

Accepted Version, 10/2009

Please refer only to the edited 2011 Demography copyrighted version for quotes.

Son preference in Indian Families: Absolute Versus relative Wealth Effects*

Sylvestre Gaudin
Oberlin College

Sylvestre Gaudin
Oberlin College, Dept. of Economics
10, North Professor St.
Oberlin, Ohio 44074

Email: sylvestre.gaudin@oberlin.edu/slygaudin@yahoo.com

Tel: 440-775-8483 / 301-547-5095

Fax: 440-775-6978

* Initial work for this paper was done while the author was visiting research associate with the University of Maryland, AREC, and research affiliate with the World Bank, Washington D.C. In addition to her host institutions, the author thanks Jayati Datta-Mitra and Gautam Datta for substantial comments. Acknowledgements also go to David Bishai, Barbara Craig, Hirschel Kasper, Kala Krishna, Imran Lalani, Yana Rogers, and Abdo Yazbeck. This revised version owes a great deal to comments of three anonymous referees. All analyses and remaining errors are the author's responsibility.

ABSTRACT

The desire for male children is prevalent in India, where son preference has been shown to affect fertility behavior and intrahousehold allocation of resources. Economic theory predicts less gender discrimination in wealthier households, but demographers and sociologists have argued that wealth can exacerbate bias in the Indian context. I argue that these apparently conflicting theories can be reconciled and simultaneously tested if one considers that they are based on two different notions of wealth: one related to resource constraints (absolute wealth), and the other to notions of local status (relative wealth). Using cross-sectional data from the 1998–1999 and 2005–2006 National Family and Health Surveys, I construct measures of absolute and relative wealth by using principal components analysis. A series of statistical models of son preference is estimated by using multilevel methods. Results consistently show that higher absolute wealth is strongly associated with lower son preference, and the effect is 20%–40% stronger when the household’s community-specific wealth score is included in the regression. Coefficients on relative wealth are positive and significant although lower in magnitude. Results are robust to using different samples, alternative groupings of households in local areas, different estimation methods, and alternative dependent variables.

Keywords: Gender, India, Multilevel Model, Relative Wealth, Son preference

INTRODUCTION

To “promote gender equality and empower women” is one of the eight Millennium Development Goals that 191 UN countries pledged to meet by 2015. Gender equality is seen as both a development goal and a necessary condition for sustainable development. Economists tend to think of economic development and gender equity as reinforcing: reduced gender discrimination facilitates income growth and higher incomes move societies toward less gender bias. While many have convincingly argued and empirically shown the negative impact of gender discrimination on economic development (e.g. Klasen 1999; Scanlan 2004; Sen 1984; Sweetman 2002; World Bank 2001), the other side of the causality, i.e. the role of income growth in altering gender biases, has received little attention. Recent evidence indicates that gender bias — measured using juvenile sex ratios and mortality — is generalizing and intensifying in India (Agnihotri 2003; Basu 1999, 2000; Murthi, Guio, and Dreze 1995; Premi 2001; Rajan, Shudha, and Mohanachandran 2000). Taking son preference as a barometer of girl neglect in India, a recent publication of the International Center for Research on Women using data from the first wave of the National Family and Health Survey (NFHS-1) concluded that “wealth and economic development do not reduce son preference” (Pande and Malhotra 2006). In light of these facts, it is important to reconsider the linkages between wealth and anti-female biases.

Different theories drawn from the economics, sociology, demography, and anthropology literatures provide conflicting predictions about the impact of increased wealth on gender bias in India. This paper does not add to these theories but argues and empirically shows that they rely on different notions of wealth: theories predicting less bias consider absolute notions of wealth based on resource constraints; those predicting more bias consider relative notions of wealth based on local status. Using a single measure of wealth in multivariate analyses of gender bias could capture either or both effects. If the effects are indeed opposite, omitting relative wealth would cause to underestimate the effect of absolute wealth.

To test this hypothesis, a series of nested multivariate models of son preference are estimated using cross-sectional data from the 1998-1999 and 2005-2006 NFHS. The dependant variable, son preference, is measured using survey responses on “ideal” family composition. Two objections can be raised about this choice of variable to capture gender bias: one is that answers are biased by rationalization (for example a mother may declare preferring more sons

because she has more sons); the other is that son preference does not measure actual manifestations of gender bias.¹ The first issue of rationalization can be treated using appropriate controls (as in Bhat and Zavier 2003, later referenced as B&Z, and Pande and Astone 2007, later referenced as P&A). The second objection is valid in the sense that stated son preference will never be perfectly correlated with manifestation of bias in any specific area; son-preferring households may still treat their daughters as well as their sons, and girl-preferring households may still be biased against girls when it comes to education and health. However, measures of revealed bias in one area are also problematic when generalizing to overall bias; for example, finding gender bias in education does not imply bias in health decisions. Stated son preference, on the other hand, can serve as a good barometer of trends in gender bias in all areas: girls in son-preferring households, *ceteris paribus*, are at greater risk of discrimination than girls in neutral or girl-preferring households. B&Z show that the summary statistics on stated son preference from NFHS-1 and NFHS-2 are consistent with sex ratios by state from the 1990 Indian Census.

Using dependant variables based on the same NFHS questions, neither B&Z nor P&A find strong support for a negative relationship between individual wealth and son preference. P&A construct a measure of wealth using Principal Components Analysis (PCA) on a wide array of assets as is done here for absolute wealth; they find no significant positive or negative effect of wealth on son preference in NFHS-1. B&Z use NFHS-1 and NFHS-2 data and a coarser measure of standards of living; they find “[...] weak support for the possibility that the preference for sons will decline with the increase in wealth and rising wage employment of women” (Bhat and Zavier 2003: 649). Here, I construct two different measures of wealth using PCA: the absolute wealth score is obtained using the full India sample while relative wealth is calculated using a sample of households living nearby. Both variables are included in a multivariate model of son preference inspired by B&Z. I use multilevel statistical methods to take account of the hierarchical structure of the model and complex survey design. Consistent with economic theory, I find that higher standards of living at the household and macroeconomic levels are associated, *ceteris paribus*, with lower son preference. Incorporating relative wealth into the model reinforces the negative relationship between absolute wealth and son preference,

¹ A related issue is that “pure” son preference, i.e. wanting a son based on personal taste rather than a response to structural motivating factors, will be confounded in the measure. Statistically, however, such preference should cancel out with girl-preference, on average.

increasing elasticities by 20 to 40% in magnitude, depending on the estimation method and sample. Higher relative wealth is significantly associated with lower son preference although the effect applies primarily to landed households. To test whether results generalize to gender bias beyond son preference, a similar model is estimated using stated educational preferences for boys and girls as the dependant variable. The same pattern is revealed, although the role of land ownership is no longer prevalent. Results are checked using alternative estimation methods with alternative definitions of son preference as well as an alternative delineation of local areas for relative wealth calculations. Details of these analyses are given in supplements available online on the Demography website.² In all cases, qualitative results support the main hypothesis.

THEORETICAL FOUNDATIONS

There is no consensus across disciplines about the impact of increased wealth on gender bias in India. Economists generally predict a decrease in gender bias with increased wealth, at least after a certain level of development, whereas sociologists and anthropologists predict an increase in gender bias with increased wealth, based on features of Indian society. In this section, I review the arguments on both sides, separating out those based on changes in resource constraints from those related to one's economic position relative to others.

Theories Based on Absolute Notions of Wealth

Economic models of resource allocation within the family (including fertility choices) generally consider some type of maximization of expected returns from having and raising male/female children given a budget constraint. Increases in wealth are represented by outward shifts of the budget constraint. The shifts represent increases in the quantity of resources available relative to a previous state, not relative to other households. The budget constraint may shift out due to pure income effects—family resources increase while relative expected returns from boys and girls do not change— or from both income and substitution effects—e.g. when expected net benefits from sons increase less rapidly than expected net benefits from daughters, an evolution that would be expected with economic development if wage gaps narrow. Larger budget sets may reduce bias

² In the remainder of the article, I will make frequent reference to five online supplements available on the Demography Website (<http://www.springerlink.com/content/x24126468214t642/supplementals/>.) I will simply refer to them as “Supplement 1-5”.

through several direct or indirect interrelated channels: reduced competition for resources within the household, a narrowing of real gender-based wage gaps, and reduced fertility. Theories range from intra-household allocation decisions motivated by utility maximization and/or bargaining, to representative-agent macroeconomic growth models.

