Wealth, Marriage, and Sex Selection*

Girija Borker† Jan Eeckhout‡ Nancy Luke§ Shantidani Minz¶ Kaivan Munshi‖
Soumya Swaminathan**

January 20, 2022

Abstract

We collect unique data on permanent income, marriage and child sex ratios in a rural population of 1.1 million individuals in South India. We document two novel and robust facts: (i) relatively wealthy households in the marriage market, defined by the caste in India, engage more in sex selection; and (ii) hypergamy – the wealth-gap between grooms and brides – is increasing as we move down the wealth distribution. We show that these inter-linked facts emerge from a model in which marital sorting, sex selection and dowries are jointly determined. The model provides novel micro-foundations for sex selection, based on the marriage institution in India, that complement and can be disentangled from intrinsic son preference. We evaluate counterfactual policies with the estimated model and find that some existing policies inadvertently increase sex selection, while also identifying other policies that could be especially effective.

Keywords. Family Economics. Social Norms. Marriage Market. Sex Selection. Caste. Assortative Matching. Wealth Distribution. Inequality. Control Function.
JEL. J12. J16. D31. I3.

*Akhil Lohia and Johannes Maywald provided exceptional research assistance and Swapnil Singh graciously helped us with the cross validation. We are grateful to Alexei Onatskiy, Debraj Ray, and numerous seminar participants for their constructive comments. Research support from the National Institutes of Health through grant R01-HD046940, the Keynes Fund at the University of Cambridge, Cambridge-INET, the Agence Nationale de la Recherche (ANR) under the EUR Project ANR-17-EURE-0010, ERC Grant 339186, and grant ECO2015-67655-P is gratefully acknowledged. We are responsible for any errors that may remain.

†World Bank
‡UPF Barcelona (ICREA-BSE-CREi)
§Pennsylvania State University
¶Christian Medical College, Vellore
‖Yale University
**World Health Organization

1 Introduction

Sex selection through female feticide, infanticide, or neglect is a serious problem in many parts of the world. Amartya Sen brought sex selection to public attention over 30 years ago when he famously claimed that 100 million women were "missing" in Asia (Sen, 1990). Since that time, India has made tremendous economic progress. Economic development has been associated with greater gender equality on many dimensions (e.g. Geddes and Lueck (2002), Doepke and Tertilt (2009)). However, on the sex selection dimension, the problem has persisted. Child (aged 0-6) sex ratios in India were 109 boys per 100 girls in the most recent 2011 population census, which implies, based on the natural benchmark of 102 set in our analysis, that 2.4 million girls were missing in that age cohort.[1]

Our research provides new empirical and theoretical insights into this longstanding problem. Using unique data we have collected from rural South India, we document two novel and robust facts: (i) relatively wealthy households within each marriage market, defined by the caste in India, engage more in sex selection; and (ii) hypergamy – the wealth-gap between grooms and brides – is increasing as we move down the wealth distribution. We show that these two inter-linked facts emerge from a theoretical model in which marital sorting, sex selection and dowries are jointly determined.

Our analysis also sheds light on the underlying motivations for sex selection. Two mechanisms have been proposed to explain sex selection in India: *son preference* in which parents desire a male heir, either to provide old-age support or to inherit their wealth, and *daughter aversion*, in which marriage payments or dowries make parents worse off with a girl than with a boy (Guilmoto, 2009). The son preference mechanism has been examined theoretically (Edlund, 1999; Bhaskar, 2011) and has also received empirical support (Bhalotra et al., 2019). In contrast, although there is some recent evidence that exogenous increases in dowries worsen the sex ratio (Alfano, 2017; Bhalotra et al., 2020), the theoretical foundations underlying the daughter aversion mechanism are less well understood. In particular, there is no obvious reason why girls' families, who are on the buyer's side of the market when marriage payments flow to the boy, should be necessarily disadvantaged in equilibrium. The model that we develop in this paper provides novel micro-foundations for daughter aversion, or what we call the marriage market mechanism, based on the organization of the marriage institution in India.

In our model, parents are endowed with a particular level of wealth that is available for consumption. Parents are altruistic and so if their children were single, they would distribute their wealth so that they and their children consume the same amount. However, this is not the allocation of resources that emerges in equilibrium. Marriage in India is patrilocal, which means that the girl leaves her natal home when she marries. Her altruistic parents must thus use the dowry to share wealth with their daughter. The dowry is given to the in-laws and is combined with their own wealth before being subsequently distributed. We will see that the inability of the girl's parents to make the *premortem* bequest directly can under specific conditions

[1] The natural sex ratio at birth favors boys and is typically assumed to be 105 (Guilmoto, 2009). Subsequent mortality favors girls, but there is no similar widely agreed upon benchmark for the 0-6 age group, which will be the focus of our analysis. As discussed below, widespread sex selection only commenced in the 1980s in South India, which is the setting for our research. Drawing on population censuses prior to that time, we set the benchmark natural sex ratio for children aged 0-6 at 102 boys per 100 girls.

(discussed below) leave parents worse off with a girl than with a boy, which, in turn, results in sex selection.

We solve the model, incorporating intrinsic son preference and the marriage market mechanism described above, in three steps. In the first step, we establish that there is positive assortative matching on wealth, which is a well documented feature of the marriage market in India; e.g. Bloch and Rao (2002). The marriage market is precisely defined by the endogamous caste or *jati*, with many castes coexisting in a given area, and this feature of Indian society will play an important role in the empirical analysis that follows. Given that parents are altruistic, all girls' parents want their daughters to match with wealthy boys, whose families have greater resources, thus allowing the daughter-in-law to consume at a higher level. The dowry is a price that equilibrates the demand and supply of marriage partners. It must be increasing steeply enough in wealth to ensure that less wealthy girls' parents are not willing to deviate from the assortative equilibrium and pay the price that is needed to match up in wealth. The dowry has been characterized in the literature as a price in the marriage market (Becker, 1973) or as a bequest from the girl's parent to her (Botticini, 1999). A distinguishing feature of the marriage transfer in our model is that it serves both purposes, allowing girls' parents to make (indirect) bequests, while simultaneously clearing the marriage market (see also Anderson and Bidner (2015)).

The second step in solving the model is to establish that there is sex selection at every wealth level within the marriage market (caste). In general, sex selection arises when the utility from having a boy exceeds that from having a girl. Assuming that there is heterogeneity in the pecuniary and non-pecuniary cost of sex selection, parents below a cutoff cost will then choose to select the sex of their child. With the son preference mechanism, the utility surplus from having a boy is generated exogenously. With the daughter aversion or marriage market mechanism, three conditions must be satisfied for this surplus to be endogenously generated: (i) There must be an absence of commitment on the boy's side. If the girl's parent could transfer the bequest to her directly or if the boy's side could commit to making the optimal transfer *ex post*, then sex selection would not arise. (ii) The girl's bargaining position in her marital home must be relatively weak, which results in her receiving a less than equal share of the available resources. If those resources were divided equally across all members of the household, then there would be no sex selection even in the absence of commitment. (iii) The social norm in India that all girls must marry must be binding. If that were not the case, a girl's parent could avoid the disutility associated with the marriage market mechanism by leaving her single.[2]

Once there is sex selection, the wealth distribution on the two sides of the marriage market (which determines the pattern of matching) becomes endogenous. As in the family economics literature; e.g. Greenwood, Guner, and Vandenbroucke (2017), we must account for the two-way interaction between the family's sex selection decision and the marriage market equilibrium; i.e. sex selection, the wealth distribution, and the dowry must be solved simultaneously. The expression for the dowry, in addition, holds a fixed point. This is a challenging problem, which has not been previously solved in the matching literature. We are nevertheless able to show analytically that sex selection exists, under reasonable parameter restrictions, at the top of the wealth distribution (where the wealthiest boys and girls match with each other) and at the bottom of

[2]Although social norms have historically received less attention in the micro-development literature, there have recently been some attempts to understand how such norms can affect the functioning of the marriage institution (Ashraf et al., 2020; Corno et al., 2020).

the wealth distribution (where the dowry of the last boy to match is pinned down by his outside option of remaining single). In addition, we solve the model numerically and observe that sex selection exists at each point in the wealth distribution.

The third and final step in solving the model is to establish that sex selection is declining as we move down the wealth distribution within a marriage market (caste). The intuition for this result is provided by the following example. Suppose that there are 100 wealth levels and two boys and one girl at each wealth level. Except for the number of boys and girls, the wealth distribution is the same and uniform on [1,100]. Then under positive assortative matching, one of the boys with wealth 100 marries the girl with wealth 100 and the other boy marries the girl with wealth 99, the boys with wealth 99 marry the girls with wealth 98 and 97, and so on, until the last boy to be matched, with wealth 50, marries the girl with wealth 1. The key insight is that the wealth-gap or hypergamy is increasing as we move down the wealth distribution: at the top (100,100) there is no wealth gap and at the bottom (50,1) the wealth gap is 49. This is obviously not an equilibrium. Poorer parents are less disadvantaged by having a girl and, thus, they will have less incentive to select the gender of their child. As a result, the sex ratio will adjust and be less biased as we move down the wealth distribution (although the wealth-gap will continue to increase in equilibrium). Note that this decline in sex selection is not driven by an associated decline in equilibrium dowries. Although we assume that dowries are always positive, in line with the evidence that dowries are universal in India (Chiplunkar and Weaver, 2021), the model does not generate implications for the relationship between dowries and relative wealth. The reason why sex selection is declining as we move down the wealth distribution is because girls are marrying higher up *and* because they are in increasingly short supply (which allows the girl's side to appropriate an increasing share of the marital surplus).

