#### **THE GRADUATE INSTITUTE | GENEVA**

**INSTITUT DE HAUTES ETUDES** INTERNATIONALES ET DU DÉVELOPPEMENT

GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

#### Graduate Institute of International and Development Studies Working Paper No: 05/2011

# Parental Height and the Sex Ratio

Jean-Louis Arcand

Graduate Institute of International Studies and Development Studies

## Matthias Rieger

Graduate Institute of International Studies and Development Studies

#### Abstract

This paper tests the generalized Trivers Willard hypothesis, which predicts that parents with heritable traits that increase the relative reproductive success of males compared to females will have relatively more males than females. As in Kanazawa (2005) we test if taller mothers have relatively more sons in a pooled sample of Demographic Health Surveys (DHS) from 46 developing countries. Despite using a rich dataset and an array of statistical models that address some of the concerns raised by Gelman (2007), we provide further evidence against the hypothesis.

© The Authors. All rights reserved. No part of this paper may be reproduced without the permission of the authors.

# <span id="page-1-0"></span>Parental Height and the Sex Ratio

Jean-Louis Arcand<sup>∗</sup> Matthias Rieger†

December 2010

#### Abstract

This paper tests the generalized Trivers Willard hypothesis, which predicts that parents with heritable traits that increase the relative reproductive success of males compared to females will have relatively more males than females. As in [Kanazawa](#page-6-0) [\(2005\)](#page-6-0) we test if taller mothers have relatively more sons in a pooled sample of Demographic Health Surveys(DHS) from 46 developing countries. Despite using a rich dataset and an array of statistical models that address some of the concerns raised by [Gelman](#page-6-1) [\(2007\)](#page-6-1), we provide further evidence against the hypothesis.

Keywords: Evolutionary psychology; sex ratio; Generalized Trivers Willard hypothesis (gTWH); height

<sup>∗</sup>Graduate Institute of International and Development Studies, Pavillion Rigot, Avenue de la Paix 11A, 1202 Genève. E-mail: Jean-Louis.Arcand@graduateinstitute.ch

<sup>†</sup>Corresponding Author: Graduate Institute of International and Development Studies, Pavillion Rigot, Avenue de la Paix 11A, 1202 Genève. E-mail: matthias.rieger@graduateinstitute.ch

## 1 Introduction

This paper tests the generalized Trivers Willard hypothesis as proposed by [Kanazawa](#page-6-0) [\(2005\)](#page-6-0), which predicts that parents with heritable traits that increase the relative reproductive success of males compared to females will have a lower-than-expected offspring sex ratio. The hypothesis is based on Trivers and Willard (1973), who argue that parents may alter the offspring sex ratio to increase reproductive success, depending on their material and nutritional conditions. More specifically, for species in which the male fitness variance is higher than the female fitness variance, male offspring of parents in good conditions will reproduce more successfully than their female siblings, since they are physically superior to their reproductive competitors and can thus monopolize mating opportunities. Conversely, their female siblings cannot increase their output of offspring with ease, as for them reproduction is associated with lengthy and costly investments. Consequently, parents in good material and nutritional conditions should prefer boys over girls to maximize reproductive success. In a variant of this hypothesis, [Kanazawa](#page-6-0) [\(2005\)](#page-6-0) finds that taller and bigger parents have relatively more boys than girls in Britain's National Child Development Survey and the British Cohort Survey. This findings is striking, as one expects that in a modern society body mass and height has lost some of its importance for reproductive success. However subsequent studies by [Denny](#page-6-2) [\(2008\)](#page-6-2) on British data, as well as [Pollet and Nettle](#page-6-3) [\(2010\)](#page-6-3) on British and Guatemalan data have rejected the hypothesis. Here we provide the first evidence against the hypothesis for a large number of developing countries.

As in to [Kanazawa](#page-6-0) [\(2005\)](#page-6-0) we test if taller mothers have relatively more sons in a pooled sample of Demographic Health Surveys(DHS) from 46 developing countries. Despite using a rich data set and an array of statistical models that address some of the concerns raised by [Gelman](#page-6-1) [\(2007\)](#page-6-1), we find no clear evidence in favor of the hypothesis in this particular sample of developing countries.