In Rosenzweig and Schultz (1982), the household chooses investment expenditures on boys and girls to maximize expected utility. Anticipations of higher returns motivate a larger share of expenditures on male children. If having a child is itself an investment, this could also explain son preference. Since the inequality is generated through budget constraints, relaxation of the constraint should reduce bias. Galor and Weil (1996) model the relationship between gender gaps, fertility, and growth in an overlapping-generations model. As the stock of capital grows, women's relative wages increase (they assume that capital is more complementary to women's labor than to men's) so the opportunity cost of having children increases, lowering fertility. Lower fertility in turn increases the amount of capital per worker, resulting in higher relative wages for female workers. Their argument implies that women participate in the labor force, or at least that the increase in the wage of women relative to men provides greater incentives to do so. Such participation in the labor force creates positive feedbacks toward wage equality. The model does not address son preference but it is clear that the difference in expected lifetime earnings between boys and girls would decrease as the wage gap narrows, and this should in turn contribute to reducing son preference. Similarly, higher expected earnings would motivate greater educational investments in girls (Behrman *et al.* 1999, Kingdon 2005). Zhang, Zhang, and Li (1999) propose an endogenous growth model in which parent's expectation of support for old-age perpetuates culturally and religiously induced son preference because greater support is expected from children who can accumulate greater human capital. Son preference is an argument of the parent's utility function as well as a result of differential expected returns. If old-age support comes only from sons, bias continues with growth but if both children can accumulate human capital and girls can provide old age support, gender bias decreases with growth. Taking the argument further and assuming growth promotes better functioning capital markets, old age support would become less of a concern in wealthier households who would have less reason to be son-preferring.

The relationship between wealth and gender bias may become negative only after a certain level of wealth is reached. Kanbur and Haddad (1994) introduce the possibility of such

U-shaped relationship between wealth and intra-household inequality using a bargaining framework. They do not consider resources spent on children but the division of additional resources (pure income effects) within the household. Theoretical solutions to the bargaining model result in Kuznets' effects when there are increasing returns in production and differences in productivity between the two household members (in particular when the members' outside options do not change proportionately with joint production). If intra-household equality is associated with less gender bias, the theory would predict an eventual leveling of expenditures on boys and girls and lower son-preference as household resources continue to grow. At the aggregate level, we may also find Kuznets effects if income effects first dominate substitution effects in the labor supply decision of women (Goldin 1995). When income effects dominate, higher incomes reduce female labor force participation, thus reducing women's contribution to household wealth. Considering some evidence that gender bias at the household level is negatively related to the contribution of the female member to household resources (Agnihotri 2000), such reduction contributes to son preference. As economic activity expands, however, opportunities for women in the labor market become more lucrative. Substitution effects start working toward increased female labor force participation, creating positive feedbacks toward gender equality as in Galor and Weil (1996).

Theories Based on Relative Notions of Wealth

This section requires some background specific to the Indian context. There are three interrelated dimensions to socio-economic status in India: economic, political, and ritual rankings (Srinivas 1962). Srinivas describes "high" caste or castes that have achieved a high ritual status as "Sanskritized". In Sanskritized castes, women are subordinate to men in religious rituals; they do not work outside the home (or do not leave the home in some cases); they need to be married, have male children, and have male grandchildren. They cannot initiate a divorce and they cannot remarry as widows. Parents are responsible for finding a suitable husband for their daughter, generally within the caste, and finding a husband of higher status than one's own, a practice called hypergamy, is looked upon positively. Not being successful in arranging a marriage for one's daughter, however, reflects badly on the unmarried daughter and her family. Parents of the bride pay all marriage costs and give some dowry and/or some substantial cash payment to the family of the groom. Once married, the daughter is completely committed to the family of her

husband and can no longer provide emotional, physical, or economic support to her own family, so parents can only count on male children as support for old age (Arnold, Chow, and Roy 1998; Dyson and Moore 1983). The more Sanskritized a caste and the higher one's status in the caste hierarchy (status being determined on all three axes mentioned above), the more costly it is to have a daughter instead of a son and the more barriers there are for women to integrate into the productive economy. Modern dowry (Srinivas 1989) or groomprices (Billig 1992) have become more and more common and have spread to all areas of India.³

Within this context, several (interrelated) channels have been identified linking rising gender biases to rising income. A wide body of research has pointed to the persistence of son preference combined with increased availability of sex determination techniques and voluntary pregnancy terminations as largely responsible for the increasing number of "missing women" (e.g. Agnihotri 2003; Basu 1999; Das Gupta 2005; Sen 2003). It is a fact that wealth increases the ability to use such techniques but the desire to do so is a phenomenon of son-preferring households only. If son preference is shown to decrease with wealth, fewer families would practice sex selection and sex-ratios would eventually improve. The next arguments relate to the prevalence of son preference itself. Most compelling is the "prosperity effect" argument, also often loosely called "Sanskritization".⁴ A simplified version of the argument is that individuals in families or castes of lower ritual status who have the economic means to do so feel inclined to emulate the behavior of higher castes (Agnihotri 2000, Basu 1999, 2000; Berreman 1993; Srinivas 1962, 1989). Following from above, while ritual status (related to caste), economic status (related to wealth), and political status are different concepts that need not be related, the three axes of the hierarchy are very much intertwined. Srinivas (1962) argues that the highest possible place in the social hierarchy includes a high ranking on all three axes. A family with high ritual status will likely seek a strong economic position to reinforce its social status, and a family or a group of families that have achieved relative economic prosperity will likely seek to improve their ritual status. The "prosperity effect" implies that gender bias/son preference is likely a feature of richer households, regardless of caste. Another possible channel through

³ A groomprice is, according to Billig (1992), any payment in kind or cash that can be used by the husband's family as they please (whereas a dowry remains the property of the daughter). Srinivas calls it "modern dowry."

⁴ The term "Sanskritization", as it was introduced by Srinivas (1962), refers to a phenomenon deeply rooted in the caste system and more specifically related to group behavior; "prosperity effect" is used more generally to apply to individual households.

which wealth can increase son preference is related to rising marriage costs (groomprices, dowries, and wedding ceremonies). Billig (1992) notes that groomprices have become common practice in the South and the amounts of money involved have risen significantly (see also Basu 1999). Anderson (2003) shows theoretically that economic development can explain increases in modern dowries in India due to the effect of wealth dispersion in a caste-based society. Higher groomprices and dowry values can cause greater son preference (or daughter aversion) since they increase the expected costs of having a daughter and the expected benefits of having a son. At the household level, this argument is not sufficient to explain why richer households would be more son-preferring, unless they are expected to spend relatively more. Given the hypergamous tradition, however, one must also take account that the expected gains from marriage may depend on one's opportunity to gain status via marriage. If wedding ceremonies are themselves used to signal social status (as argued in Bloch, Rao, and Desai 2004), higher ranked families will be expected to give grand weddings just to maintain their status while families of lower social status will incur the expense in order to gain full social benefits when their daughter marries into a family of higher status. Given that it is looked upon positively to marry upward for a girl but not for a boy, a daughter can become more valuable than a boy in terms of gaining status, as long as there are wealthier families to marry into:

"[...] families seeking social advancement compete among themselves to amass a dowry large enough to secure a place for their daughter in an elite household. This brings a prestigious alliance for parents along with prospects of well-endowed grandsons. [...] At the top of the hierarchy, however, hypergamy dooms daughters. There is no family for them to marry into." (Blaffer Hrdy 1999: 325-6).

The problem faced by higher ranked families in this context can contribute to greater son preference, while aspiration for social advancement via marriage in lower-ranked families may explain some preference for daughters if they can afford such marriage –the caveat being related to notions of absolute wealth.⁵

⁵ The relevance of this argument depends on conditions of the marriage market. The less geographically segmented the market and the less costly it is to establish contact with families further away, the less important narrowly defined relative wealth should be. Further information on this point is provided online (Supplement 5.)

Arguments related to Sanskritization, rising marriage costs, and hypergamy, all suggest that the positive link between wealth and gender bias relates more to relative notions of socio-economic status within the community rather than to standards of living.

Absolute versus Relative Wealth

Although there is often confusion between the two concepts, the case for distinguishing between absolute and relative wealth in economics has been made before. Frank (1985) sees the value of local status as a major driver of economic decisions. He stresses, in particular, the role of imitative behavior: “Indeed, the most sensible strategy might involve focusing on the behavior of those positioned above ourselves on whatever index of “success” we find most important. After all, if what they are doing seems to work for them, then why not for us?” (Frank 1985:19). In issues related to health access and outcomes, Deaton (2003) highlights the importance of distinguishing between income and other related dimensions of well-being such as education, wealth, control, or rank. He argues that all these variables have a potential effect on health, independently of income or even income dispersion; “control” and “rank”, in particular, are dimensions directly linked to relative wealth.

If the theories expositied above hold true in the data, the direction of the effect from wealth to son preference must depend on the reference base on which wealth indices are constructed. If households in the sample are not connected by a notion of community and if wealth is measured relative to the whole sample, one would expect less son preference in richer household. If, however, the country is highly segmented by community, and wealth is measured relative to households in the same community, one would expect more son preference in richer households. Since measures of absolute and relative wealth are necessarily positively correlated, the estimated effect of a unique wealth variable on son preference shall include both effects. Estimating the effect of absolute levels of wealth on son preference independently of the confounding effect of relative wealth is the focus of the empirical analysis that follows.

