)	3.22 (2.98)
Mother's age	30.94 (3.36)	30.88 (3.34)	31.00 (3.37)
Father's age	36.48 (4.81)	36.42 (4.79)	36.54 (4.82)
Mother's years of schooling	2.99 (4.06)	3.05 (4.09)	2.93 (4.03)
Father's years of schooling	5.48 (4.74)	5.56 (4.76)	5.40 (4.72)
Family size	3.54 (1.33)	3.70 (1.33)	3.40 (1.31)
Rural	0.82 (0.39)	0.81 (0.39)	0.82 (0.39)
Low caste (SC & ST)	0.41 (0.49)	0.41 (0.49)	0.41 (0.49)
Middle caste (OBC)	0.39 (0.49)	0.39 (0.49)	0.39 (0.49)
Low wealth	0.49 (0.50)	0.48 (0.50)	0.49 (0.50)
Medium wealth	0.39 (0.49)	0.4 (0.49)	0.39 (0.49)
No. of observations	393,597	193,263	200,334
No. of districts	601		

*Notes:* Standard deviations are shown in parentheses. All sampled children were 5-20 years old at the time of survey (2007-08). The analytical sample is restricted to 20-35 years old mother.

## Descriptive Statistics of the Health Sample

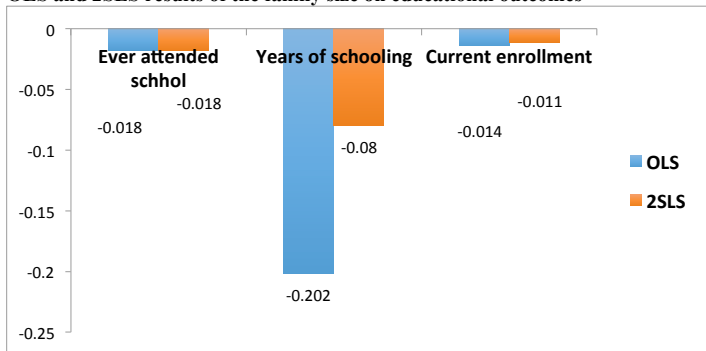
	All	First-born girl	First-born boy
Child Age (months)	28.0 (17.01)	28.31 (17.04)	28.57 (16.99)
<b>Gender of first child (female=1)</b>	<b>0.51</b> <b>(0.49)</b>		
Weight (Gram)	10251.72 (3123.52)	10142.56 (3061.12)	10363.79 (3182.72)
Height (Centimeter)	81.85 (13.34)	81.62 (13.27)	82.10 (13.42)
WAZ	-1.63 (1.16)	-1.64 (1.17)	-1.63 (1.16)
HAZ	-1.47 (1.50)	-1.46 (1.52)	-1.48 (1.48)
WfH	-0.92 (1.13)	-0.92 (1.12)	-0.94 (1.13)
Child is underweight	0.39 (0.49)	0.39 (0.49)	0.39 (0.49)
Child is stunted	0.35 (0.48)	0.35 (0.48)	0.35 (0.48)
Child is wasted	0.15 (0.36)	0.15 (0.36)	0.15 (0.36)
Family size	2.17 (0.42)	2.20 (0.44)	2.16 (0.39)
Rural	0.61 (0.49)	0.62 (0.49)	0.60 (0.49)
Low caste (SC & ST)	0.33 (0.47)	0.33 (0.47)	0.32 (0.47)
Middle caste (OBC)	0.33 (0.47)	0.33 (0.47)	0.33 (0.47)
Low wealth	0.29 (0.45)	0.29 (0.45)	0.29 (0.45)
Medium wealth	0.22 (0.41)	0.22 (0.41)	0.22 (0.41)
Mother's age	24.09 (3.75)	24.14 (3.75)	24.04 (3.74)
Father's age	29.38 (4.79)	29.46 (4.85)	29.30 (4.73)
N	10090	5111	4979
No. of states	29		

## Regression of Gender of First Born on household characteristics

	Dependent variable: First-born is a girl	
	LPM	Probit
	(1)	(2)
Rural	-0.003 (0.004)	-0.008 (0.011)
Low wealth	-0.003 (0.007)	-0.004 (0.017)
Medium wealth	-0.0002 (0.005)	-0.0005 (0.013)
Religion (Hindu=1)	0.006 (0.004)	0.016 (0.011)
Scheduled caste/tribe (Yes=1)	0.004 (0.004)	0.009 (0.011)
Other backward caste	0.002 (0.004)	0.006 (0.010)
Mother's years of schooling	0.002 (0.001)	0.004 (0.003)
Mother's years of schooling (square)	-0.00007 (0.00008)	-0.0002 (0.0002)
Father's years of schooling	-0.00008 (0.0009)	-0.0002 (0.002)
Father's years of schooling (square)	0.0001 (0.00006)	0.0002 (0.0001)
Mother's age	0.036*** (0.006)	0.092*** (0.016)
Father's age	0.002 (0.003)	0.004 (0.008)

Notes: \*, \*\*, and \*\*\* represent significance levels of 10, 5, and 1 percent. Robust standard errors, clustered by district, are shown in parentheses. All models include district fixed-effects. Column 2 reports marginal effects from the probit model.