## 2 Empirical Model

<span id="page-2-0"></span>We employ two empirical models to test whether maternal height has an impact on the offspring sex ratio. In a first instance, we follow [Kanazawa](#page-6-0) [\(2005\)](#page-6-0) by estimating two separate equations for the total number of girls and boys for each mother. Thereafter, we examine the impact of maternal height on the fraction of boys out of total births for each mother as suggested by [Gelman](#page-6-1)  $(2007)$ . Let m denote mothers, h households, and let  $N$  be the sample of mothers:

$$
girls_{mh} = x_{mh}\alpha + height_{mh}\beta_{girls} + \varepsilon_{mh} \tag{1}
$$

$$
boys_{mh} = x_{mh}\alpha + height_{mh}\beta_{boys} + \varepsilon_{mh} \tag{2}
$$

<span id="page-3-0"></span>where  $girls_{mh}$  and  $boys_{mh}$  is the  $N \times 1$  vector associated with the number of female and male births of mother m in household h,  $x_{mh}$  is an  $N \times K$  matrix of mother and household characteristics, and  $\varepsilon_{mh}$  is a disturbance term. In this setting the test of the generalized Trivers Willard hypothesis boils down to a simple test of the equality of the coefficients on maternal height across the two regressions.[1](#page-1-0)

The test is only meaningful if we can consistently estimate the impact of maternal height on our two outcome variables.To see this compose the disturbance term into two components:

$$
\varepsilon_{mh} = \lambda_{mh} + \eta_h \tag{3}
$$

where  $\lambda_{mh}$  represents mother-level unobservables that affect the outcome, while  $\eta_h$  are household-level unobservables.

There is a possibility that OLS estimates of [\(1\)](#page-2-0) and [\(2\)](#page-3-0) will lead to inconsistent estimates of  $\beta_{bops}$  and  $\beta_{girls}$ , since maternal height may be correlated with mother-level unobservables  $\lambda_{mh}$  and household-level unobservables  $\eta_h$ . To control for household-level unobservables we make use of our rich and extensive dataset, which features an average of 1.2 mothers per household due to extended families and polygamous households in developing countries. Mother-specific fixed effects cannot be introduced, as variables such as  $height_{mh}$  can then no longer be identified. As a result, we only include fixed effects at the hierarchically higher household level and thus rely on within-household variation. To control for remaining bias we include a series of mother level controls such as age, education in years, a marriage dummy, a household head dummy and a variable indicating whether the mother is married to the household head. Addressing the statistical critique of [Gelman](#page-6-1) [\(2007\)](#page-6-1), we will not include highly endogenous co-variates such as the number of girls in the number of boys regression and vice versa. Finally, contrary the classical regression model, our response variable is discrete and its distribution features nonnegative integer values only. The natural choice to analyze fertility is a Poisson regression, which we estimate as a further robustness check for the OLS and household fixed effects models.

The second empirical strategy estimates the impact of maternal height on the fraction

<sup>&</sup>lt;sup>1</sup>Here we use a simple t-statistic:  $t = \frac{(\beta_{boys} - \beta_{girls})}{\sqrt{(N_{axis} - \beta_{girls})} \sqrt{(N_{axis} - \beta_{phys})}}$  $\frac{(p_{boys} - p_{girls})}{(Var(beta_{boys}) + Var(beta_{girls}))}$ . An alternative is the F-test proposed in the classic textbook by [Sokal and Rohlf](#page-6-4) [\(1964\)](#page-6-4). Also if one estimated the two equations simultaneously we could account for the correlation between unobservables in the two equations.

of boys out of total births for each mother:

<span id="page-4-0"></span>
$$
fraction_{mh} = x_{mh}\alpha + height_{mh}\gamma + \varepsilon_{mh} \tag{4}
$$

where  $fraction_{mh} = \frac{bogs}{bogs + girls}$  is the  $N \times 1$  vector associated with the fraction of male births out of total births of mother m in household h,  $x_{mh}$  is an  $N \times K$  matrix of mother and household characteristics, and  $\varepsilon_{mh}$  is a disturbance term. Now the test of the generalized Trivers Willard hypothesis amounts to testing whether  $\gamma$  is statistically significant and positive. As in the count data models we introduce maternal control variables and household fixed effects. Our baseline results are estimated by OLS and using fixed effects. In a series of robustness checks, we estimate a fractional logit, a Mundlak-procedure fractional logit<sup>[2](#page-1-0)</sup> and a fixed effects poisson to take into account that the left hand side variable is a fraction.