EMPIRICAL MODEL: DATA AND METHODS

The degree of son preference (SP) of an individual (i) at one point in time is assumed to be a function of various characteristics of the individual, her household, and the macroeconomic context in which she lives. Individual-level demographic and socio-economic characteristics are

borrowed from B&Z. At the household level, two measures of wealth are considered: absolute wealth (W) measures the household's standard of living and relative wealth (WR) measures the economic standing of the household in its community. Land ownership, generally considered a prime element of socio-economic status in rural areas, is also considered independently of its contribution to W and WR . The macroeconomic level is included to account for the effect of economic development in terms of goods and services available to all.

Data

Individual and household-level variables are constructed from the 1998-1999 and 2005-2006 National Family and Health Surveys of India (NFHS-2 and NFHS-3). NFHS is the Indian version of the Demographic and Health Surveys (DHS); it has been widely used for academic research and by international organizations since its first wave (1992-1993).⁶ The NFHS data are organized as individuals within households within primary sampling units (PSUs), within states. PSUs are villages (in rural areas) or census blocks in urban areas) selected with probability inverse to size to form the rural and urban samples in each states. Households are selected randomly within each PSU. The full sample of households is used to calculate principal component (PC) absolute and relative wealth scores. The household data and associated wealth scores are then merged with the individual level data. All ever-married women between the ages of 15 and 49, present in the selected household at the time of the visit are interviewed. After excluding visitors and a few observations with missing data or unusable information on key variables, the samples consist of 77, 886 women in NFHS-2 and 83,785 in NFHS-3. Overall, dropped observations on independent variables are evenly distributed in the data in terms of levels of son preference, education, and income (Supplement 1, section 1).

Dependant variable: Son preference

Both NFHS-2 and NFHS-3 questionnaires included questions on ideal family size. In particular, women with living children were asked: "If you could go back to the time when you did not have any children and could choose exactly the number of children to have in your whole life, how many would that be?"; women with no living children were asked: "If you could choose exactly the number of children to have in your whole life, how many would that be?" Women who

⁶ Details of the surveys can be found in IIPS (2000 & 2007); both reports as well as the full data sets are available online at www.measuredhs.com.

answered the first question with a numerical answer were then asked: “How many of these children would you like to be boys, how many would you like to be girls, and for how many would the sex not matter?”⁷ *SP* is calculated using the difference between ideal number of boys and ideal number of girls divided by the ideal family size.⁸ The ratio was chosen as a measure of son preference because it provides an intuitive ordering and can capture the full range of preferences, including preference for girls (counterbalancing some “pure” preferences for boys not related to a societal notion of gender bias.) The proportion of sons in ideal family size chosen by B&Z for their OLS model was not retained because it eliminates variation on the girl preference side and, more importantly, dampens variation on the son preference side.⁹ Figure 1 gives the distribution of the dependant variable *SP*; it ranges between -1 and +1 (by construction) with a large number of observations clustered at zero, indicating no preference for either boys or girls. Means are respectively 0.13, 0.09, and 0.11 for the NFHS-2, NFHS-3, and pooled samples with standard deviations around 0.27. The numerator, the difference between ideal number of boys and ideal number of girls, lies mostly between -1 and +2 (99% of the data). The denominator, ideal family size, averages between 2 and 3. Very few women gave answers of zero (1%) and above 10 (less than 0.02%); they were excluded without creating detectable selection problems (Supplement 1, section 1). Dividing by the ideal number of total children avoids giving more weight to large ideal family sizes; this would especially be problematic here because desired total number of children and wealth level are strongly negatively correlated ($\rho = -.31$ for NFHS-2 and $-.27$ for NFHS-3).

Explanatory Variables

Please refer to Table 1 for descriptions of explanatory variables and summary statistics. Variables are separated into three groups depending on the level at which they are measured: the macroeconomic level (state), the household, and the individual. I focus here on describing the methodology used to measure variables that distinguish this paper from the existing literature.

⁷ Poorer women with no education were disproportionately represented in non-numerical answers but the number of cases was small enough to justify ignoring the bias (1, section 1).

⁸ In robustness analyses alternative definitions of the dependant variable are used: a binary variable defined as in B&Z for Logit models and a three-level ordered categorical variable defined as in P&A for the ordered Logit alternative (Supplement 4, section 3).

⁹ Consider three responses: A. 3 boys and 3 of either sex; B. 3 boys, 2 girls, and 1 of either sex; C. 3 boys and 3 girls; most people would rank A above B and B above C in terms of son preference. Using B&Z’s measure, all three responses get a value of 1/2; with the measure chosen here A scores 1/2, B scores 1/6, and C scores 0.

Economic development (macroeconomic level). A single variable is used to approximate the level of development in each state. State-level gross domestic products for 1999 and 2005 are converted in constant Rupees (thousands) and divided by state populations using the 1991 and 2001 census to obtain the real gross per capita state domestic product (*GSP/c*).¹⁰ It is anticipated that overall economic development would decrease son preference, or possibly increase son preference in the first stages of development and eventually decrease it.

Absolute Wealth. To properly measure resource constraints one would like to have consumption or income data but it is well-known that the data are not available or poorly measured in the developing world. Montgomery *et al.* (2000) critically assess the use of standard of living indices based on a small number of assets arbitrarily weighted; despite weaknesses in approximating consumption patterns, the measures are found appropriate for hypothesis testing. A finer index, based on a larger set of asset information and systematic weights is likely to better approximate standards of living. The NFHS household questionnaire provides extensive data on asset ownership (over thirty asset categories are used). Principal components analysis (PCA) weighs each asset according to its contribution to the overall variance. The full set of asset variables is used to estimate these weights (or factor scores), taking account of NFHS sample weights for households. Assets that are owned in addition to others (most assets) receive a positive factor score while assets that are owned instead of others receive a negative score. Assets owned by all households receive a zero weight since they do not contribute to the variance. Household wealth scores are then calculated as a weighted sum of all the assets they own, taking asset factor scores as weights.¹¹ Household scores are then normalized so that coefficient estimates can be compared. Let Z_i be the total score obtained for household i and let Z_{min} and Z_{max} be the minimum and maximum scores in the sample considered, the normalized score for household i is

$$W_i = \frac{Z_i - Z_{min}}{Z_{max} - Z_{min}} \quad (1)$$

¹⁰ Details on the construction of *GDP/c* are given in Supplement 1, section 2.

¹¹ The procedure is the same as in Gwatkin *et al.* (2000), Pande and Yazbeck (2003), Gaudin and Yazbeck (2006a). For a list of asset variables and results on asset weights and scores obtained with NFHS-2 data, see Gaudin and Yazbeck (2006b).

These normalized scores provide an ordinal measure of wealth levels that can be used as an index of wealth. The measure is fundamentally relative by construction but the reference base being all of India, it can be used to approximate the variation in absolute wealth within the country at a given point in time. It must be noted, however, that household wealth scores cannot be compared across the two surveys since PCA is carried out using each sample separately (asset variables differ between surveys). Because of this, the pooled analysis treats the absolute wealth scores of the two samples as two different variables, *W99* and *W06*.

Relative wealth. The relative wealth variable (*WR*) is constructed using PCA on subsamples of households living nearby. The area of influence one would ideally choose depends on the size of the population and specific customs but in general, it must be large enough to constitute a community of people but not too large so it can be differentiated from measures of absolute wealth (the larger the area considered, the greater the correlation with absolute wealth and the less meaningful the distinction). The only geographical identifiers available in both NFHS-2 and NFHS-3 that can place households as “neighbors” at the community level are the PSUs.¹² There are over 3,000 PSUs in each sample with a median size of 30 households per PSU (Table 2). PCA is run PSU by PSU using the same list of assets as for the full sample. Because PSU samples are small, tests are run to verify the behavior of PCA in small samples; they indicate that the procedure is relatively stable down to sample sizes of ten (Supplement 2). Scores obtained from the PSU-level analysis are normalized in a manner similar to equation (1) using PSU specific minimum and maximum scores. Consequently, wealth scores are completely independent between PSUs but dependent within. The overall distribution of *WR* includes a large number of zeros and ones by construction (normalization is done PSU by PSU); in between, it is close to uniformly distributed with slightly more mass below 0.4. More interesting statistically is the distribution of the variable within each PSU; Figure 2 summarizes this distribution by PSU size. Each point on the graph is obtained by taking the average of the PSU-specific *WR* means and standard deviations among all PSUs of equal size. Standard deviations stabilize around sample sizes of fifteen; since close to 99% of PSUs are larger, samples with less than fifteen households can be dropped without biasing results (Supplement 1, section 1).

¹² Supplement 3 online discusses the construction of an alternative grouping of households based on geographical identifiers available in NFHS-2.