### OLS and 2SLS results of the family size on educational outcomes



*Notes:* Robust standard errors, clustered by district, are shown in parentheses. Children's controls include age, age squared, gender and birth order. Parents' control includes education levels of father and mother, household religion, household caste, rural, and household socioeconomic status. Family size is total number of 0-20 years old children in the family at the time of the survey.

### IV Estimates of the Effect of Family Size on Children's Educational Outcomes

	Instrument: First child is a girl (G)			
	Literate	Ever attended school	Years of schooling	Currently enrolled
	(1)	(2)	(3)	(4)
First Stage	0.219*** (0.007)	0.219*** (0.007)	0.219*** (0.007)	0.228*** (0.007)
Family size	-0.028*** (0.007)	-0.018*** (0.006)	-0.081** (0.033)	-0.011*** (0.004)
<b>Weak-Identification Tests</b>				
Kleibergen-Paap Wald rk F-stat	814.61	814.61	421.51	932.80
P-value	0.00	0.00	0.00	0.00
<b>Weak-Instrument-Robust – Inference</b>				
Anderson-Rudin F	11.77	11.77	0.02	5.86
P-value	0.00	0.00	0.88	0.016
Stock-Wright S stat	11.36	11.36	0.02	5.79
P-value	0.00	0.00	0.88	0.016
Children's control	yes	yes	yes	yes
Parents' controls	yes	yes	yes	yes
District fixed-effect	yes	yes	yes	yes

*Notes:* \*, \*\*, and \*\*\* represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age<sup>2</sup>, gender, birth order, religion, caste, SES and district fixed effect. Parents' controls include age, age<sup>2</sup>, gender, birth order, religion, caste, SES and district fixed effect.

# Threats to identification

- Son-preferring, differential stopping behaviour (SP-DSB), may alter the sex composition in the family
  - Results are robust to inclusion of number of girls in the model
- Increased probability of mother's employment/saving due to dowry payment
  - First-born girl does not predict mother's employment or asset accumulation
- First-born girl may also increase maternal and adult mortality after age 30
  - Restricting the sample to age 30 does not change the result

# Heterogeneous Results

Q-Q trade-off is higher

- among socially disadvantaged caste households
- in rural areas
- among poor households
- for less-educated mother

OLS and 2SLS Estimates of the Effects of Family Size on Education by Caste and Residence

Dependent variables	Instrument: First child is a girl (FG)									
	Low caste		Middle caste		High caste		Rural		Urban	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Literate	-0.021*** (0.002)	-0.046*** (0.012)	-0.019*** (0.001)	-0.021* (0.011)	-0.018*** (0.002)	-0.004 (0.013)	-0.020*** (0.002)	-0.030*** (0.007)	-0.018*** (0.002)	-0.023 (0.017)
R-square	0.25	0.16	0.25	0.19	0.23	0.17	0.25	0.18	0.25	0.19
N	161380	161380	153015	153015	79202	79202	321140	321140	72457	72457
Ever in school	-0.019*** (0.002)	-0.036*** (0.010)	-0.017*** (0.001)	-0.012 (0.009)	-0.016*** (0.002)	0.007 (0.010)	-0.018*** (0.002)	-0.018** (0.007)	-0.016*** (0.002)	-0.021 (0.014)
R-square	0.16	0.074	0.14	0.088	0.14	0.073	0.14	0.084	0.16	0.097
N	161380	161380	153015	153015	79202	79202	321140	321140	72457	72457
Years of schooling	-0.181*** (0.009)	-0.162** (0.054)	-0.201*** (0.008)	-0.089 (0.050)	-0.210*** (0.013)	0.093 (0.064)	-0.197*** (0.006)	-0.107** (0.035)	-0.197*** (0.013)	-0.046 (0.086)
R-square	0.67	0.637	0.71	0.683	0.78	0.758	0.69	0.663	0.78	0.765
N	161380	161380	153015	153015	79202	79202	321140	321140	72457	72457
Currently enrolled	-0.011*** (0.001)	-0.005 (0.006)	-0.015*** (0.001)	-0.019*** (0.005)	-0.016*** (0.001)	-0.008 (0.007)	-0.014*** (0.0007)	-0.010** (0.004)	-0.016*** (0.001)	-0.026** (0.009)
R-square	0.16	0.140	0.16	0.147	0.15	0.131	0.16	0.004)	0.15	0.131
N	138272	138272	135014	135014	72699	72699	279847	279847	66138	66138
Children's control	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Parents' controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
District F.E.	yes	yes	yes	Yes	yes	yes	yes	yes	yes	yes

Notes: \*, \*\*, and \*\*\* represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age square, gender, birth order, religion, caste, SES and rural dummies. Parent controls include age, age square, and education levels of father and mother. Family size is total number of 0-20 years old children in the family at the time of the survey. Low caste is scheduled caste (SC) and scheduled tribe (ST) households while middle caste is other backward caste (OBC) category. Poor is households in bottom two quintiles based wealth index constructed from assets, amenities and durables.