It is generally believed that wealthy (landowning) households are more likely to practice sex selection (Murthi et al., 1995). There is also anecdotal evidence that this practice is more prevalent in landowning upper castes; e.g. the Jats and Rajputs in North India (Jeffery et al., 1984) and the Gounders and Kallars in South India (George et al., 1992). Recent research that exploits exogenous changes in property rights (Bhalotra et al., 2019) provides statistical support for the postulated association between sex selection and (land) wealth, which in our model would increase intrinsic son preference. However, this channel is distinct from the implication derived above from our model, and from the models of Edlund (1999) and Bhaskar (2011), which is that sex selection is increasing in relative wealth *within* the marriage market (defined by the caste in India). Our research breaks new ground by verifying this implication.

The empirical analysis uses data from the South India Community Health Study (SICHS). This study covers a rural population of 1.1 million individuals residing in Vellore district in the South Indian state of Tamil Nadu. The study area is representative of rural Tamil Nadu and rural South India with respect to socioeconomic and demographic characteristics; e.g. age distribution, marriage patterns, literacy rates, labor force participation, child and adult sex ratios, and religious composition. The analysis makes use of two components of the SICHS: a census of all 298,000 households drawn from 57 castes residing in the study area, completed in 2014, and a detailed survey of 5,000 representative households, completed in 2016. The survey collected information on the marriage of the primary respondent (the household head) and the marriages of his children (in the preceding five years) which we use for the analyses of dowry and hypergamy. To test the predictions of the model for sex selection, we use the SICHS census data, which include nearly 80,000 children aged 0-6. In general, extremely large data sets are needed to detect sex selection with the required level of statistical confidence. The SICHS census is the only data set we are aware of that is large enough to estimate the relationship between family wealth and sex selection *within* castes.

The terms wealth and permanent income, which determines the resources available for consumption, are used interchangeably in our analysis. The SICHS census and the SICHS survey collected information on the household's realized income in the preceding year. This income includes a transitory component, which we purge to construct measures of permanent income.[3] In addition, we account for the fact that family size varies in practice, in contrast with the model in which each household consists of a single parent and a single child, by using per capita wealth to measure the family's position in the caste wealth distribution.

We begin the empirical analysis by documenting that dowries are always positive, as assumed in the model. Although the model does not have unambiguous implications for the association between dowries and relative wealth, it does predict that dowries given by girls will exceed dowries received by boys at each wealth level (because girls marry up when there is sex selection). Using the relative wealth measures described above, we verify this prediction with SICHS survey data. We then proceed to test the key implications of the model for hypergamy and sex selection. We verify that hypergamy is increasing (using SICHS survey data) and sex selection is decreasing (using SICHS census data) as we move down the wealth distribution within castes.

Although our focus is on relative wealth, other factors, including absolute household wealth, could also determine sex selection. We take advantage of two features of the data to identify the relative wealth effect: (i) the data cover multiple castes, within which independent marriage markets are organized, and (ii) the per capita wealth distribution varies across castes. Once caste fixed effects are included in the estimating equation, the threat to identification is that household wealth or family size, which are used to construct per capita wealth, directly determine or are correlated with other independent determinants of sex selection. Following Das et al. (2003), we thus include a flexible control function, with household wealth and family size as arguments, in the estimating equation. With this research design, we effectively compare families that have the same wealth and the same size, but are located at different positions in their caste's per capita wealth distribution. While we verify that the observed increase in sex selection with relative wealth is robust to alternative specifications of the control function, our preferred specification is determined nonparametrically (based on the fit to the data) by cross validation.

The current child sex ratio in the study area, obtained from the SICHS census, is 109, which is just slightly higher than the corresponding statistic for South India from the 2011 census (108). Although the sex ratio is biased, based on our natural benchmark of 102, it is not exceptional. Nevertheless, within castes, our estimates indicate that the sex ratio varies from 101 in the bottom decile of the wealth distribution, to as

[3] If multiple income realizations are available, the average will provide a measure of permanent income. We take this approach in the dowry analysis, where incomes for relevant households are available from the census and the survey. However, this is infeasible for the analyses of hypergamy and sex selection. For these analyses, household income predicted by (potential) agricultural productivity, based on a detailed historical assessment of the villages in the study area, is used to measure permanent income.

high as 118 in the top decile (which is comparable to the sex ratios in the worst states in the country). The marriage market is organized in the same way in all castes and the positive association between sex selection and relative wealth is obtained caste by caste across the social spectrum in our study area. We would thus expect our results to hold more widely.

We complete the analysis by estimating the structural parameters of the model and conducting counter-factual simulations. The first simulation decomposes the contribution of the son preference mechanism and the marriage market mechanism: based on our estimates, the latter mechanism accounts for 52% of the variation in sex ratios within castes. The next set of simulations examines the effectiveness of alternative programs that attempt to reduce sex selection by targeting the marriage market mechanism. This analysis is especially relevant, given our new findings on the extent of sex selection within castes. In recent years, the central government and many state governments have introduced cash transfer programs rewarding parents if they have a girl, with the objective of reducing daughter aversion. We find that some existing programs, which target specific (low income) households, could actually worsen sex selection overall, by changing the equilibrium marriage price within castes. While this is an important finding in itself, our counter-factual analysis indicates, in addition, that transfers that flow directly to married women and thus avoid the marriage market channel could be substantially more effective than transfers to parents (even if they are altruistic towards their children). Based on this finding, we argue below that an easily implementable change in existing programs, with regard to the timing of the transfers, could substantially increase their impact.

2 A Model of Wealth, Marriage, and Sex Selection

2.1 Marriage in India

We take the following features of the marriage institution in India as given in our analysis. First, marriages are endogamous, matching individuals almost exclusively within their caste or *jati*.[4] Second, marriages are patrilocal, with women moving into their husbands' homes, which are often outside their natal village (Dyson and Moore, 1983; Rahman and Rao, 2004). Third, marriages are arranged by the parents and relatives of the groom and bride, with family wealth being a major consideration when forming a match (Prasad, 1994; Bloch and Rao, 2002; Desai and Andrist, 2010).[5] Fourth, the social norm is that all girls must marry (Caldwell et al., 1983; Arnold et al., 1998; Bhat and Halli, 1999; Basu, 1999). Bloch and Rao (2002) note that "getting one's daughter married is considered an Indian parent's primary duty and to have an older unmarried daughter is a tremendous misfortune with large social and economic costs." The consequence of this social stigma is that marriage is universal for women in India, whereas the marriage rate for men has varied by region and over time, depending on the availability of brides (as documented by Gupta (2014), using census data over

[4]Evidence from nationally representative surveys such as the 1999 Rural Economic Development Survey (REDS) and the 2005 India Human Development Survey (IHDS) indicates that over 95% of Indians marry within their caste. Recent genetic analyses have established that these patterns of endogamous marriage have been in place for over 2,000 years (Moorjani et al., 2013).

[5]Rao (1993a) finds that household characteristics, especially land wealth, matter more for matching than individual characteristics in rural India. Our own results, reported below, indicate that matching on family wealth is independent of matching on education.

the course of the twentieth century).

Although the level of the dowry is determined endogenously in our model, the way in which the payment is organized is determined by the marriage institution. As discussed, marriages are arranged by the bride and groom's parents. In addition, more than 90% of new couples begin their married life co-residing with the groom's parents (Desai and Andrist, 2010). Over time, cash and gold, which can be easily appropriated by the in-laws, have come to constitute an increasing share of the dowry payment (see Prasad (1994) and the references cited therein). Given these features of the marriage institution, it follows that the "dowry is not transferable to the bride, nor does the daughter gain control of the dowry in the way in which the son gains control over land following the partition of his parents' estate. In fact, even the groom's control over the dowry is likely to be subordinate to that of his parents as long as the latter are alive" (Sharma, 1980). The fact that the marriage payment flows to the in-laws before it is distributed within the household, will have important consequences for sex selection in our analysis.

2.2 Model Set Up

The model that we develop in this section isolates those elements of the Indian marriage institution that are responsible for sex selection. The model is thus set up to be as parsimonious as possible, abstracting away from many features of the family and the marriage market that are not directly relevant for the analysis.

Population. Consider a population of families with measure 2. We assume that a family consists of one parent and one child. The gender of the parent is irrelevant. The gender of the child is the purpose of our analysis. Under natural circumstances, without sex selection, a child is born a boy or a girl with equal probability, and the distribution of children would each have measure one.[6] Families are indexed by their wealth z which is distributed according to the measure $\Gamma(z)$ on $[\underline{z}, \overline{z}]$, with $\Gamma(\overline{z}) = 2$. Denote the boy's family wealth by x and the girl's by y. The measure of families with boys and with girls will be endogenous, as will be the distribution of wealth. We denote the wealth distribution of families with boys by $F(x)$ and with girls by $G(y)$. Under natural circumstances, without sex selection, and with equal probability of having a boy or a girl, the wealth distribution of boys is identical to that of girls: $F(\cdot) = G(\cdot) = \frac{1}{2}\Gamma(\cdot)$.

Preferences. Denote the wealth-contingent consumption of parents by C_x, C_y and that of the children by c_x, c_y. All individuals have logarithmic preferences over consumption, and we assume that families maximize the sum of their members' utilities $U = \log(C_i) + \log(c_i) + \mathbb{I}_x u_b, \forall i = \{x, y\}$.[7] Although the focus of the analysis is the marriage market channel for sex selection, we incorporate a coexisting intrinsic son preference channel by assuming that families who have a boy get a utility boost u_b. Denote the maximized utility of

[6] Although the sex ratio at birth is not exactly equal to one in practice, what matters for the analysis is the sex ratio in the marriage market. Under natural circumstances, the adult sex ratio is indeed equal to one.

[7] Equivalently, parents have altruistic preferences over the utility of their children. The assumption that preferences are logarithmic is broadly consistent with Euler equation estimates of the inter-temporal elasticity of substitution; e.g. Attanasio and Weber (1993), Blundell et al. (1994). Although there are only two generations in the model, we can capture the steady state of a fully dynamic model with overlapping generations by interpreting the weight on the child's consumption utility as reflecting the cumulative (discounted) weight on all future generations. It is possible that the parent's and the child's consumption would then no longer receive equal weight, and this would also be true if parents were not perfectly altruistic, but this extension to the model would not change the results that follow.

the groom's family with wealth x marrying a bride with wealth y by $u(x, y)$ and the associated utility for the bride's family by $v(x, y)$.