Absolute and relative wealth scores are obviously positively correlated but a well-to-do individual by national standards may score lower in relative wealth than a much poorer individual living in a different area, if the poorer individual is relatively richer than her neighbors (as in Jaffe *et al* 2004).¹³ The correlation coefficients between the all-India principal components household wealth scores and the PSU-level scores are approximately .75 in both samples. Categorical variables were constructed to create PSU-level wealth quintiles both based on number of households and on deviations from the PSU means. The regressions did not change qualitative results nor add important information relative to using the scores.

Land. Ownership of agricultural land and the size of land holdings (cultivated land) enter in the PC calculation of both the absolute and relative measures of wealth described above, along with other asset variables. There are two reasons to give them additional consideration, both related to relative wealth considerations. First, land ownership, beyond its contribution to wealth, has traditionally been an important consideration when families in the same community evaluate each other's worth (for example in the hypergamous tradition, it is well regarded to marry a daughter in a family with lots of land). Second, ownership of agricultural land (regardless of size) may identify a group of households more likely to behave according to the theories outlined above that argue a positive link between wealth and son preference. Indeed, higher caste families traditionally owned much of the land in agricultural village economies (Chakraborty and Kim 2008) and lower-caste land-owners have been identified as the group most likely to Sanskritize (Staal, 1963). To account for the first effect, I include a continuous variable that records the number of acres of cultivated land that a household owns. Among households with agricultural land, correlation between acres of land and absolute wealth scores is around .14 in both samples; correlations with relative wealth scores are .17 in NFHS-2 and .14 in NFHS-3. The relationship between son preference and land holdings is likely very different in rural and urban areas so I allow the slopes to differ. To account for the second effect, I allow the impact of relative wealth to differ between landed and non-landed households.

Other Determinants of Son preference, Control Variables, and Fixed Effects.

Socioeconomic and demographic variables measured at the level of individuals follow closely

¹³ Jaffe *et al* (2004) also distinguish between absolute and relative measure of socio-economic status in their analysis of mortality in Israel.

those in B&Z and I refer the reader to their explanations (Bhat and Zavier 2003).¹⁴ Modifications where made regarding education variables and age; they are explained in Supplement 1, section 3.

Variables from B&Z related to family demographics are included to account for rationalization effects (possible endogeneity issues are addressed in B&Z.) Another variable from B&Z identifies observations with odd-numbered ideal family sizes; it is used because women who reported an odd-number in their ideal family size generally preferred one more son than one more daughter (true in both samples). The inclusion of this variable does make a significant difference in the estimation results, although qualitative results relative to the effects of absolute and relative income are robust to the change.

Since state-level random effects are used in the multilevel regression (explained later), I include regional fixed effects to account for differences in cultural norms that are known to vary closely with regions. In particular, the North, home of the Rajputs, is traditionally more son-preferring whereas the South's cultural traditions have been more favorable to women. I follow the Indian Census to group states into regions. Finally, as in B&Z, I include a dichotomous variable for urban residence. In the pooled regression, another binary variable captures additional differences between the two samples; in addition to time, it can account for differences in general surveying procedures.

Multilevel Estimation

Although the distribution of the dependant variable would favor a discrete probability model, the linear model was chosen for several reasons: (i) it preserves more variance in terms of degrees of son preference; (ii) it allows inferences about differential causal effects from estimated coefficients when residual variation varies across groups, which is not the case in Logit or Probit models (Allison 1999); (iii) as will be shown later and as found in B&Z, OLS results compare favorably to Logit and ordered Logit models; and (iv) the linear model facilitates the use of

¹⁴ Additional background on socioeconomic and cultural variables is given in Pande (2003) and Pande and Astone (2007).

multilevel methods.¹⁵ Compared to feasible alternatives, the linear multilevel procedure proved to be the most conservative in hypothesis testing (Supplement 4.)

Determinants of son preference come from different levels. At the top (level 1) is the state; within states are local areas (level 2); within local areas are households (level 3). Individuals (base level observations) are therefore nested into households, nested into local areas, nested into states.¹⁶ There are several well documented problems associated with the estimation of such hierarchical multilevel relationship with OLS: first, assigning values to base level observations as if they were observed for each individual when in fact they are observed at the higher level, artificially reduces standard errors and therefore increases the likelihood of type I errors (finding significant relationship when, in fact, they are not); second, observations within a household, local area, and/or state are expected to share some common unobserved characteristics, violating assumptions of independently distributed errors, unless fixed and/or random effects are included for each group. Multilevel modeling takes account of the hierarchical structure of the model and the data, focusing on between-level relationships when there is no variation within a level. Explanatory variables are separated into three categories: variables in \mathbf{X} includes explanatory variables measured at the individual level; variables in $\mathbf{X}^{(1)}$ are measured at the state level (invariant for individuals in the same state in a survey year), and variables in $\mathbf{X}^{(3)}$ are measured at the household level. There are no variables with values common to all individuals at the local level but a random effect is included because the relative wealth variable is constructed local area by local area and each area coincides with a PSU, itself constituting a cluster in the NFHS design. Intercept fixed effects are included in \mathbf{X} using dummy variables for region, urban/rural residence, and survey wave (for the pooled regression). Using subscripts s , l , and h to represent respectively states, local areas, and households groups, i to index individuals, and parenthesized superscripts to emphasize the multilevel structure, the following model is estimated:

¹⁵ Multilevel ordered Logit is theoretically feasible but proved to be too computationally intensive with such a large data set and a four-level structure.

¹⁶ The multilevel modeling literature is divided in terms of numbering levels from the top (as here) and from the bottom. Models that consider base level observations as level one would call this model a four-level hierarchical models rather than three-level.

$$SP_{slhi} = \alpha + X_s^{(1)}\lambda + X_{slh}^{(3)}\gamma + X_{slhi}\beta + u_s^{(1)} + u_{sl}^{(2)} + u_{slh}^{(3)} + \varepsilon_{slhi}, \quad (2)$$

where α is a common regression intercept, β , γ , and λ are vectors of coefficients and fixed effects to be estimated (after organizing the data into nested groups); X , $X^{(1)}$, and $X^{(3)}$ represent rows of the corresponding matrix, and the u 's are random intercepts at each level. The random effects are assumed to be normally distributed with mean zero and their own independent group specific variance to be estimated; ε_{slhi} are individual-level residuals with the classical OLS assumptions.¹⁷ The important distinction with a single-level model is that the variances of random intercepts at a given level are estimated so that they are shared by all individuals at that level. This means that coefficients in λ are estimated from the between state variance, not within, limiting the effective sample size to the number of states in each sample. The same is true for household-level coefficient estimates in γ that represent the between household variance while the variance of u_{slh} captures the within household variance.¹⁸ The hierarchical structure implies that individuals from different states cannot be in the same household nor local area and therefore do not share any common residual variance. Levels of the hierarchical model being identical to the different levels of the multistage sampling design of the data set, the estimation procedure also corrects for statistical issues created by the survey design itself. The estimation is performed using Maximum Likelihood.¹⁹

In the pooled regressions, state level effects are taken to be random and different for NFHS-2 and NFHS-3. The choice was made for two reasons. First B&Z found that coefficients and significance levels on state dummy variables were different between NFHS-2 and NFHS-3 without following any logical pattern; they concluded that the effects were likely measuring differences in data quality across states. Including a random effect at the state/survey level should account for differences in surveying procedures while unobserved regional characteristics can be captured by regional fixed effects. The second reason is that the state GDP per capita

¹⁷ The notation is adopted to facilitate the intuition behind the estimation procedure. For a clear exposition of hierarchical linear multilevel analysis and advantages of its use in policy analysis, see Leyland and Groenewegen (2003); for a more rigorous treatment see, for example, Rabe-Hesketh and Skrondal (2006).

¹⁸ The number of household is large and 75 to 80% of households contain a single observation. Although it complicated the procedure, the household level residual was kept because its variance was consistently found significantly different from zero and higher in magnitude than variances for upper levels.

¹⁹ Maximum Likelihood is easier to implement for unbalanced panels and allows comparisons between models with different fixed portion. Restricted Maximum Likelihood did not yield noticeable differences on coefficient estimates and standard errors.

figures do not always change the same way state by state from 1999 to 2005 because a few state delineations were changed in 2000 (Supplement 1, section 2).

Using the same mixed-effects linear estimation methods, several models are estimated, each one nested into the other in the sense that new variables are added to the $X^{(1)}$ matrix (household-level variables) while other variables are unchanged. All models are estimated for the NFHS-2, NFHS-3, and pooled samples. The base model, MW1, is the closest to B&Z (aside, notably, from the estimation procedure and the inclusion of the macroeconomic context). The model is first used as a benchmark to evaluate whether there is evidence of U-shaped relationships in aggregate and household wealth; it is then compared to models with variables related to dimensions of economic status. The Akaike Information Criteria (AIC), calculated as $-2 \ln(L) + 2k$ (where L is the value of the minimized likelihood function and k is the number of parameters) are used to compare the different models; the lower the AIC, the better the fit for a given sample.