### 3 Data

We use cross-sectional data from the Demographic and Health Surveys (DHS). This allows one to test the generalized Trivers Willard hypothesis across many countries using highly comparable data. To increase comparability, we restrict ourselves to countries that have had a standard DHS during the last two rounds (at least DHS-5 or DHS-4) and feature maternal height. The final data set pools surveys from 46 developing countries, which are listed in Table [1.](#page-7-0) Our sample consists of 399,733 mothers for which we have quality information on height, a complete birth history at the time of the survey as well as a wide array of control variables. Table [2](#page-8-0) gives basic summary statistics of the pooled sample. Mothers in the sample have on average 1.80 boys and 1.72 girls, which amounts to a mean sex ratio of 0.5158437. The mean height of mothers is 1.56. Mean age is 32.5 years, ranging from 13 to 49 years. In the analysis we provide estimates for the full sample as well as for mothers aged 35 years or more, as well as 40 years and older. A majority of mothers are wives of household heads and the mean household in the sample features 1.28 mothers.

### 4 Results

Baseline estimates of the impact of maternal height on the number of male and female births as modeled in equations [\(1\)](#page-2-0) and [\(2\)](#page-3-0) are presented in Table [3.](#page-9-0) In columns (1) and (2) give simple OLS estimates, in columns (3) and (4) we add maternal and household

<sup>&</sup>lt;sup>2</sup>More specifically, we are running a random effects fractional logit with household means to control for unobserved heterogeneity at the household level. This involves a relatively mild linearity assumption in terms of the manner in which household-specific unobservables enter the specification (see for instance [Wooldridge](#page-6-5) [\(2001\)](#page-6-5), pp.487 for a discussion).

controls, while the last two columns include household fixed effects. Across specifications, maternal height has a positive and significant effect on the number of boys and girls. In the OLS estimates the coefficient in the regression for girls is slightly lower than in the regression for boys, however this difference is not statistically different from zero at usual levels of confidence. Adding control variables to the regressions increases the coefficients in both regressions, suggesting that omitted variables bias OLS estimates. However both coefficients are still statistically indistinguishable. Introducing household fixed effects leads to substantially lower coefficient estimates. In particular, the coefficient in the regressions for boys is substantially lower. Nevertheless, the t-statistic suggests that both coefficients are equal viz. we reject the generalized Trivers Willard hypothesis.

We provide two robustness check of these first results by restricting the age range of mothers and estimating a poisson regression to take into account the count nature of our dependent variables. Restricting the age range to  $[35 - 49]$  and  $[40 - 49]$  leads to a rejection of the hypothesis, as we fail to reject the equality of coefficients across OLS and fixed effect estimates. The poisson estimates, as well as the fixed effects poisson estimates on the full and restricted samples provide no evidence of the hypothesis that maternal height has a differential impact on the number of male and female births.

Estimates of equation [\(4\)](#page-4-0), with the fraction of boys as dependent variable leads, to an even clearer rejection of the hypothesis. OLS and fixed effect estimates are presented in Table [7.](#page-13-0) For the full sample, the impact of maternal height on the fraction of boys is in stark contrast to the theoretical prediction, since the associated coefficient is negative. However, this impact is only statistically significant in the OLS models in columns (1) and (2). Once we control for household unobservables, statistical significance vanishes. In columns  $(4)$ , $(5)$ , $(6)$  and  $(7)$  of the same table, we restrict the age range of mothers. In the OLS models, the coefficient estimate is still negative and significant for the age range of  $[35 - 49]$ , but it is insignificant in the range of  $[40 - 49]$ . Once we introduce household fixed effects estimates are no longer significant and positive. These results imply that unobservables may lead to spurious results in a simple OLS model. In any case, we find no evidence in favor of the hypothesis. In a final robustness check, we take into account the fractional nature of our left-hand-side variable by estimating fractional logits, a random effects fractional logit with a simple Mundlak procedure, as well as a fixed effect Poisson. Again the coefficient on maternal height is never positive and statistically significant at usual levels of confidence.