THE MARRIAGE INSTITUTION. The model incorporates the key features of the marriage institution listed above. Castes form independent marriage markets and we can think of the model as describing one such market. Marriages are arranged, with family wealth being the major consideration when forming a match. The additional institutional feature that is especially relevant for the model is that marriage in India is patrilocal; i.e. women move into their husbands' homes. Patrilocal marriage has benefits and costs for the girl's family. The cost of patrilocal marriage to the girl's family is that the boy's parent is only willing to accept the match if the girl's parent pays a dowry d. The benefit is that the girl will get to consume a fraction of the boy's family's wealth (inclusive of the dowry payment). Matching with a well-off boy is thus especially beneficial.

Based on the previous description of the marriage payment arrangement, we assume that the boy's parent retains control of the dowry. The boy's parent's total endowment, $x + d$, is subsequently allocated in three parts: the boy and his (altruistic) parent each receive a share $\beta/2$, and the remainder, $1 - \beta$, goes to the girl. The girl's share reflects her bargaining position in the marital home. We make the standard assumption that her bargaining position is relatively weak; i.e. $\beta \geq 2/3$. We further assume that there is no heterogeneity in the bargaining position, which is determined by the outside option of being single.[8]

CONSUMPTION. Given the setup described above, the consumption of all agents (parents and children) of a married groom-bride pair (x, y) can be written as:

$$\begin{aligned} c_x &= \frac{\beta}{2}(x+d) \\ C_x &= \frac{\beta}{2}(x+d) \\ c_y &= (1-\beta)(x+d) \\ C_y &= y-d. \end{aligned} \tag{1}$$

The dowry d, and the consumption allocations are determined endogenously in the model as functions of wealth, x and y. Given that x, y are the resources available for consumption, we can equivalently interpret these variables as permanent incomes; i.e. the return on the wealth stock. In a multi-generational model, the stock would be passed on to the next generation, but these dynamic considerations are irrelevant in our model. The terms wealth and permanent income are thus synonymous in our analysis.

MATCHING. Matching in this marriage market is frictionless, with the transfer between the bride and the groom's family d determined competitively.[9] We denote the equilibrium allocation by $\mu(y)$, i.e., the family

[8]While higher education will generally improve the girl's outside options (Anderson and Bidner, 2015), unusually low female labor force participation rates in India weaken this association. Based on the evidence reported below, greater education does not improve the girl's bargaining position in her marital home. In addition, previous studies have proposed that the girl's bargaining position will vary with economic development (Anderson and Bidner, 2015), changes in sex selection technology (Hussam, 2021), and changes in dowry laws (Calvi et al., 2021). These dynamic considerations are outside the scope of our model.

[9]We are effectively ignoring delays in matching, which are likely to be small, given that all marriages occur within the tightly integrated caste. We also ignore dowry-based marital violence, which has been associated with asymmetric information in the

wealth of the groom who is married to a bride with family wealth y is $x = \mu(y)$. The timing of the decisions is as follows: the participants in the marriage market first choose their best partner, given a "Walrasian" schedule of dowries, and the marriage market subsequently clears with a resulting equilibrium price d.

For any match between a girl's family with wealth y and a boy's family with wealth x, and given a dowry d, the utility of the boy's family can be written as:

$$u = u_b + 2\log\left((x+d)\frac{\beta}{2}\right). \tag{2}$$

The utility of the girl's family satisfies:

$$v = \log(y-d) + \log\left((1-\beta)(x+d)\right). \tag{3}$$

The outside option of the child staying single for the boy's family is $S_x = u_b + 2\log\frac{x}{2} - m_b$ and for the girls' family: $S_y = 2\log\frac{y}{2} - m_g$, where m_b, m_g denote the social cost to the family when a boy, girl stays single. Note that if the child stays single, then the altruistic parent will divide the family wealth equally between the two of them. Note also that son preference applies to the family whether the boy is married or single. Given the social norm that all girls must marry, we assume that the cost of her staying single m_g is very high and hence, that the outside option for the girls' side, $S_y \to -\infty$. This difference in options outside marriage for girls and boys will play an important role in generating sex selection below.

2.3 Analytical Solution and Results

We solve the model in three steps. First, we show how families on the two sides of the marriage market match on wealth. Second, we show that there is sex selection, which implies a shortage of girls, at the top and the bottom of the wealth distribution. Third, we show that sex selection is increasing in wealth at the top and the bottom of the wealth distribution. Complementing these results, numerical simulations of the model then show that there is sex selection at each point in the wealth distribution and that sex selection is increasing in wealth across the distribution.

Equilibrium Matching. The dowry d is determined together with the equilibrium matching pattern, $x = \mu(y)$. In competitive equilibrium, the allocation must be optimal for each agent and the market must clear. In the marriage market, we derive conditions for optimality on the girl's side, taking as given the maximized utility on the boy's side, $u(x)$, for each wealth level. Notice that u is now a function of the boy's family wealth x alone because once the marriage price d has been determined in equilibrium, there will be a distinct price for each wealth level.

Using equation (2), we can write the boy's family wealth as:

$$x + d = \frac{2e^{\frac{u-u_b}{2}}}{\beta} \tag{4}$$

bargaining between the groom's family and the bride's family over the distribution of resources (Bloch and Rao, 2002). While such violence is very costly for the victim, its incidence in the general population is likely to be relatively low.

and as a result, the utility of the girl's family can be expressed as:

$$v(x,y,u) = \log\left(x + y - \frac{2e^{\frac{u-u_b}{2}}}{\beta}\right) + \log\left((1-\beta)\frac{2e^{\frac{u-u_b}{2}}}{\beta}\right). \tag{5}$$

A girl's family with wealth y will take the hedonic Walrasian price schedule, $u(x)$, as given when choosing the partner with wealth x that maximizes its utility given in equation (5). This is a matching problem with Imperfectly Transferable Utility (ITU), as analyzed in Legros and Newman (2007). The first order condition to this problem satisfies

$$v_x + v_u u' = 0. \tag{6}$$

Having established the condition for optimality, the remaining condition to be satisfied for a competitive equilibrium is market clearing. We construct the equilibrium allocation $x = \mu(y)$ to ensure market clearing, and then determine the properties of the pattern of matching μ. Denote $M_b \equiv \exp(m_b/2)$.

Proposition 1 *There is Positive Assortative Matching on wealth, i.e., $\mu'(y) > 0$, provided $M_b\beta < 1$.*

Proof. In Appendix. ■

The condition $M_b\beta < 1$ ensures that dowries are positive at every wealth level, as shown below. Given this condition, we establish that there is positive assortative matching and the market will clear from the top, with the wealthiest available girl matching with the wealthiest available boy. Without sex selection, a child is born a boy or girl with equal probability. This implies that the wealth distribution on either side of the market is the same. It follows that girls and boys of equal wealth will match with each other; $y = x$. If there is sex selection at every wealth level, as derived below, then $\overline{y} = \overline{x}$ at the top of the wealth distribution and $y < x$ at all other wealth levels when the market clears from the top.

There is no technological complementarity between the boy's and the girl's wealth in our model. The complementarity that gives rise to positive sorting is derived from the structure of the marriage institution in conjunction with the parents' preferences to leave a bequest. Wealthy parents are willing to pay a higher dowry to secure a wealthy match, which will ensure higher consumption for their daughters.[10] The first order condition, equation (6), ensures that the hedonic price, u, and, hence, the dowry is increasing sufficiently steeply in x so that the matching on wealth is stable. If the dowry was chosen on the basis of the bequest motive alone, then for a given x, a girl's family with wealth y would choose d to maximize $v(x,y,d) = \log(y-d) + \log((1-\beta)(x+d))$. It is straightforward to verify that all girls' parents would then want them to match with the wealthiest boys and the market would not clear. The dowry thus serves *both* as a bequest *and* as a price to clear the marriage market in our model.[11]

[10]The implicit assumption in our model is that girls' parents can transfer any fraction of their wealth as a dowry; i.e. there are no liquidity constraints. The evidence on this assumption is mixed. Recent evidence from India (Anukriti et al., 2022; Corno et al., 2020) indicates that income shocks affect the timing of marriage and that girls' households change their labor supply leading up to the marriage. However, there is an extensive literature, summarized in Munshi and Rosenzweig (2016), that documents (close to) full risk-sharing in rural India in the face of income shocks that include major contingencies such as illness and marriage.

[11]This dual role for the dowry distinguishes our model from existing models of marriage with dowries. In Botticini and Siow (2003) the marriage market clears by wealth matching between brides and grooms, and dowries serve only as a bequest.

Our assumption that families match on wealth alone is consistent with the empirical evidence that household characteristics, especially land wealth, matter more for matching than individual characteristics in rural India (Rao, 1993a). Our own results, reported below, indicate that matching on household wealth is independent of matching on an important individual characteristic (education). We could add individual characteristics to the model, but this would not generate additional empirical implications and the matching problem then becomes a multi-dimensional allocation problem, which is analytically intractable once the wealth distribution is allowed to be endogenous (on account of sex selection).[12] A related consideration motivates the assumption that each family consists of a single parent and a single child in our model. If the family consisted of two parents and multiple children, instead, then sex selection decisions would need to be modeled inter-dependently within the family and we would be faced with a multi-dimensional matching problem once again.[13] The advantage of our simplifying assumptions is that they allow us to focus on a single key source of variation – family wealth – in the model, although we will condition on other dimensions of heterogeneity, including education and family size, in the empirical analysis.