RESULTS

Before choosing MW1 as a benchmark to evaluate models with relative wealth variables, three alternative models are estimated to test for the presence of U-shaped relationships for aggregate and individual wealth in NFHS-2, NFHS-3 and the pooled data. All models are identical to MW1 except that they include squared terms for state wealth (MW21), household wealth (MW12), or both (MW22). Coefficients on squared terms are found insignificantly different from zero in all models, with p values greater than .25.²⁰ MW1 has the lowest AIC value and MW22 the highest in all samples (the same rankings are obtained using Bayesian Information Criteria.) I conclude that, given the nature of the data examined (in particular the fact that I am using cross-sectional data), there is no evidence of U-shaped relationships between son preference and absolute wealth at the microeconomic or macroeconomic levels.

MWR1 adds land ownership (acres of cultivated land) as a separate variable, allowing different slopes whether the household is in rural or urban areas. I start with this specification because land ownership has been used in previous models of gender bias focused on rural areas (in addition to wealth and for reasons related to social status). MWR2 adds the new relative

²⁰ Detailed results are not reported for these models (other than MW1) since coefficients estimates on all variables other than wealth and squared wealth did not change.

wealth variable. Finally, MWR3 allows the relative wealth effect to vary whether the household is landed or not. All models include absolute wealth at the macroeconomic-level. Full estimation results are presented in Table 3 for NFHS-2 data and Table 4 for the pooled data. NFHS-3 results are not reported here to save space (the table is available online), but differences are highlighted below. For ease of presentation and interpretation, p values (instead of standard errors) are given below coefficients; the p values are omitted when below .001. Likelihood ratio tests indicate conservatively that the multilevel regression is appropriate (compared to a regression without group random effects). All estimated variance components for the different levels are virtually the same across model (but not across samples) and significantly different from zero at the strictest confidence level, although their respective contribution to the overall variance is very small compared to the individual specific residual variance. All demographic control variables and regional fixed effects are the same across models and consistent with B&Z's findings except for ideal number of children and its square (discussed in Supplement D4, section 2). The interpretation of the effect of ideal-total and its square when the dependant variable is a proportion of ideal family size is, however, very different and less obvious than in discrete probability models. Coefficient estimates on socioeconomic individual-level variables change slightly across models but signs and significance remain consistent with the literature. I concentrate therefore on commenting results related to the main goal of the paper: disentangling relative and absolute wealth effects.

First, looking at the statewide wealth level, I find that higher per capita state GDP reduces the prevalence of son preference at the strictest confidence level in the NFHS-2 and pooled samples. Using the 1998-1999 data and evaluating elasticities at mean values, I find roughly that a 1% increase in state GDP per capita (measured in thousand rupees) decreases SP by 0.25%. The relationship is less strong economically in the pooled data with elasticities of -0.14 to -0.15, but the result is still statistically significant at the strictest confidence level. There are two explanations for the lower magnitude of the pooled effect: first, the relationship is weaker in NFHS-3, with elasticities between -0.16 and -0.18, and p values between .06 and .09; second, the estimated coefficient in the pooled sample depends on both time and cross-sectional variance; variation between states in a given sample is expected to be larger than variation across the six years considered. Indeed, the cross-sectional effect should capture a long-run effect in the

sense that it measures the effect of deeper, more structural changes; effects measured using changes across short periods of time (here six years) are expected to be smaller.

The specific hypothesis of this analysis relates to the direction of the effect of the household's standard of living (*W99* and *W06*) and its relative economic standing within the community (*WR*). Results on absolute wealth in all models and across samples are consistent with economic theory: wealthier households are less likely to express preference for boys. Coefficients vary between -0.045 and -0.066 , translating into elasticities of -0.12 to -0.30 (using mean values). The largest elasticities are in models MWR2 and MWR3 with NFHS-3. Estimated coefficients on *W* (*W99* and *W06*) are stable across samples all with *p* values below .001. Controlling for local status increases the strength of the relationship, mostly due to the addition of the relative wealth variable. Including land ownership alone (size of cultivated land) increases the absolute value of the coefficient on *W* by less than 6% (the largest impact is in NFHS-2); this is very small compared to the increase in the magnitude of the effect after controlling for relative wealth in MWR2 (32 to 43%).

The most compelling result that is least intuitive to economists but finds much support in the sociology and anthropology literatures on gender bias in India is the role of variables related to economic status in the relative sense. A higher standing in the local community, whether it is related to higher relative wealth or land ownership, is associated with a higher preference for sons. Looking at land ownership alone does not suffice to capture the effect of relative economic status. Land ownership variables are significant and positive in the NFHS-2 and the pooled sample when looking at rural areas only; it becomes largely statistically insignificant both for urban and rural areas when using NFHS-3 data. When the effect is statistically significant, its magnitude is quite small with elasticities not exceeding 0.009, indicating that a large percentage difference in the size of land holdings is necessary to create a noticeable change in son-preferring attitudes. The relationship between son preference and relative economic status is, however, very robust when considering the principal components measure of relative wealth that defines relative economic status along a wide array of assets (including land). The coefficients on *WR* remain stable around 0.014 ± 0.002 in all samples, with *p* values lower than .001 in the pooled sample and NFHS-3, and below .02 for NFHS-2. The relative wealth effect is about four times weaker than the household absolute wealth effect with elasticities between 0.04 and 0.07, but

including *WR* in the regression significantly enhances the magnitude of the absolute wealth effect in terms of son preference reduction.

Although the size of land holdings does not have a clearly important role in son-prefering attitudes, the impact of relative wealth appear to be a feature of landed households: in MWR3, coefficients on relative wealth for non-landed households lose significance while the significance level is reinforced for landed households.

Finally, the pooled regression includes a fixed effect for the NFHS-3 survey. The effect is clearly showing lower levels of son preference overall in the 2005-2006 sample (independently of differences captured in the model).

Results on absolute and relative wealth effects are robust across all samples. An alternative specification of local areas using the NFHS-2 geographical identifiers is also tested (the full procedure is described in Supplement 3); the level of significance of the relative wealth variable is slightly reduced but remains significant at the same level; estimation results are virtually identical to those of Table 3. Comparisons of single level OLS result with Logit and ordered Logit estimations on dependant variables defined as in B&Z and P&A and taking account of the complex survey design also validate the qualitative linear results. Table 5 present results of the three alternative estimations for the pooled sample. Changes in the absolute wealth impact with or without relative wealth were also tested with the Logit model both the single level model and a random-effect Logit model with State fixed effects. Adding the size of cultivated land decreases the odd ratio of absolute wealth by less than 5% while adding *WR* decreases it by about 25%. Details of the estimations and additional results are given in Supplement 4.

As mentioned in the introduction, although variation in son preference should be a good approximation of overall changes in attitudes toward women, the dependant variable does not measure gender bias in terms of girls' access to resources. In addition, explanatory variables on family demographics could be endogenous in a world where sex selection is increasingly practiced. In order to test the absolute/relative wealth dichotomy further, I use an alternative dependant variable (*EduBias*) based on questions on desired level of education for boys and girls. Details on the variable, summary statistics and complete results are given in Supplement 4, section 5. Although it is still a stated preference variable, it is expected to be positively correlated with son preference and gender bias in general, while mostly free of possible endogeneity

problems. Unfortunately, the information was no longer available in NFHS-3. The same empirical analysis is run with *EduBias* as the dependant variable, removing demographic controls specifically related to son preference. To facilitate comparison with the son-preference model, Table 6 reports results on wealth-related variables in terms of elasticities. The effect of absolute wealth, both at the macroeconomic and household-levels are significant and negative, with magnitudes 50% higher than for son preference. Inclusion of the relative wealth variable into the estimation also increases the magnitude of the absolute wealth impact, although by a lower percentage than in the son preference model (13%). One important difference between the two models concerns the role of land ownership. Size of land holdings, outside of its contribution to *W* and *WR*, is found to have no impact on stated educational bias. The link between relative wealth and gender bias is still positive and significant but no more a feature of landed households only. The result is not surprising if one believes that issues of marriage are those distinguishing the behavior of landed household and if education of a girl facilitates her placement in a good family.

CONCLUSION

Contradictory predictions from different disciplines about son preference in India as households become richer can be reconciled if we accept that they rely on different notions of wealth. Empirical evidence demonstrates that the link between son preference and absolute wealth is clearly negative and significantly stronger when one controls for the effect of relative wealth. The observed negative influence of “prosperity” on gender equality appears to work through notions of local economic status, i.e. it is the higher relative wealth position of a household in the local community that generates gender bias.