### OLS and 2SLS Estimates of the Effects of Family Size on Education by Household Wealth and Mother's Education

Dependent variables	Instrument: First child is a girl (G)											
	Household wealth						Mother's education					
	Bottom two quintile		Third quintile		Top quintile		Illiterate		Less than primary		Primary & above	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Literate	-0.024*** (0.002)	-0.052*** (0.012)	-0.016*** (0.001)	-0.017* (0.008)	-0.007*** (0.002)	-0.007 (0.013)	-0.026*** (0.001)	-0.046*** (0.010)	-0.008*** (0.002)	-0.027* (0.012)	-0.005*** (0.001)	-0.0004 (0.011)
R-square	0.23	0.15	0.23	0.20	0.21	0.08	0.23	0.16	0.25	0.20	0.23	0.19
N	191211	191211	154262	154262	48124	48124	227697	227697	63815	63815	102085	102085
Ever in school	-0.023*** (0.002)	-0.402*** (0.011)	-0.013*** (0.001)	-0.009 (0.006)	-0.003*** (0.0009)	0.003 (0.007)	-0.024*** (0.001)	-0.039*** (0.009)	-0.005** (0.001)	-0.005 (0.008)	-0.003*** (0.0008)	0.006 (0.006)
R-square	0.14	0.08	0.10	0.07	0.05	0.03	0.14	0.08	0.09	0.07	0.08	0.05
N	191211	191211	154262	154262	48124	48124	227697	227697	63815	63815	102085	102085
Years of schooling	-0.177*** (0.008)	-0.261*** (0.052)	-0.177*** (0.008)	-0.004 (0.039)	-0.098*** (0.012)	0.113 (0.068)	-0.205*** (0.008)	-0.295*** (0.046)	-0.134*** (0.009)	0.104* (0.057)	-0.102*** (0.007)	0.107** (0.050)
R-square	0.60	0.56	0.77	0.76	0.88	0.87	0.63	0.58	0.81	0.78	0.87	0.86
N	191211	191211	154262	154262	48124	48124	227697	227697	63815	63815	102085	102085
Currently enrolled	-0.013*** (0.0009)	-0.018*** (0.006)	-0.013*** (0.0009)	-0.007 (0.005)	-0.008*** (0.001)	-0.013** (0.006)	-0.015*** (0.0008)	-0.018*** (0.005)	-0.011*** (0.001)	-0.020*** (0.007)	-0.008*** (0.0008)	-0.005 (0.004)
R-square	0.18	0.16	0.16	0.14	0.08	0.06	0.18	0.17	0.15	0.13	0.08	0.06
N	157962	157962	142067	142067	46226	46226	189169	189169	59323	59323	97493	97493
Children's control	yes	Yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Parents' controls	yes	Yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
District F.E.	yes	Yes	yes	yes	Yes	yes	yes	yes	yes	yes	yes	yes

**Notes:** \*, \*\*, and \*\*\* represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age square, gender, birth order, religion, caste, SES and rural dummies. Parent controls include age, age square, and education levels of father and mother. Family size is total number of 0-20 years old children in the family at the time of the survey. Low caste is scheduled caste (SC) and scheduled tribe (ST) households while middle caste is other backward caste (OBC) category. Poor is households in bottom two quintiles based wealth index constructed from assets, amenities and durables.

**OLS and 2SLS Estimates of the impact of Family Size on Child's Health Outcomes**

	Weight (gram)	Height (cm)	Weight- for-age z- score	Height- for-age z- score	Weight- for-height z-score	Underweight (waz <-2)	Stunting (haz <-2)	Wasting (WfH <-2)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Panel A: OLS Results</b>								
Family size	-177.0*** (46.63)	-0.755*** (0.158)	-0.114*** (0.0295)	-0.182*** (0.0344)	-0.00350 (0.0270)	0.0371*** (0.0132)	0.0483*** (0.0119)	0.00962 (0.00915)
<b>Panel B: IV Results</b>								
<b>First stage</b>			<b>0.026** (0.010)</b>					
Family size	450.8 (1418)	7.364 (6.658)	-1.499 (1.162)	-0.858 (1.253)	-0.639 (1.252)	-0.0286 (0.417)	0.00435 (0.306)	0.143 (0.323)
N	10107	10113	10136	10136	10136	10136	10136	10136

F test of excluded instruments:

F( 1, 28) = 5.73

Prob > F = 0.0236

*Notes:* Family size is the number of 0-59 months old children in the family at the time of the survey. All models include child's age, birth order, birth size, gender, religion, caste of the household, rural dummy, mother's education, father's education, mother's age, father's age, socio-economic status of the household and state fixed-effects. Standard errors clustered by state are reported in parentheses. Data source: NFHS

# Conclusion

- We find strong evidence of Q-Q trade-off in India
- The effect differed by wealth gradient, caste, and mother's education
- Finally, we do not find any evidence of quantity-quality trade-off in health outcomes
- Better access to family planning methods might help accumulation of human capital in developing countries

# The End

Thank you!

Questions?