Sex Selection. Intrinsic son preference and the marriage market mechanism work interchangeably to leave parents better off with boys than with girls; $u(y) - v(y) > 0$ at each wealth level y in equilibrium, as made precise below. For this gain from having boys to translate into a biased sex ratio, a sex selection technology must be available. We assume that parents who are expecting a girl can replace her with a boy (with probability one) at a utility cost k, which is distributed according to the cumulative density function $H(k)$.[14] k incorporates the monetary cost, which is relatively small, and the more important legal and ethical cost of sex selection. We assume that k is uncorrelated with wealth and is bounded below at zero.[15] Although sex selection decisions are made many years before the child enters the marriage market, note that there is no uncertainty in our model. Parents correctly anticipate how the marriage market will clear in the future and, hence, sex selection and matching (with resulting equilibrium dowries and consumption allocations) can be modeled simultaneously.

Before proceeding to establish that sex selection is present, we first derive the following result:

Lemma 1 *Dowries are positive at every wealth level provided $M_b\beta < 1$.*

Proof. To see why this is the case, consider the utility of a married boy's family when the dowry is zero: $u(x) = u_b + 2\log\frac{\beta x}{2}$. This can be compared with the family's utility when he is single: $u_b + 2\log\frac{x}{2} - m_b$. The

In Anderson and Bidner (2015), the dowry serves both roles, but two separate instruments are available. Following common convention, we refer to the marriage transfer as the "dowry" throughout the paper. The technically more accurate terminology is that the price component of the transfer is the groom-price and the bequest component is the dowry (Anderson, 2007).

[12] Multi-dimensional matching problems are difficult to solve even with exogenous distributions and linear preferences. See, for example, Choo and Siow (2006) and Lindenlaub (2017).

[13] Abstracting away from family composition is not uncommon in the economics literature. In research on the labor market, for example, the vast majority of papers assume labor supply and job search decisions are made by individuals independently. However, we know that those decisions are jointly determined within the family. For instance, a spouse may choose to stay at home if his wife has a well-paid job, but not if she is unemployed.

[14] In reality, the decision is more subtle. First, if parents use sex selective abortion rather than infanticide or neglect to eliminate unwanted girls, then all parents who anticipate that they will make this decision must bear the *ex ante* cost of sex determination. Second, even if parents do eliminate a girl, there is no guarantee that the next pregnancy will result in a boy. There is thus a stochastic element to the cost of sex selection that we abstract from in our modeling choice.

[15] Given that parents are altruistic, it may be more reasonable to assume that the lower bound for k is strictly positive. This would have no qualitative bearing on the results that follow.

family will be better off with the boy remaining single if $2\log\beta < -m_b$ or, equivalently, $M_b\beta < 1$. It follows that if this condition is satisfied, the dowry must be strictly positive at every wealth level x for the marriage market to be active. ■

Given Lemma 1, which is consistent with the observation that dowries are universal in India, it can be shown that there will be sex selection at the extremes of the wealth distribution. The matching function, $\mu(y)$, and the dowry, d, are determined simultaneously in equilibrium, together with sex selection. Although this simultaneity prevents us from analytically deriving the level (or presence) of sex selection at each wealth level, this is possible at the extremes because (i) under positive sorting, matching is exogenously determined at the top of the distribution, with the wealthiest girls marrying the wealthiest boys, and (ii) the dowry is pinned down at the bottom of the distribution, where the family of the last boy to match is indifferent between him staying single or marrying.

Proposition 2 *In equilibrium, there is sex selection at the top and the bottom of the wealth distribution:*

1. *at the top $y = \overline{y}$, for some $\underline{u}_b(\beta)$, which is declining in β, whenever $u_b > \underline{u}_b$;*
2. *at the bottom $y = \underline{y}$, for some $\overline{m}_b$, whenever $m_b < \overline{m}_b$.*

Proof. In Appendix. ■

Given that the cost of sex selection, k, is bounded below at zero, we need to simply establish that $u(y) - v(y) > 0$, in equilibrium, for sex selection to be present at wealth level y. At the top of the wealth distribution, girls' families with wealth $\overline{y}$ match with boys' families with equal wealth $\overline{x}$. We first establish that $u(\overline{y}) - v(\overline{y}) > 0$; i.e. parents would be better off with a boy than a girl, when the equilibrium dowry, d, is set to zero, if u_b, β are sufficiently large. In particular, this condition is satisfied if $u_b > \underline{u}_b$, where $\underline{u}_b(\beta)$ is declining in β, with $\underline{u}_b(2/3) = -log(1/3)$ and $\underline{u}_b(1) \to 0$. We next establish that $u(\overline{y}) - v(\overline{y})$ is increasing in d, which implies that this condition must be satisfied for all d. The preceding result is consistent with Alfano (2017) and Bhalotra et al. (2020) who document that an exogenous increase in the dowry is associated with an increase in sex selection.

At the bottom of the wealth distribution, girls' families with wealth $\underline{y}$ match with boys' families with wealth x^*, where x^* is determined by the level of sex selection higher up the distribution. We first exploit the fact that the family of the last boy to match is indifferent between him staying single or marrying to derive the dowry $d(x^*)$, where $d'(x^*) > 0$, that the least wealthy girl's parent must pay. Second, we show that for $x^* = \underline{x} = \underline{y}$, $u(\underline{y}) - v(\underline{y}) > 0$ if $u(\overline{y}) - v(\overline{y}) > 0$. Third, we show that $u(\underline{y}) - v(\underline{y})$ is decreasing in x^* if $x^* \leq \underline{x}^*$, where $d(\underline{x}^*) = \underline{y}/2$ and increasing in x^* for $x^* > \underline{x}^*$. The nonmonotonicity arises because as girls match up in wealth; i.e. as x^* increases, they consume at a higher level, but their parents also pay a higher dowry. The preceding result indicates that the first effect dominates initially, but the second effect takes over when x^* crosses a threshold level. To complete the proof we thus need to establish that $u(\underline{y}) - v(\underline{y}) > 0$ when x^* is equal to $\underline{x}^*$. We show that this condition will be satisfied if m_b is sufficiently small.

Boys with wealth less than x^* are left unmatched. If there was bride-price in equilibrium; i.e. payments from boys to girls at the time of marriage, and the population was growing, then the deficit of girls could be cleared by boys "buying" younger girls (Tertilt, 2005; Neelakantan and Tertilt, 2008; Bhaskar, 2011).[16] In our model, with dowry payments in equilibrium, the marginal boy that stays single prefers that state to marrying the least wealthy girl (and receiving the dowry that comes with her, conditional on his wealth). If the population is stationary, as it has been in South India since the mid-1990's, and assuming that the wealth distribution is unchanged in the short-run, that marginal boy will similarly prefer being single to marrying the least wealthy girl in any successive cohort. Thus, the deficit of girls in our model cannot be cleared by allowing boys to match across cohorts.[17]

Wealth and sex selection. Proposition 2 establishes that there will be sex selection at the top and at the bottom of the wealth distribution if u_b, β are sufficiently large and m_b is sufficiently small. Numerical simulations reported below show that if these conditions are satisfied, then there will be sex selection at each point in the wealth distribution. We next proceed to describe how the level of sex selection changes with wealth. The example that we constructed above with 100 wealth classes indicates that sex selection will decline in equilibrium as we move down the wealth distribution. The analysis that follows formalizes this argument.

With positive assortative matching, girls with family wealth y match with boys with family wealth $\mu(y)$, where $d\mu(y)/dy > 0$. When a family with wealth y that is expecting a girl decides to have a boy instead, it will receive utility $u(y) - k$. Note that the boy will then match with a poorer girl with family wealth $\mu^{-1}(y)$. If the family had chosen instead to keep the girl, it would have received $v(\mu(y), y; u(\mu(y)))$, which we know from Proposition 2 (and the numerical simulations that follow) is less than $u(y)$. Thus, the family will proceed with sex selection if its cost $k < u(y) - v(\mu(y), y; u(\mu(y)))$. In general, for families with wealth y there is a critical cutoff $k^\star$ satisfying

$$k^\star(y) = u(y) - v(\mu(y), y; u(\mu(y))), \tag{7}$$

where families with $k < k^\star$ choose sex selection. Given that the cost of sex selection, k, is distributed according to the cumulative density function, $H(k)$, the fraction of families with wealth y that choose sex selection is thus $H(k^\star(y))$. For what follows and without loss of generality, we assume H uniform on $[0, a]$. The pattern of sex selection at every wealth level generates an endogenous and distinct distribution of wealth for girls and boys. The economy-wide distribution of wealth z is $\Gamma(z)$. The measure of families with boys whose wealth exceeds x and the measure of families with girls whose wealth exceeds y can thus be described

[16] If the population is stationary, then the age-gap between husbands and wives will widen over successive cohorts until the girls are too young to marry. To illustrate this argument, consider the following thought experiment. Suppose that we are out of steady state and the number of boys is double the number of girls and all boys marry at the age of 25. Then the first cohort of boys will marry girls aged 25 and 24, the second cohort will marry girls aged 24 and 23, and so on. Eventually the girls will be too young and some boys must remain unmarried. This is independent of the sex ratio as long as it is greater than one.

[17] Consistent with this argument, we see in Appendix Figure A1, using data from the SICHS census, that the age-gap between spouses has actually narrowed over time (across successive age cohorts).

as follows:

$$F(x) = \int_{x}^{\overline{x}} (1 + H(k^{\star}(z))d\Gamma(z)/2 \;\text{ and }\; G(y) = \int_{y}^{\overline{y}} (1 - H(k^{\star}(z))d\Gamma(z)/2, \tag{8}$$

where $\overline{x} = \overline{y} = \overline{z}$. With Positive Assortative Matching, the market clearing condition is

$$\int_{\mu(y)}^{\overline{x}} dF(x) = \int_{y}^{\overline{y}} dG(y) \tag{9}$$

or equivalently:

$$\int_{\mu(y)}^{\overline{x}} (1 + H(k^{\star}(z))d\Gamma(z)/2 = \int_{y}^{\overline{y}} (1 - H(k^{\star}(z))d\Gamma(z)/2. \tag{10}$$

Sex selection determines the distribution of wealth for boys and girls, which, in turn, determines the pattern of matching in equation (10). The pattern of matching determines sex selection in equation (7). Sex selection and assortative matching must thus be solved simultaneously. This two-way interaction between a particular family decision and the sorting equilibrium is a common feature of marriage models in the family economics literature.