Including both absolute and relative wealth in the empirical analysis is not only necessary to avoid biased results, it also allows us to think differently about the types of development strategies that are most likely to mitigate son preference and gender bias in general. In addition to the general goal of increasing standards of living and specific targeting of women in education and health, development strategies that decrease the importance of local status are expected to be instrumental in reducing son preference. Such policies could include, for example, the development of transportation infrastructure and the promotion of their use, de-localization of information (for example with the use of modern communication technology) and access to

education. Their potential effect on labor mobility and geographical reach of marriage searches is most likely to minimize the role of local status-seeking behavior and to reinforce the positive effect of increased standards of living in son preference reduction.

In order to better understand the links between son preference and economic development it is important to estimate how much stated preference for sons effectively translates into de-facto daughter neglect. Tarozzi and Mahajan (2007) showed that nutritional status improved more for boys than girls in India and that the trend was mostly marked in son-preferring rural areas of Northern states. Further research is needed to understand the strength of the relationship and whether it generalizes to other domains. Indeed, the channels linking gender bias to slower economic development are all related to the treatment of “surviving” girls -i.e. the likelihood of girls to be given appropriate health care and access to education rather than son-preference or unbalanced sex ratios.

REFERENCES

- Agnihotri, S.B. 2000. *Sex Ratio Patterns in the Indian Population: A Fresh Exploration*. New Delhi: Sage.
- _____ 2003. "Survival of the Girl Child: Tunneling out of the Chakravayuha." *Economic and Political Weekly* October 11: 4351-560.
- Allison, P.D. 1999. "Comparing Logit and Probit Coefficients Across Groups." *Sociological Methods and Research* 28 (2): 186-208.
- Anderson, S. 2003. "Why Dowries Payments Declined with Modernization in Europe but Are Rising in India." *Journal of Political Economy* 111(2): 269-310.
- Arnold, F., M. Kim Choe, and T.K. Roy. 1998. "Son preference, the family-building process and child mortality in India." *Population Studies* 523: 301-15.
- Basu, A.M. 1999. "Fertility Decline and Increasing Gender Imbalance in India, Including a Possible South-Indian Turnaround." *Development and Change* 302: 237-63.
- _____ 2000. "Fertility Decline and Worsening Gender Bias in India: A response to S. Irudaya Rajan *et al.*" *Development and Change* 31: 1093-95.
- Behrman, J., A. Foster, M. Rosenzweig, and P. Vashishtha. 1999. "Women's Schooling, Home Teaching, and Economic Growth." *Journal of Political Economy* 1074: 682-714.
- Berremán, G. 1993. "Sanskritization as female oppression in India." In *Sex and gender hierarchies*, edited by Barbara Diane Miller. Cambridge: Cambridge University Press.
- Bhat, P.N.M. and F. Zavier. 2003 "Fertility Decline and Gender Bias in Northern India." *Demography* 404: 637-57.
- Billig, M.S. 1992. "The Marriage Squeeze and the Rise of Groomprice in India's Kerala State." *Journal of Comparative Family Studies* 232:197-216.
- Blaffer Hrdy, S. 1999. *Mother Nature: a History of Mothers, Infants, and Natural Selection*. New York: Pantheon Books (Random House).
- Bloch, F., V. Rao, and S. Desai. 2004. "Wedding Celebrations as Conspicuous Consumption: Signaling Social Status in Rural India." *Journal of Human Resources* 39(3): 675-95.
- Chakraborty, T. and S. Kim. 2008. "Caste, Kinship, and Sex Ratios in India." *NBER Working Paper* no. 13828. Available online at <http://www.nber.org/papers/w13828>.
- Das Gupta, M. 2005. "Explaining Asia's 'Missing Women': A New Look at the Data." *Population and Development Review* 31(3): 529-35.
- Deaton, A. 2003. "Health, Inequality, and Economic Development." *Journal of Economic Literature* 41: 113-158.
- Dyson, T. and M. Moore. 1983. "On kinship structure, female autonomy and demographic behavior in India." *Population and Development Review* 91: 35-60.
- Frank, R.A. 1985. *Choosing the Right Pond*. New York: Oxford University Press
- Galor, O. and D.N. Weil. 1996. "The Gender Gap, Fertility, and Growth." *American Economic Review* 86(3): 374-87.

- Gaudin, S. and A. Yazbeck. 2006a. "Immunization in India 1993-1999: Wealth, Gender, and Regional Inequalities Revisited." *Social Science and Medicine*, 623: 694-706.
-
- 2006b "Immunization in India: An Equity-Adjusted Assessment" *Health Nutrition and Population Publication Series* March 2006, World Bank, Washington D.C.
- Goldin, C. 1995. "The U-shaped Female Labor Force Function in Economic Development and Economic History." In *Investment in Women's Human Capital*. Edited by T.P. Schultz, Chicago: University of Chicago Press.
- Gwatkin, D., S. Rutstein, K. Johnson, R. Pande, and A. Wagstaff. 2000. "Socio-economic differences in health, nutrition and population in [44 countries]." *Health, Nutrition and Population Publication Series*, HNP/Poverty Thematic Group, World Bank, Washington D.C.
- International Institute for Population Sciences (IIPS) and Orc Macro. 2000. *National Family Health Survey (NFHS-2) 1998-99: India*. Bombay: IIPS.
- IIPS and Macro International. 2007. *National Family Health Survey (NFHS-3) 2005-06: India: Volumes I and II*. Mumbai: IIPS.
- Jaffe, D.H., Z. Einsenbach, Y.D. Neumark, and O. Manor. 2004. "Individual, household and neighborhood socioeconomic status and mortality: a study of Absolute and Relative Deprivation." *Social Science and Medicine* 605: 989-97.
- Kanbur, R. and L. Haddad. 1994. "Are Better Off Households More Unequal or Less Unequal?" *Oxford Economic Papers* 463: 445-58
- Kingdon, G.G.. 2005. "Where Has All the Bias Gone? Detecting Gender Bias in the Intrahousehold Allocation of Educational Expenditure." *Economic Development and Cultural Change* 532: 409-51
- Klasen, S. 1999. "Does Gender Inequality Reduce Growth and Development? Evidence from Cross-Country Regressions." *Policy Research Reports on Gender and Development Working Paper Series*. World Bank, Washington D.C.
- Leyland A.H. and P. Groenewegen. 2003. "Multilevel Modeling and Public Health Policy." *Scandinavian Journal of Public Health* 31: 267-74.
- Montgomery, M.R., M. Gragnolati, K.A. Burke and E. Paredes. 2000. "Measuring Living Standards with Proxy variables." *Demography* 37(2):155-174.
- Murthi, M., A-C. Guio and J. Dreze. 1995. "Mortality, Fertility and Gender Bias in India: A district level analysis." *Population and Development Review* 21: 742-82.
- Pande, R. 2003. "Selective Gender Differences in Childhood Nutrition and Immunization in Rural India: The Role of Siblings." *Demography* 40(3): 395-418.
- Pande R. and N.M. Astone 2007. "Explaining Son Preference in Rural India: the Independent role of structural versus individual factors." *Population Research and Policy Review* 26(1): 1-29.

- Pande, R. and A. Malhotra. 2006. "Son Preference and Daughter Neglect in India: What Happens to Living Girls?" International Center for Research on Women publication, ICRW, Washington D.C.
- Pande, R. and A. Yazbeck. 2003. "What's in a Country Average? Wealth, Gender, and Regional Inequalities in Immunization in India." *Social Science & Medicine* 57(11): 2075-88.
- Premi, M.K. 2001. "The Missing Girl Child" *Economic and Political Weekly* 3621 (May 26): 1875-80.
- Rabe-Hesketh S. and A. Skrondal 2006. *Multilevel and Longitudinal Modeling Using Stata*. College Station, TX: Stata Press.
- Rajan, S.I., S. Shudha, and P. Mohanachandran. 2000. "Fertility Decline and Worsening Gender Bias in India: Is Kerala no longer an exception?" *Development and Change* 31: 1085-92.
- Rosenzweig, M.R. and T.P. Shultz. 1982. "Market Opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in Rural India." *American Economic Review* 72(4): 803-15.
- Scanlan, S. J. 2004. "Women, Food Security, and Development in Less-Industrialized Societies: Contributions and Challenges for the New Century." *World Development* 32(11): 1807-29.
- Sen, A.K. 1984. *Resources, Values, and Development*. Cambridge MA: Harvard University Press.
- _____ 2003. "Missing Women Revisited." *British Medical Journal* 327(7427): 1297-98.
- Srinivas, M. N. 1962. *Caste in India and Other Essays*. Bombay: Asia Publishing House.
- _____ 1989. *The Cohesive Role of Sanskritization and Other Essays*. Delhi: Oxford University Press.
- Staal, J.F. 1963. "Sanskrit and Sanskritization." *The Journal of Asian Studies*, 22(3): 261-75.
- Sweetman, C., ed. 2002. *Gender, development, and poverty*. (Oxfam focus on Gender series) Oxford: Oxfam (distributed by Stylus, Herndon, Va.).
- Tarozzi A. and A. Mahajan. 2007 "Child Nutrition in India in the Nineties." *Economic Development and Cultural Change* 55(3): 441-86.
- World Bank. 2001. "Engendering development: Through gender equality in rights, resources, and voice." *World Bank Policy Research Report series*. New York and Oxford, Washington D.C.: World Bank and Oxford University Press.
- Zhang, J., J. Zhang, and T. Li. 1999. "Gender Bias and Economic Development in an Endogenous Growth Model." *Journal of Development Economics* 59(2): 497-525.