# 5 Conclusion

In the spirit of [Kanazawa](#page-6-0) [\(2005\)](#page-6-0) we tested whether taller mothers have relatively more sons in a pooled sample of Demographic Health Surveys (DHS) from 46 developing countries. In this particular sample we find no clear evidence in favor of the hypothesis despite employing a wide range of empirical models and robustness checks. This also confirms findings by [Denny](#page-6-2)  $(2008)^3$  $(2008)^3$  $(2008)^3$  on British data, as well as [Pollet and Nettle](#page-6-3)  $(2010)$  on British and Guatemalan data.

# References

- <span id="page-6-2"></span>Denny, K. (2008): "Big and tall parents do not have more sons," Journal of Theoretical  $Biology, 250(4), 752-753.$
- <span id="page-6-1"></span>Gelman, A. (2007): "Letter to the editors regarding some papers of Dr. Satoshi Kanazawa," Journal of Theoretical Biology, 245(3), 597 – 599.
- <span id="page-6-0"></span>Kanazawa, S. (2005): "Big and tall parents have more sons: Further generalizations of the Trivers-Willard hypothesis," Journal of Theoretical Biology, 235(4), 583 – 590.
- <span id="page-6-3"></span>POLLET, T. V., AND D. NETTLE (2010): "No evidence for the generalized Trivers-Willard hypothesis from British and rural Guatemalan data," Journal of Evolutionary Psychology, 8(1), 57–74.
- <span id="page-6-4"></span>SOKAL, R. R., AND F. J. ROHLF (1964): *Biometry: the principles and practice of* statistics in biological research. W. H. Freeman and Co, San Francisco.
- <span id="page-6-5"></span>WOOLDRIDGE, J. (2001): *Econometric Analysis of Cross-Section and Panel Data.* MIT Press, Cambridge, MA, 1st edn.

<sup>&</sup>lt;sup>3</sup>[Denny](#page-6-2) [\(2008\)](#page-6-2) uses a different strategy of testing the theory by estimating the probability of a male birth as a function of the size of each parent.

<span id="page-7-0"></span>

Country	DHS-Code	N	Perc. of Sample		
Armenia	AM4	4092	1.02		
Azerbaijan	$\rm AZ5$	5076	1.27		
Bangladesh	BD <sub>5</sub>	9723	2.43		
Burkina Faso	BF4	9328	2.33		
Benin	BJ5	13070	3.27		
<b>Bolivia</b>	BO <sub>5</sub>	11456	2.87		
<b>DRC</b>	CD5	3438	0.86		
Congo	CG5	5034	1.26		
Cameroon	CM4	3729	0.93		
Columbia	CO <sub>4</sub>	24968	6.25		
Egypt	EG5	14690	3.67		
Ethiopia	ET4	4431	1.11		
Gabon	GA <sub>3</sub>	2791	0.7		
Ghana	GH <sub>5</sub>	3243	0.81		
Guinea	GN4	3129	0.78		
Honduras	HN <sub>5</sub>	13336	3.34		
Haiti	HT <sub>5</sub>	3195	0.8		
India	IA <sub>5</sub>	81146	20.3		
Jordan	JO5	4759	1.19		
Kenya	KE5	6013	1.5		
Cambodia	KH <sub>5</sub>	5382	1.35		
Kazakhstan	KK3	1626	0.41		
Lebanon	LB5	5596	1.4		
Lesotho	LS4	2349	0.59		
Morocco	MA4	8595	2.15		
Moldova	MB4	4828	1.21		
Madagascar	MD <sub>5</sub>	6311	1.58		
Mali	ML <sub>5</sub>	11406	2.85		
Malawi	MW4	8878	2.22		
Mozambique	MZ4	9241	2.31		
Nicaragua	NC4	8974	2.24		
Nigeria	NG5	23115	5.78		
Niger	NI5	3555	0.89		
Namibia	NM5	6466	1.62		
Nepal	NP <sub>5</sub>	7753	1.94		
Peru	PE4	18208	4.56		
Rwanda	RW4	3510	0.88		
Sierra Leone	SL5	2872	0.72		
Senegal	SN4	2971	0.74		
Swaziland	SZ5	3390	0.85		
Chad	TD4	3582	0.9		
Turkey	TR4	3288	0.82		
Tanzania	TZ4	7517	1.88		
Uganda	UG5	2169	0.54		
Zambia	ZM <sub>5</sub>	5340	1.34		
Zimbabwe	ZW <sub>5</sub>	6164	1.54		
Total		399,733	100		

Table 1: Sample of Demographic and Health Surveys (DHS) across 46 countries.