If we knew the payoff $u(x)$ at every level of wealth x on the boys' side, then we could solve for sex selection and matching recursively, starting at the top of the wealth distribution and moving down. We would know $\mu(y)$ at any wealth level y on the girls' side, given the pattern of sex selection at higher wealth levels, and so would be able to compute $u(y) - v(\mu(y), y; u(\mu(y)))$ and, hence, $H(k^{\star}(y))$. However, the hedonic price schedule $u(x)$ must also be derived endogenously in the model. To do this we integrate the first order condition in equation (6), $v_x + v_u u' = 0$, which implies $u' = -\frac{v_x}{v_u}$, with respect to x:

$$u(x) = \int_{x^{\star}}^{x} -\frac{v_x(x, \mu^{-1}(x); u(x))}{v_u(x, \mu^{-1}(x); u(x))} dx + u(x^{\star}) \tag{11}$$

where the denominator is negative, and where $x^{\star}$ is the lowest wealth boy who is matched (and is indifferent between marrying and staying single). From the outside option, we know that $u(x^{\star}) = u_b + 2\log\frac{x^{\star}}{2} - m_b$.

The equilibrium is fully defined by the sex selection condition, the matching condition, and the payoff condition, as specified in equations (7), (10), and (11). This system of equations must be solved simultaneously. The additional consideration is that the payoff condition holds a fixed point because $u(x)$ appears on both sides of equation (11). We cannot solve the system of equations analytically to determine sex selection at each wealth level. However, the model can be solved numerically (see Section 2.4). We can, moreover, obtain analytical results at the very top of the wealth distribution where the matching pattern is exogenously determined; $\overline{y} = \overline{x}$, and at the lowest wealth level at which boys match, $x^{\star}$, where $u(x^{\star}) = 2\log\left(\frac{x^{\star}}{2}\right) - m_b$.

Proposition 3 *Sex selection is increasing in wealth (i.e., $\frac{dk^{\star}(y)}{dy} > 0$):*

1. *at the top $y = \overline{y}$, whenever $d < \frac{\overline{y}}{2}$;*

2. at the bottom $y = \underline{y}$, whenever $d(x^*) < \frac{\underline{y}}{2}$.

Proof. In Appendix. ■

To prove Proposition 3, we differentiate the expression for $k^*(y)$ in equation (7) with respect to y and then use the first order condition (6), together with expressions for v_x, v_u, v_y derived from equation (5), to establish a general condition under which $k^*(y)$ is increasing in y. We then show that this condition is satisfied at the top of the wealth distribution if girls with wealth $\overline{y}$, who match with the wealthiest boys with wealth $\overline{x}$, pay a dowry that is less than $\overline{y}/2$. We also show that this condition is satisfied at the bottom of the wealth distribution if girls with wealth $\underline{y}$ pay a dowry that is less than $\underline{y}/2$. Given that the dowry $d(x^*)$ is increasing in the wealth, x^*, of the last boy to match, this condition will be satisfied if x^* is not too large; i.e. if sex selection higher up the wealth distribution is not too severe.

If the result in Proposition 3 holds over the entire wealth distribution, then this implies that sex selection increases monotonically with relative wealth. The intuition for this conjecture, which is validated by the numerical results that follow, is that the shortage of girls grows as we move down the wealth distribution (because more boys are left unmatched above them) which allows their families to retain an increasing share of the marital surplus. In the extreme, at the bottom of the wealth distribution, the family of the last boy to match is pushed down to its outside option; i.e. it is indifferent between him marrying and staying single. The least wealthy girls benefit the most from the surplus of boys and, hence, sex selection is lowest at the bottom of the wealth distribution.

2.4 Numerical Solution and Results

The algorithm. The numerical solution of the model assumes that there is a finite number of wealth classes. This implies that boys and girls in a given wealth class could potentially match across multiple wealth classes. The matching allocation then looks like a step function instead of a smooth curve. With a continuum of wealth classes, the first order condition in equation (6), $\frac{dv}{dx} = 0$, ensures that the allocation and transfers are optimal for girls' families in each wealth class. With a finite number of wealth classes, the equivalent condition is that girls' families in a given wealth class will obtain the same utility across all the wealth classes that they match with. Given that the equilibrium payoff for the boys' families, $u(x)$, is a function of their wealth alone, the symmetric condition is that boys' families in a given wealth class receive the same utility across all the wealth classes that they match with.

The solution to the model must satisfy the sex selection condition, the measure preserving allocation or matching condition, and the payoff condition simultaneously. The algorithm that we use to solve the model numerically begins with an initial guess for the payoff at the top of the wealth distribution, $u(\overline{x})$, and for the pattern of matching. We know from Proposition 2 that there will be sex selection at the top and the bottom of the wealth distribution. This implies that there will be a shortage of girls in the highest wealth class and so girls in the next to highest wealth class will match up (with boys one wealth level higher than themselves) and horizontally (with boys in their own wealth class). As we move down the wealth distribution, the excess of boys accumulates and it is possible that below some wealth level, girls match exclusively with wealthier boys.

Given any initial guess for $u(\overline{x})$ and the matching pattern, we can solve for $u(x)$ and $v(y)$ in each wealth class. $v(y)$ is a function of y, x, and $u(x)$, as specified in equation (5). Given that girls in the highest and the next to highest wealth class match with the wealthiest boys, with family wealth $\overline{x}$ and payoff $u(\overline{x})$, we can solve for v in both wealth classes. Girls' families in the next to highest wealth class must receive the same utility, v, from matching with the wealthiest boys and boys in their own wealth class. This allows us to solve for u in the next to highest wealth class. We continue to solve recursively in this way down the wealth distribution.

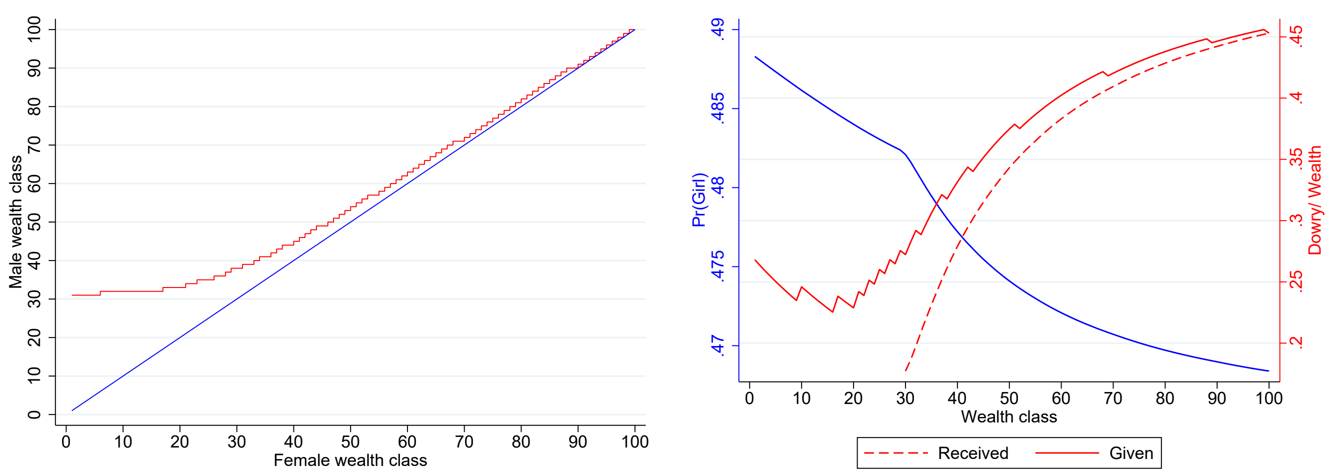
With sex selection, boys below a wealth level $x^\star$ will remain unmatched. A comparison of $u(x^\star)$ derived in the first iteration with the outside option, $u_b + 2\log\left(\frac{x^\star}{2}\right) - m_b$, is used to adjust the guess for $u(\overline{x})$ in the next iteration. Given u and v derived in the first iteration, the level of sex selection $H(k^*(y))$ can be determined in each wealth class y. The pattern of matching implied by this sex selection is used as the starting point for the next iteration. This iterative process continues until there is convergence. The numerical solution thus simultaneously satisfies the sex selection condition, the matching condition, and the payoff condition.

Numerical simulations. The wealth distribution is assumed to be log-normal in the numerical simulations, with the parameters selected to match the SICHS census data (within castes). The wealth distribution is divided into 100 classes for the simulations. As assumed above, $k \sim U[0, a]$, which implies that there are four parameters in the model: β, u_b, m_b and a. We select values for these parameters: $\beta = 0.8$, $u_b = 0.7$, $m_b = 0.12$, and $a = 23$ that are in line with the values estimated below.

Figure 1: **Simulated Model**

(a) Marital Matching

(b) Sex Selection and Dowry



The matching pattern generated by the model is reported in Figure 1a. Notice that the plot is not a smooth function and has small steps. This is due to the discreteness of the wealth distribution, which results in each wealth class matching with multiple wealth classes of the opposite sex as described above. At higher wealth classes, girls and boys match horizontally as well as up and down, respectively. This is why the plot touches the 45 degree line at those wealth levels. However, below a certain wealth level, girls match exclusively with wealthier boys, shifting the plot above and away from the 45 degree line. Hypergamy, or the wealth-gap, increases as we move down the wealth distribution because of the growing stock of unmatched boys, with the intercept of the plot measuring the fraction of boys that end up being single.