Table 1. Explanatory Variables: Description and Summary Statistics (Final Samples)

Variable	Description	Mean (St. Dev.) ^a	
		NFHS-2	NFHS-3
State Level:			
GSP/c	State level real gross domestic product per capita in constant Rps (thousands) of 2001	19 (8.3)	39 (19)
Household Level - Absolute and Relative Wealth:			
<i>W</i> (<i>W99</i> and <i>W06</i>)	Normalized p.c. wealth scores based on all India (range between 0 and 1).	0.38 (0.23)	0.39 (0.24)
<i>WR</i>	Normalized p.c. wealth scores based on own PSU (between 0 and 1 in <i>each</i> PSU)	0.45 (0.31)	0.46 (0.31)
<i>WR</i> ×Land	Mean <i>WR</i> conditional on owning land	0.46 (0.30)	0.46 (0.31)
Land	0/1 Household owns agricultural land	0.51	0.42
Land-acres x urban	Acres for households in urban areas only	1.2 (9.2)	0.9 (7.5)
Land-acres x rural	Acres for households in rural areas only	3.5(1.14)	2.2 (7.0)
Individual Level: Socio-Economic/Cultural:			
Illiterate	0/1 Respondent is illiterate (see footnote xx)	0.54	0.43
Education-self	Education of respondent in completed years	3.9 (4.7)	5.2 (5.2)
Edu-partner	Education of partner in completed years	6.5 (5.1)	7.3 (5.2)
Paid Work	0/1 Respondent works and is paid in cash	0.24	0.28
Other Work	0/1 Respondent works but is not paid in cash	0.13	0.078
Media Exposure	0/1 Respondent Watches TV, listens to radio, or read newspaper at least once a week	0.63	0.70
Religion	Reference: Hindu	0.78	0.75
	0/1 Muslim	0.11	0.13
	0/1 Sikh	0.025	0.022
	0/1 Christian	0.057	0.072
	0/1 Other	0.026	0.030
Scheduled Caste	Belongs to a historically disadvantaged caste	0.17	0.17
Scheduled Tribe	Belongs to a historically disadvantaged tribe	0.12	0.12
Individual Level: Demographic			
Sons	Number of sons alive	1.4 (1.1)	1.3 (1.1)
Daughters	Number of daughters alive	1.3 (1.2)	1.2 (1.2)
Sons Dead	Number of sons dead	0.19 (0.51)	0.14 (0.45)
Daughters Dead	Number of daughters dead	0.17 (0.50)	0.13 (0.42)
Age	Age of respondent in years	32 (8.6)	33 (8.4)
Ideal-total	Ideal total number of children	2.7 (1.1)	2.4 (0.94)
Odd ideal	0/1 Ideal total number of children is odd	0.35	0.30
Fixed Group Effects:			
Regions	Census definitions. Ref.: East	0.31	0.31
	0/1 North	0.33	0.28
	0/1 Central and West	0.19	0.22
	0/1 South	0.17	0.19
Urban	0/1 place of residence is urban	0.31	0.45
NFHS-3	0/1 data from NFHS-3	0	1

^a Means and standard deviations are calculated based on samples used in the regressions and without sample weights. Standard deviations are omitted for dummy variables (0/1).

Table 2. Primary Sampling Units (PSU) by Size

PSU	NFHS-2	NFHS-3
All	3,215	3,850
≥50 households	112	175
40-49	232	239
35-39	312	312
30-34	637	658
25-29	868	1,083
20-24	650	857
15-19	316	398
10-14	85	126
<10 households	3	2
Mean size	31.4	31.4
Median size	30	29

Table 3. Multilevel Linear Estimation of Son preference Models: NFHS-2 sample

Variable	Model			
	MW1	MWR1	MWR2	MWR3
State level:				
GSP/c	-0.0017*** (0.002)	-0.0017*** (0.002)	-0.0017*** (0.003)	-0.0016*** (0.004)
Household level:				
<i>W</i>	-0.045***	-0.048***	-0.064***	-0.059***
<i>WR</i>			0.012** (0.018)	0.0039 (0.531)
<i>WR</i> ×Land				0.011** (0.011)
Land-acres × urban x 100		0.025 (0.158)	0.024 (0.170)	0.017 (0.343)
Land-acres× rural x 100		0.032*** (0.001)	0.030*** (0.003)	0.026** (0.011)
Individual level				
Illiterate	0.0057 (0.100)	0.0058* (0.089)	0.0062* (0.074)	0.0064* (0.064)
Education-self	-0.0023***	-0.0023***	-0.0022***	-0.0022***
Edu-partner	-0.0004 (0.117)	-0.00042* (0.096)	-0.00046* (0.073)	-0.0005** (0.050)
Paid Work	-0.0062*** (0.007)	-0.0059** (0.011)	-0.0057** (0.015)	-0.0055** (0.017)
Other Work	0.0063** (0.037)	0.0058* (0.057)	0.0055* (0.072)	0.0049 (0.104)
Media Exposure	-0.0061** (0.011)	-0.0061** (0.011)	-0.0063*** (0.009)	-0.0065*** (0.007)
Religion: Ref. Hindu				
Muslim	-0.0042 (0.223)	-0.0039 (0.265)	-0.0038 (0.276)	-0.0033 (0.334)
Sikh	0.033***	0.033***	0.033***	0.032***
Christian	-0.026***	-0.026***	-0.025***	-0.026***
Other	-0.0012 (0.853)	-0.0012 (0.859)	-0.00088 (0.892)	-0.00078 (0.905)
Scheduled Caste	-0.00012 (0.962)	0.00028 (0.915)	0.00058 (0.825)	0.001 (0.689)
Scheduled Tribe	-0.011*** (0.002)	-0.011*** (0.002)	-0.012*** (0.001)	-0.011*** (0.002)
Age (respondent)	-0.00015 (0.247)	-0.00015 (0.270)	-0.00014 (0.291)	-0.00014 (0.301)
Sons	0.025***	0.025***	0.025***	0.025***
Daughters	-0.015***	-0.015***	-0.015***	-0.015***
Sons-dead	0.0092***	0.0091***	0.0091***	0.0091***
Daughters-dead	-0.0036* (0.059)	-0.0035* (0.060)	-0.0035* (0.060)	-0.0036* (0.059)

Ideal-total	-0.025***	-0.025***	-0.025***	-0.025***
(Ideal-total) ²	0.0022***	0.0022***	0.0022***	0.0022***
Odd ideal	0.18***	0.18***	0.18***	0.18***
Fixed Effects				
Region (East reference)				
North	0.052***	0.052***	0.052***	0.052***
Central & West	0.033**	0.033**	0.033**	0.033**
	(0.025)	(0.028)	(0.027)	(0.027)
South	-0.034**	-0.034**	-0.034**	-0.034**
	(0.012)	(0.012)	(0.012)	(0.012)
Urban Residence	-0.015***	-0.014***	-0.012***	-0.0099***
Constant	0.16***	0.16***	0.16***	0.15***
Random Components (standard deviations)				
Level 1: State	0.022***	0.022***	0.022***	0.022***
Level 2: PSU	0.037***	0.037***	0.037***	0.037***
Level 3: Household	0.055***	0.055***	0.055***	0.055***
Residual error	0.25***	0.25***	0.25***	0.25***
Regression Statistics				
Akaike Information Criterion	1,439	1,431	1,427	1,423
Nested groups (unbalanced)				
States	26	26	26	26
Local areas (PSUs)	3,127	3,127	3,127	3,127
Households	68,114	68,114	68,114	68,114
N	77886	77886	77886	77886

note: p values in parentheses (p>|z|), omitted when p<0.001.