<span id="page-8-0"></span>

Variable	Mean	Std. Dev.	Min	Max
Nr. of Sons	1.80	1.50	$\mathcal{O}$	14
Nr. of Daughters	1.72	1.51	$\left( \right)$	14
Fraction of Sons	0.52	0.33	$\left( \right)$	1
Height of Mother	1.56	0.07	1	$\overline{2}$
Age of Mother	32.54	8.58	13	49
Education of Mother in Yrs.	5.08	4.86	$\left( \right)$	27
Married	0.73	0.44	$\mathbf{0}$	1
Mother is Household Head	0.13	0.33	∩	
Wife of Household Head	0.65	0.48	$\mathbf{0}$	1
Age of Household Head	43.51	13.15	13	97
Sex of Household Head	0.80	0.40	$\mathbf{0}$	
Mothers per Household	1.28	0.62		13

Table 2: Summary statistics of the pooled sample of 46 Demographic and Health Surveys from 46 countries.

<span id="page-9-0"></span>

Table 3: OLS and household fixed effect models of the impact of mother's height on the total number of girls and boys per mother. Robust standard errors are given below coefficient estimates and have been clustered at the household level for the fixed effect models.<br>Significance levels are denoted + 0.10 \* 0.05. Table 3: OLS and household fixed effect models of the impact of mother's height on the total number of girls and boys per mother. Robust standard errors are given below coefficient estimates and have been clustered at the household level for the fixed effect models. Significance levels are denoted  $+$  0.10  $*$  0.05.



different motherly age samples. Robust standard errors are given below coefficient estimates and have been clustered at the household level for the fixed effect models. Significance levels are denoted  $+0.10 * 0.05$ . Table 4: OLS and household fixed effect models of the impact of mother's height on the total number of girls and boys per mother for Table 4: OLS and household fixed effect models of the impact of mother's height on the total number of girls and boys per mother for different motherly age samples. Robust standard errors are given below coefficient estimates and have been clustered at the household level for the fixed effect models. Significance levels are denoted + 0.10 \* 0.05.



Table 5: Poisson and household fixed effect poisson models of the impact of mother's height on the total number of girls and boys per Table 5: Poisson and household fixed effect poisson models of the impact of mother's height on the total number of girls and boys per mother. Standard errors are given below coefficient estimates. Significance levels are denoted  $+0.10 * 0.05$ . mother. Standard errors are given below coefficient estimates. Significance levels are denoted + 0.10 \* 0.05.



mother over various motherly age samples. Standard errors are given below coefficient estimates. Significance levels are denoted + 0.10 Table 6: Poisson and household fixed effect Poisson models of the impact of mother's height on the total number of girls and boys per Table 6: Poisson and household fixed effect Poisson models of the impact of mother's height on the total number of girls and boys per mother over various motherly age samples. Standard errors are given below coefficient estimates. Significance levels are denoted + 0.10 \* 0.05.

<span id="page-13-0"></span>

Table 7: OLS and household fixed effect models of the impact of mother's height on the fraction of boys of total births per mother for different motherly age samples. Robust standard errors are given below coefficient est Table 7: OLS and household fixed effect models of the impact of mother's height on the fraction of boys of total births per mother for different motherly age samples. Robust standard errors are given below coefficient estimates and have been clustered at the household level for the fixed effect models. Significance levels are denoted + 0.10 \* 0.05.



Table 8: Fractional logit, random effects fractional logit and poisson fixed effect models of the impact of mother's height on the fraction of boys of total births per mother. Standard errors are given below coefficients estimates. Significance levels are denoted  $+0.10 * 0.05$ .