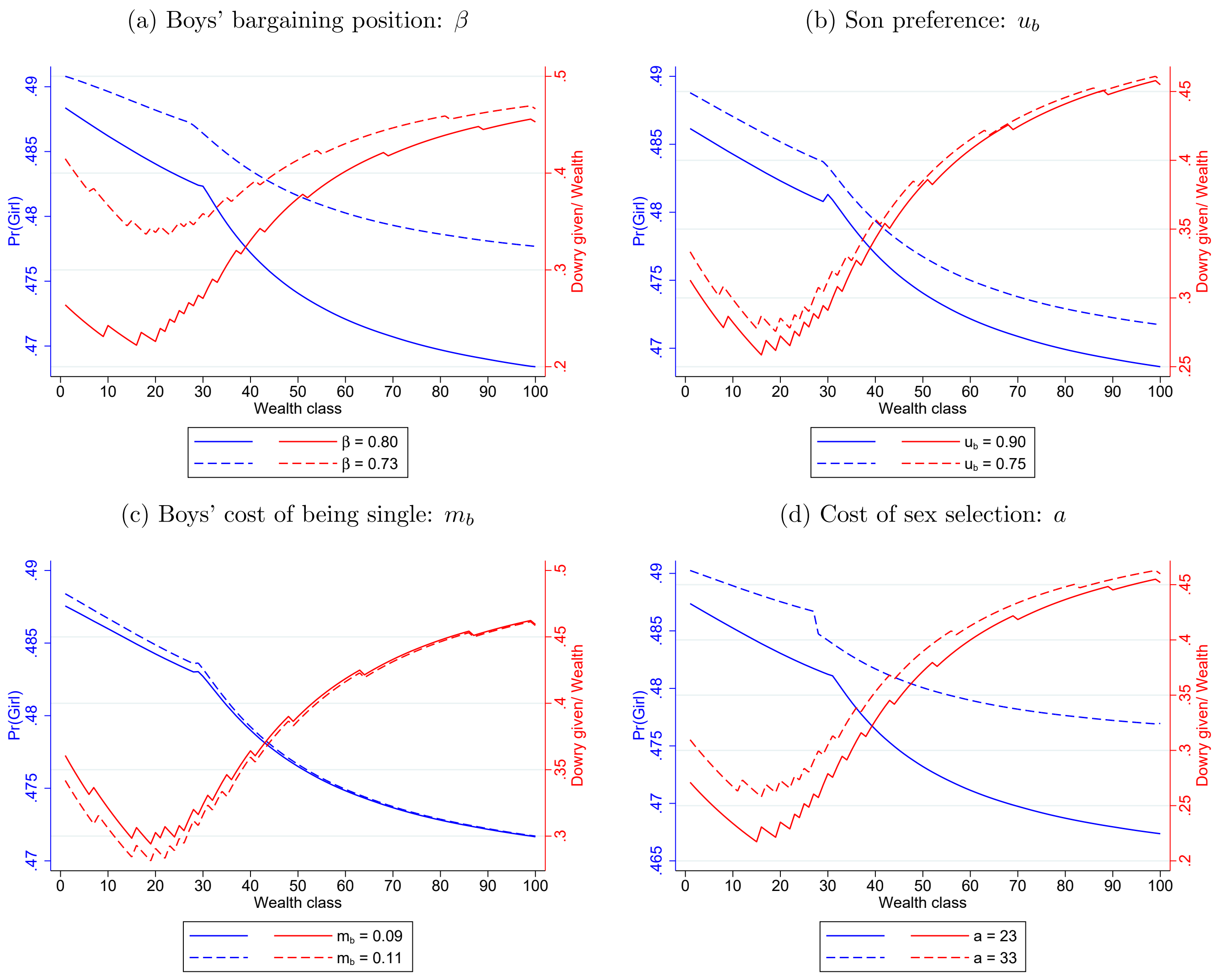
We can compute the dowry that boys of a given wealth class x receive directly from the value of $u(x)$ derived for that wealth class. Recall that $u(x) = u_b + 2\log\left(\frac{\beta}{2}(x+d)\right)$. While this is the same, regardless of the wealth of the girls' families that they match with, the dowry paid by girls in a given wealth class will vary with the wealth of the families that they match with. It is thus necessary to take account of the matching pattern in each wealth class when computing the average dowry paid by girls' families over the wealth distribution. Given that girls are matching up on average, the model unambiguously predicts that the dowry given must be greater than the dowry received at each wealth level. This is indeed what we observe in Figure 1b. Moreover, the dowry is positive at all wealth levels if the condition specified in Lemma 1 is satisfied and this is also true in our numerical simulation. Note that the model does not have a clear prediction for how the dowry will vary across the wealth distribution. As we move down the distribution, girls match with increasingly wealthy boys, but their family's share of the marital surplus is also increasing. The first effect shifts up the dowry, whereas the second effect works in the opposite direction and, hence, the net effect on the dowry is ambiguous. We see in Figure 1b that the dowry (as a ratio of household wealth) is initially decreasing and then increasing in wealth.

We have shown analytically that there will be sex selection at the top and the bottom of the wealth distribution (Proposition 2) and that sex selection is increasing in wealth at the extremes (Proposition 3). In contrast with the model's ambiguous implications for the dowry, we expect sex selection to be monotonically declining as we move down the wealth distribution because girls are in increasingly short supply (and their families thus receive an increasing share of the marital surplus). This is indeed what we observe in Figure 1b. Notice that the dowry is less than half the family's wealth at the top and the bottom of the wealth distribution. This satisfies the condition for sex selection to be increasing in wealth at both points, as derived in Proposition 3.

Figure 2 describes how sex selection and dowries vary across the wealth distribution for different values of the model's parameters. Boys' bargaining position, β, and the intrinsic utility from having a boy, u_b, work interchangeably to generate sex selection, via the marriage market and son preference channels, and we see that an increase in either of these parameters, in panels (a) and (b), results in an increase in sex selection. Notice that there is an accompanying decline in dowry payments, this is especially pronounced with the marriage market channel (panel (a)) at low wealth levels, reflecting the fact that as sex selection increases further up the wealth distribution, girls' families receive an increasing share of the marital surplus.

Panel (c) describes the corresponding comparative statics exercise with the m_b parameter, which measures the cost to the boy's side from being single. The value of this parameter is directly relevant for the last boy to match – his family is indifferent between him marrying and staying single, and as m_b increases, the value of staying single declines. An increase in m_b thus reduces the utility of the family of the last boy to match, increasing the share of the marital surplus on the girl's side, with a resulting decline in sex selection at the

Figure 2: **Comparative Statics**



bottom of the wealth distribution. This reduces sex selection at higher wealth levels through changes in the marriage price and we will observe similar general equilibrium effects below when examining the impact of counter-factual policy experiments. Finally, panel (d) examines the effect of an exogenous change in the cost of sex selection. Recall that the cost of sex selection, $k \sim U[0, a]$. As the parameter a increases, the cost of sex selection also increases, with an accompanying increase in the fraction of girls at each wealth level.

A consistent finding from the comparative statics exercises in Figure 2 is that exogenous changes in the model's parameters that lead to an increase in sex selection are accompanied by a decline in the dowry in equilibrium. While this negative correlation can be explained by the fact that girls are in shorter supply, and thus their families have greater bargaining power, it is not at odds with recent evidence cited above, which indicates that exogenous increases in the dowry increase sex selection. Moreover, the latter result does not tell us why sex selection is observed in equilibrium. To generate sex selection through the marriage market mechanism, what is required is the following: (i) the absence of commitment (ii) $\beta > 2/3$ (iii) $m_b < m_g$

(which we assume goes to infinity). If the boy's parent could commit to transferring the dowry in its entirety to the daughter-in-law, then the girl's parent would make a marriage payment of $y/2$. The dowry would be increasing in wealth, but there would be no sex selection. This would be true even in the absence of commitment if $\beta = 2/3$, with positive assortative matching on wealth in equilibrium. Thus, we need (i) and (ii) to generate sex selection. We also need (iii), because without the norm that all girls must marry, a girl's parent could avoid the disutility of having a girl (due to (i) and (ii)) by leaving her single.

3 Empirical Analysis

3.1 Descriptive Evidence

Demographic and socioeconomic characteristics. The South India Community Health Study (SICHS) covers a rural population of 1.1 million individuals residing in Vellore district in the state of Tamil Nadu. There are 298,000 households drawn from 57 castes in the study area. The study area is representative of rural Tamil Nadu (with a population of 37 million) and rural South India (comprising Tamil Nadu, Andhra Pradesh, Karnataka, and Maharashtra with a total population of 193 million) with respect to demographic and socioeconomic characteristics.[18]

Appendix Table A1, Panel A, reports the age distribution, marriage patterns, literacy rates, and labor force participation, separately for males and females, in the study area, rural Tamil Nadu, and rural South India, respectively. Statistics for Tamil Nadu and South India are based on official Government of India data, while the corresponding statistics for the study area are derived from the SICHS census. The age distribution and marriage patterns are combined in a composite statistic that measures the number of married individuals in 5-year age categories as a fraction of the total population, separately for men and women. If this statistic is the same across two populations, then it follows that both the age distribution and marriage rates must be the same in these populations. A Kolmogorov-Smirnov test cannot reject the null hypothesis that the age distribution of married individuals is equal, for men and for women, between the study area and both rural Tamil Nadu and rural South India. Literacy rates and labor force participation rates, for men and for women, are similarly comparable between the study area and both rural Tamil Nadu and rural South India. Literacy rates are much higher for men than for women, 80% versus 60%, although this gender gap has largely disappeared for children currently enrolled in school (see Appendix Table A2). Labor force participation rates match the patterns for literacy; 80% for men versus 40% for women.