*p<.10 ** p<.05 ***p<.01

Table 4. Multilevel Linear Estimation of Son preference Models: Pooled Sample

Variable	Model			
	MW1	MWR1	MWR2	MWR3
State level:				
GSP/c	-0.00055*** (0.009)	-0.00055** (0.010)	-0.00052** (0.013)	-0.00051** (0.015)
Household level:				
W99	-0.047***	-0.049***	-0.066***	-0.062***
W06	-0.048***	-0.049***	-0.066***	-0.062***
WR			0.013***	0.0063 (0.134)
WR×Land				0.011*** (0.004)
Land-acres × urban x 100		0.0036 (0.770)	0.0027 (0.826)	-0.0037 (0.766)
Land-acres × rural x100		0.026*** (0.002)	0.023*** (0.007)	0.019*** (0.032)
Individual level				
Illiterate	0.0046** (0.045)	0.0047** (0.040)	0.0050** (0.031)	0.0051** (0.026)
Education-self	-0.0021***	-0.0020***	-0.0020***	-0.0020***
Edu-partner	-0.00036** (0.044)	-0.00037** (0.036)	-0.00040** (0.023)	-0.00044** (0.013)
Paid Work	-0.0059***	-0.0057***	-0.0055***	-0.0054*** (0.001)
Other Work	0.0081***	0.0077*** (0.001)	0.0073*** (0.002)	0.0069*** (0.003)
Media Exposure	-0.0065***	-0.0065***	-0.0066***	-0.0068***
Religion: Ref. Hindu				
Muslim	-0.004* (0.086)	-0.0038 (0.101)	-0.0038 (0.110)	-0.0034 (0.147)
Sikh	0.023***	0.022***	0.022***	0.022***
Christian	-0.028***	-0.028***	-0.028***	-0.028***
Other	-0.0023 (0.592)	-0.0022 (0.599)	-0.0019 (0.649)	-0.0018 (0.668)
Scheduled Caste	0.0011 (0.536)	0.0014 (0.452)	0.0016 (0.374)	0.002 (0.277)
Scheduled Tribe	-0.010***	-0.010***	-0.010***	-0.010***
Age (respondent)	-0.000032 (0.732)	-0.000028 (0.764)	-0.000022 (0.811)	-0.00002 (0.830)
Sons	0.025***	0.025***	0.025***	0.025***
Daughters	-0.015***	-0.015***	-0.015***	-0.015***
Sons-dead	0.0078***	0.0077***	0.0077***	0.0077***
Daughters-dead	-0.0031** (0.031)	-0.0030** (0.031)	-0.0031** (0.031)	-0.0031** (0.031)
Ideal-total	-0.011***	-0.010***	-0.011***	-0.011***
(Ideal-total) ²	0.00055* (0.054)	0.00054* (0.057)	0.00055* (0.054)	0.00055* (0.051)

Odd ideal	0.18***	0.18***	0.18***	0.18***
Fixed Effects				
Region (East reference)				
North	0.034***	0.034***	0.034***	0.034***
West	0.031**	0.031**	0.031**	0.031**
	(0.012)	(0.013)	(0.013)	(0.012)
South	-0.038***	-0.038***	-0.038***	-0.038***
Urban Residence	-0.013***	-0.012***	-0.011***	-0.0091***
NFHS-3 sample	-0.016**	-0.016**	-0.017**	-0.017**
	(0.043)	(0.040)	(0.029)	(0.026)
Constant	0.12	0.12	0.12	0.12
Random Components (standard deviations)				
Level 1: State	0.022***	0.022***	0.022***	0.022***
Level 2: PSU	0.037***	0.037***	0.037***	0.037***
Level 3: Household	0.043***	0.043***	0.043***	0.043***
Residual error	0.023***	0.023***	0.023***	0.023***
Regression Statistics				
Akaike Information Criterion	4,417	4,412	4,400	4,393
Nested groups (unbalanced)				
States	55	55	55	55
Local areas (PSUs)	6,849	6,849	6,849	6,849
Households	143,457	143,457	143,457	143,457
N	161,671	161,671	161,671	161,671

note: *p* values in parentheses ($p > |z|$), omitted when $p < 0.001$.

* $p < .10$ ** $p < .05$ *** $p < .01$

Table 5. Comparison of OLS, Logit, and Ordered Logit Specifications (Pooled Sample)

Variable:	Estimation method ^a		
	OLS	Logit ^b	Ordered Logit ^b
GSP/c	-0.00038***	0.99***	0.99***
W99	-0.041***	0.57***	0.69***
W06	-0.027***	0.63***	0.72***
	(0.004)	(0.001)	(0.001)
WR (PSU)	0.0011	1.04	0.998
	(0.839)	(0.635)	(0.979)
WR×Land	0.014***	1.23***	1.19***
Land-acres × urban x 100	-0.00086	0.92	0.97
	(0.943)	(0.670)	(0.843)
Land-acres × rural x 100	0.041***	1.60***	1.33***
		(0.006)	(0.004)
Illiterate	0.0034	0.99	1.00
	(0.257)	(0.848)	(0.722)
Education-self	-0.0030***	0.95***	0.97***
Edu-partner	-0.00040*	1.00	1.00
	(0.059)	(0.323)	(0.104)
Paid Work	-0.0042**	0.99	1.002
	(0.038)	(0.830)	(0.943)
Other Work	0.013***	1.15***	1.10***
Media Exposure	-0.011***	0.88***	0.91***
Religion: Ref. Hindu			
Muslim	-0.0128***	0.7413***	0.85***
Sikh	0.030***	1.36***	1.29***
Christian	-0.017***	0.63***	0.70***
Other	-0.017***	0.87	0.87**
	(0.006)	(0.140)	(0.037)
Scheduled Caste	0.00202	1.03	1.02
	(0.348)	(0.356)	(0.276)
Scheduled Tribe	-0.013***	0.79***	0.86***
Age (respondent)	-0.00025**	0.992***	0.996***
	(0.027)		
Sons	0.024***	1.36***	1.25***
Daughters	-0.014***	0.84***	0.87***
Sons-dead	0.0074***	1.086***	1.06***
Daughters-dead	-0.00090	0.999	0.99
	(0.585)	(0.971)	(0.641)
Ideal-total	-0.0066	3.21***	2.29***
	(0.170)		
(Ideal-total) ²	0.00069	0.90***	0.95***
	(0.261)		
Odd ideal	0.19***	39.8***	16.9***
Region (Ref. East)			

North	0.021***	1.35***	1.22***
West	0.018***	1.36***	1.25***
South	-0.045***	0.39***	0.50***
Urban Residence	-0.0064**	0.91**	0.94**
	(0.013)	(0.012)	(0.037)
Time (sample): NFHS-3	-0.024***	0.75***	0.81***
Constant	0.12***	0.019***	
Constant 0-1			34.6***
Constant 1-2			550***
Regression Statistics:			
N	161,671	161,671	161,671
F	933	686	618
R-Squared	0.17	n.a	n.a.

Note: p values ($p > |z|$) are in parentheses below the coefficient estimate, they are omitted when $p < .001$.

^a All standard errors corrected for group heteroskedasticity caused by the NFHS complex survey design. Strata are rural/urban areas of each state in each NFHS-sample. Household effects are ignored.

^b Odd ratios are reported; numbers <1 indicate negative relationships.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 6. Son Preference Versus. Educational Bias: Wealth Elasticities (NFHS-2)

Variable:	<i>SP Models</i>			<i>EduBias Models</i>		
	MW1	MWR2	MWR3	MW1	MWR2	MWR3
GSP/c	-0.25*** (0.002)	-0.24*** (0.003)	-0.24*** (0.004)	-0.38*** (0.012)	-0.37*** (0.013)	-0.38*** (0.011)
W	-0.13***	-0.18***	-0.17***	-0.45***	-0.49***	-0.51***
WR		0.04*** (0.018)	0.01 (0.531)		0.04* (0.063)	0.07*** (0.002)
WR×Land			0.05** (0.011)			0.009*** (0.004)

note: Elasticities are calculated at sample mean values. *p* values are in parentheses ($p > |z|$), *p* values are omitted when $p < 0.001$.

* $p < .10$ ** $p < .05$ *** $p < .01$

Figure 1. Distribution of the Son Preference Variable, $SP = \text{Ideal}\{(\text{boys}-\text{girls})/\text{total}\}$

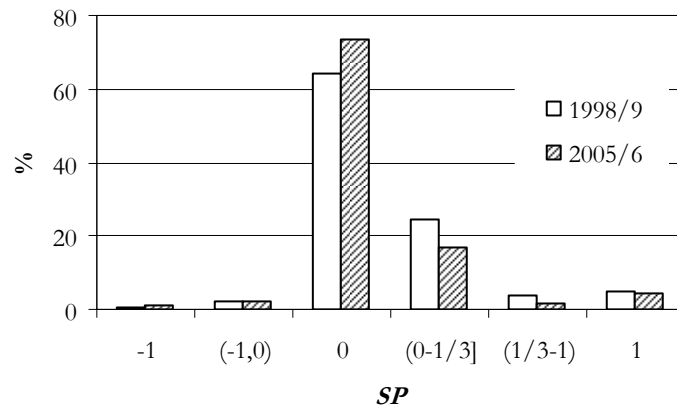


Figure 2. Means and St. Dev. of Normalized PSU-level PC Scores by Sample Size