Appendix Table A1, Panel B compares the religious composition across the three populations. Over 90% of these populations is Hindu and thus our characterization of the marriage institution in the model, based on Hindu social norms, applies to almost the entire population. We complete the description of demographic and socioeconomic characteristics in Panel B by examining overall and child (aged 0-6) sex ratios. Overall sex ratios in the population are close to parity, which can be explained by the fact that sex selection in South

[18]The SICHS was designed to examine a variety of socioeconomic phenomena and health problems, including the treatment of tuberculosis. The study area thus comprises three Tuberculosis Units (TU's) within Vellore district that were purposefully selected to be representative of rural South India. Munshi and Rosenzweig (2016) define the South Indian region by the same set of states. Kerala is excluded from the list of South Indian states because it is an outlier on many socioeconomic characteristics.

India is a relatively recent phenomenon and because life expectancy is greater for females than males.[19] In contrast, child sex ratios, which are comparable across the three populations, are clearly above the natural fertility benchmark, which we define as 102 boys per hundred girls, based on statistics for South India from the 1961 and 1971 rounds of the population census.

The benchmark we have chosen reflects the history of sex selection in South India. Marriages in this region were traditionally between close-kin (Dyson and Moore, 1983). The most preferred match for a girl was her mother's younger brother or, if he was unavailable, one of her mother's brothers' sons (Kapadia, 1995). Given the extremely close pre-existing relationship between the girl's natal family and her husband's family, the two families effectively functioned as a cooperative unit. There were no major payments at the time of marriage, just a ritual gift or *stridhan* from the groom's side to the girl (Srinivas, 1989; Anderson, 2007). Having a girl did not put parents at a disadvantage in this long-term arrangement in which the two families (dynasties) sequentially traded girls across the generations, and thus there was no sex selection. Caldwell et al. (1983) and Srinivas (1984) attribute the demise of this system to economic development and the resulting changes in wealth within castes. Families that had traded girls over many generations no longer had the same level of wealth. Close-kin marriage declined (Caldwell et al., 1983; Kapadia, 1993) and a marriage market consequently emerged to match unrelated families within the caste on wealth, with a marriage price or dowry clearing the market.[20] By the 1980's, the practice of dowry was observed across the caste distribution in South India (Caldwell et al., 1983). The widespread emergence of dowries in South India in the early 1980's coincided with the onset of sex selection, which is why sex ratios prior to that point in time can be assumed to be unbiased.[21]

Marriage patterns. The analysis in this paper makes use of two components of the SICHS: a census of all households and a detailed survey of 5,000 households who are representative of the castes in the study area.[22] The survey collected information on key aspects of the marriage institution: (i) whether marriage was within the caste, (ii) whether marriage was between close-kin, (iii) whether the marriage was arranged, and (iv) whether the female spouse was born in a different village. This information was collected from the (male) primary respondent for his own marriage and for the marriages of his children in the five years preceding the survey.

Table 1 provides information on marriages over the two generations based on data from the SICHS survey.[23] Consistent with nationally representative survey evidence and genetic evidence for the country as

[19]Sex selection only commenced in South India in the early 1980's and so the first cohorts to be affected would be 36 years old when SICHS data collection was completed in 2016.

[20]In related research, Anderson (2003) links dowry inflation to economic development and increased income inequality on the male side of the marriage market.

[21]It is possible that sex selection due to son preference predates the 1980's in which case we will underestimate the extent of sex selection.

[22]The sampling frame for the household survey included all ever-married men aged 25-60 in the SICHS census plus (a small number of) divorced or widowed women with "missing" husbands who would have been aged 25-60, based on the average age-gap between husbands and wives. The sample was subsequently drawn to be representative of each caste in the study area, excluding castes with less than 100 households in the census.

[23]The larger number of marriages for daughters versus sons in the last 5 years is because girls marry younger than boys in India, as observed with the SICHS census data in Appendix Figure A2. Although most girls marry in their twenties, men will marry into their thirties. Men who marry in their thirties will have fathers in their sixties. Given that the fathers in the survey are aged 25-60, there are more girls of marriageable age in the households in our sample.

Table 1: **Marriage Patterns**

Generation	Parents	Children	
		Males	Females
	(1)	(2)	(3)
Same caste	0.97	0.95	0.95
Related	0.48	0.35	0.35
Arranged	0.86	0.80	0.88
Female moved outside natal village	0.75	0.78	0.81
Mean dowry [as a fraction of annual income]	–	138.32	187.46
	–	[1.46]	[1.84]
Minimum dowry (in thousand Rupees)		0	0
Maximum dowry (in thousand Rupees)		1,117.2	1,417.2
Observations	3,524	421	611

Source: SICHS household survey.

a whole, 97% of the parents and 95% of the children married within their caste. The incidence of close-kin marriage declines, in line with the general trend in South India described above, from 48% in the parents' generation to 35% in the current generation. However, most marriages continue to be arranged. Girls moved from their natal village in a substantial fraction of the marriages. We will take advantage of this feature of the marriage institution in India, by exploiting information on the natal villages, to test the model's predictions for hypergamy below.

Table 1 also reports the dowry in levels and as a fraction of the household's annual income, for the marriages of the children that took place in the last five years. The dowry amount is computed by summing up the monetary value of gifts, such as household items, vehicles, and gold, as well as the cost of the wedding celebration.[24] The annual income is measured by the profit in the past year from land owned, leased, or rented plus the wage earnings of all adult members. Dowries in South India are now as high as they are in the North (Caldwell et al., 1983; Srinivas, 1984; Rahman and Rao, 2004; Anderson, 2007) and in line with past studies; e.g. Rao (1993b), Jejeebhoy and Sathar (2001), Rahman and Rao (2004), and Chiplunkar and Weaver (2021) the dowry is 1.5 – 2 times the household's annual income on average, which is a substantial sum in an economy where access to market credit is severely restricted.[25] Notice that dowries paid by girls are larger than dowries received by boys. This observation is consistent with our model when there is sex selection because girls will then marry wealthier boys on average. Hypergamy results in girls paying a higher dowry than boys receive at each level of wealth, as documented in our theoretical simulations in Figure 1b.[26] Finally, notice that the minimum dowry given and received is equal to zero, consistent with our assumption

[24]The list of items for the dowry include bed, bureau, kitchen utensils (bronze and stainless steel), grinder, mixer, refrigerator, TV, microwave, washing machine, silk saris, groceries, motorcycle, bicycle, car, gold jewelry (in grams), and cash (in Rupees).

[25]Most households will receive support from their close relatives and other caste members to pay the dowry. Munshi and Rosenzweig (2016) use data from the Rural Economic and Development Survey (REDS) to document that gifts and loans within the caste are the primary source of support for meeting major contingencies, including marriage, in rural India.

[26]If wealthier girls (boys) marry relatively early (late), then wealthier girls and less wealthy boys will be over-represented in the marriage statistics. Although we report unconditional comparisons of girls and boys in Table 1, the analysis of dowries and hypergamy that follows will be conditional on family wealth.

in the model that dowries are always positive (Lemma 1).

Table 2: **Hypergamy**

Sex of the child	Males (1)	Females (2)
Partner's parental household		
Wealthier	0.09	0.18
Same wealth	0.62	0.64
Less wealthy	0.29	0.17
Kolmogorov-Smirnov test of equality	P-value = 0.001	
Observations	421	611

Source: SICHS household survey.

Hypergamy. Table 2 provides evidence indicative of hypergamy based on the SICHS survey data. The survey respondents were asked whether their child's spouse's family had the same wealth, more wealth, or less wealth than their own. These are coarse categories and the majority of marriages, for sons and daughters, are reported to be with families of equal wealth. However, the respondents are more likely to report that their daughters married up in wealth than their sons. Conversely, they are more likely to report that their sons married down in wealth than their daughters. The Kolmogorov-Smirnov test easily rejects the null hypothesis that the distribution of responses is equal for sons and daughters. Upper caste marriages in North India have long been associated with hypergamy (Bhat and Halli, 1999). Hypergamy has also been associated with the emergence of dowry in South India (Caldwell et al., 1983; Srinivas, 1984). These are all settings with sex selection. However, previous studies have failed to make the connection between hypergamy and sex selection. Indeed, given that marriages are almost exclusively within the caste, girls cannot marry up on average without sex selection.

Sex selection. Although the model assumes that each parent has a single child, a couple will usually have multiple children. The marriage market channel will bias the sex ratio of children at all birth orders. In contrast, if parents have access to a relatively certain sex selection technology and they want a single male heir, either to support them in old age or to inherit their wealth (property), then they will postpone sex selection until they are close to their desired family size. In particular, first births will not be biased due to the son preference channel as long as parents want at least two children, which is typically the case in rural India.

Table 3 reports child (aged 0-6) sex ratios from three sources: the 2005-2006 and 2015-2016 rounds of the Demographic and Health Survey (DHS) and our own SICHS census, which was conducted 2012-2014, between the two DHS rounds. All three data sources include the birth order of each child and, hence, it is possible to compute sex ratios for first-born children and for all children. We see that the sex ratios are very similar across the three data sets in Table 3.[27] The sex ratio for first born children is elevated relative to the

[27] The DHS statistics are computed for the entire South Indian region because sample sizes are too small to measure sex ratios

Table 3: **Sex Ratios**

Population	Rural South India		Rural Vellore
Data Source	DHS 2005-06	DHS 2015-16	SICHS census 2012-14
	(1)	(2)	(3)
First-born children	105	107	106
All children	109	110	109
Observations	5,750	27,072	79,027

Note: The India Demographic and Health Survey (DHS) is also known as the Indian national Family health Survey (NFHS). Sex ratios are computed for children aged 0-6 as the number of boys per 100 girls.

natural sex ratio of 102, indicating that the marriage market channel is active.[28] Moreover, the sex ratio for all children is larger than for first born children, implying that the son preference channel is also present. If we assume that the marriage market mechanism affects the sex ratio at all birth orders equally, while the son preference mechanism only operates at higher birth orders, then 55% of sex selection in the study area can be attributed to the marriage market mechanism.

3.2 Measuring Wealth

The model generates predictions for variation in hypergamy (the wealth-gap between grooms and brides) and sex selection across the wealth distribution within castes. It also predicts that dowries given by girls will be greater than dowries received by boys at each wealth level (assuming that dowries are positive everywhere). Recall that the terms wealth and permanent income are used interchangeably in our analysis. The SICHS census and the SICHS survey both collected information on the household's income in the preceding year.[29] In an agrarian economy, this income will include a substantial transitory component, which must be purged to measure permanent income.

In general, there are two ways to purge the transitory component from the observed income realization. First, if multiple income realizations are available, then the average over time provides a measure of permanent income. We can take this approach in the dowry analysis because households in the SICHS survey, which

in Tamil Nadu. The comparison with SICHS sex ratios is still valid because we saw in Appendix Table A1 that the study area is representative of rural Tamil Nadu and rural South India with respect to socioeconomic and demographic characteristics. Child sex ratios for rural Tamil Nadu and rural South India, based on the 2011 population census, were also seen to be very similar to the corresponding SICHS statistic. Despite the fact that we are including an entire region in the country, notice that the number of children in the DHS is still substantially smaller than in the SICHS census, emphasizing the novelty of our data.

[28]This observation does not necessarily contradict the common assumption that the sex ratio at birth for first-born children is unbiased in India. Jayachandran and Kuziemko (2011) and Jayachandran and Pande (2017) document differential nutritional inputs by gender and birth order among children in India, which could elevate the child (aged 0-6) sex ratio even if the sex ratio at birth is unbiased.

[29]Household income is measured by the profit from land owned, leased, or rented plus the total labor income of all members, including those that have temporarily migrated to work. Profit is measured over the entire year, whereas labor income is measured in the month prior to data collection (and then scaled up to the annual level).

provides the information on dowries, are also included in the SICHS census. The same approach cannot be utilized, however, for the analysis of sex selection and hypergamy. The analysis of sex selection is based on the SICHS census and the vast majority of households in the census do not appear in the SICHS survey. The analysis of hypergamy is based on SICHS survey data, but information on the spouse's family wealth, which is needed to examine the pattern of matching, is not directly available. All we know is its caste affiliation and the village in which it resides.

To construct a consistent measure of permanent income that can be used for the inter-linked analyses of sex selection and hypergamy, we take a second approach in which household income predicted by a set of fixed factors provides us with a measure of permanent income. In a rural economy, permanent income will be largely determined by agricultural productivity. Although the SICHS census does not provide detailed information on farm characteristics, we have collected historical records on agricultural productivity in each village in the region encompassing the study area from the British Library in London. There are 377 *panchayats* or village governments in the SICHS study area. These *panchayats* were historically single villages, which over time sometimes divided or added new habitations. The *panchayat* as a whole, which often consists of multiple modern villages, can thus be linked back to a single historical village. The data we have collected includes information on the agricultural revenue tax, per acre of cultivated land, that was levied on each village in 1871 by the British colonial government. The tax assessment was based on the *potential* output per acre. This, in turn, was determined by growing conditions such as soil quality, crop suitability, temperature, and rainfall, which are effectively permanent (fixed). In addition, it is well known that historical wealth, which would have been correlated with the tax assessment, will have persistent effects in a developing economy. For both of these reasons, we expect the 1871 tax revenue to be a strong predictor of current income.

The 1871 tax revenue does not account for heterogeneity within the village. Land ownership and occupations vary by caste and we thus expect the relationship between current household income and village-level tax revenue in 1871 to vary by caste. We allow for this by including the 1871 tax revenue, caste fixed effects, and a full set of revenue-caste interactions in the estimating equation when predicting current household income.[30] Net of the fixed effects, our measure of historical wealth is a strong predictor of current household income; the F-statistic measuring joint significance of the village-level tax revenue and the revenue-caste interactions is as high as 20.4. Predicted income based on this specification will thus be our benchmark measure in the analysis of both sex selection and hypergamy. This measure is especially convenient for the hypergamy analysis because permanent income can be predicted for spouses whose villages lie outside the study area. Additional tests, which can be implemented with sex selection (but not hypergamy) as the outcome, will verify that the results are robust to the inclusion of village fixed effects and the education of the household head and his wife in the estimating equation when predicting permanent income.

In the model, each family consists of a single parent and a single child. In reality, family sizes vary and this will determine the resources that are available to each member. We account for this feature of the data

[30]The implicit assumption underlying this relationship is that households, or dynasties, have remained in the same village over many generations. This assumption is supported by recent evidence that permanent migration from rural to urban areas is extremely low in India (Munshi and Rosenzweig, 2016). Standard errors in this estimating equation are clustered at the *panchayat* level.

in the empirical analysis by using per capita wealth to determine the family's position in the caste wealth distribution. Households consisting of a single couple and their children, but possibly including other adults (typically a grandparent) account for 96.2% of all households with children in the census. The empirical analysis is based on these households and per capita wealth is computed by dividing household wealth by the number of family members; i.e. the two parents plus their children.[31] Note that we cannot make this adjustment in the hypergamy analysis because the spouse's family size is unobserved. We will thus examine marital matching on the basis of household (rather than per capita) wealth, interpreting the results accordingly.

Although there is a single cohort in the model, in practice the age-gap between partners can vary across marriages. In lieu of a clear partition of age cohorts into independent marriage markets within the caste, we compute the family's relative wealth with respect to the entire caste in the benchmark measure. The implicit assumption with this measure is that the distribution of per capita wealth within the caste is stable across age cohorts.[32] In addition, we construct an alternative measure of relative wealth, which is based on the set of families within each caste that are included in the estimation sample for a given outcome. For example, with sex selection as the outcome, the sample consists of children aged 0-6. The alternative measure of relative wealth for that outcome would be based on their families. The implicit assumption when constructing this measure is that the 0-6 year olds in each caste will form an independent marriage market in the future. Although our benchmark measure is based on a definition of the marriage market that may be too expansive and the alternative measure may be based on a definition that is too narrow, the results below are similar with both measures for each outcome.

3.3 Testing the Model

Evidence on Dowries. Table 4 reports the association between dowries and relative wealth. The model does not have unambiguous implications for the sign of this association; as observed in Figure 1b, the association is initially negative and then positive. In addition, our measure of the dowry in the analysis that follows is the amount given or received and not the ratio with respect to household wealth, as in Figure 1b, because wealth also appears on the right hand side of the estimating equation and thus any measurement error in that variable will bias the coefficient on relative wealth.

The estimation sample in Table 4 consists of all marriages of the primary respondents' children that took place in the five years preceding the SICHS survey. When the child is a girl, the dowry is based on the amount that was given, and when the child is a boy, the dowry is based on the amount that was received. With hypergamy, the amount that is given by girls (who are marrying up) will exceed the amount that is received by boys (who are marrying down) at each wealth level, as described in Figure 1b. We thus include

[31]The implicit assumption is that other adults in the household do not receive a share of the wealth. For example, a grandfather living with his son's family would have already distributed his wealth among his children. The results that follow are robust to constructing per capita wealth on the basis of household size.

[32]Fertility has been close to replacement in Tamil Nadu since the mid-1990's. Family size will thus be stationary across age cohorts. The household wealth distribution, which is determined by variation in historical village-level tax revenue will also be stationary across age cohorts.

Table 4: **Dowry and Relative Wealth**

Dependent variable	Dowry		
Wealth measure	SICHS survey (1)	Average of SICHS survey and census (2)	Average per capita (3)
Relative wealth	37.045**	79.761***	98.716***
	(15.675)	(16.137)	(15.814)
Mean of dependent variable	167.63	167.63	167.63
Female dummy	Yes	Yes	Yes
Caste FE	Yes	Yes	Yes
Observations	991	991	991

Source: SICHS survey. Sample based on all children's marriages in the past 5 years. Dowry measured in thousands of Rupees. Relative wealth measured by rank in the caste wealth distribution, from 0 (poorest) to 1 (wealthiest). In Column 1, the unadjusted income reported in the SICHS household survey is used as a measure of household wealth, Column 2 uses the average of the reported income from the SICHS survey and the SICHS census, Column 3 uses average per capita income. Standard errors in parentheses. *** p<0.01, ** p<0.05

a gender dummy in the estimating equation. We also include a full set of caste dummies to account for the fact that some castes will be wealthier than others and thus report higher dowries on average.[33]

Table 4, Column 1 reports estimates with the most basic specification, where the family's relative wealth is based on the unadjusted income reported in the SICHS household survey. We observe that the coefficient on relative wealth is positive and significant. The preceding estimates are subject to two sources of bias: (i) the transitory component is not purged from the household's income, and (ii) we do not account for variation in family size to construct a measure of per capita income. Table 4, Column 2 accounts for the first source of bias by using the average of the reported income from the SICHS survey and the SICHS census to measure household wealth. This will purge the wealth statistic of the transitory component, which would generate measurement error and accompanying attenuation bias, and indeed we observe that the coefficient on relative wealth increases from Column 1 to Column 2. Table 4, Column 3 addresses the second source of bias by adjusting for family size. Using per capita income to construct the relative wealth measure, the coefficient on relative wealth increases even further. Although a similar sequential exercise cannot be implemented with the hypergamy analysis that follows, these results will help us interpret the observed matching patterns.

Figure 3a reports nonparametric estimates corresponding to Table 4, Column 3 (our preferred specification).[34] The family's position in its caste's wealth distribution in Table 4 and in the benchmark specification shown in blue, with the accompanying 95% confidence interval, is based on all (surveyed) households in the caste. The alternative construction of relative wealth, shown in red, where the wealth distribution is based only on households with marriages in the five years preceding the survey, yields very similar estimates. Our

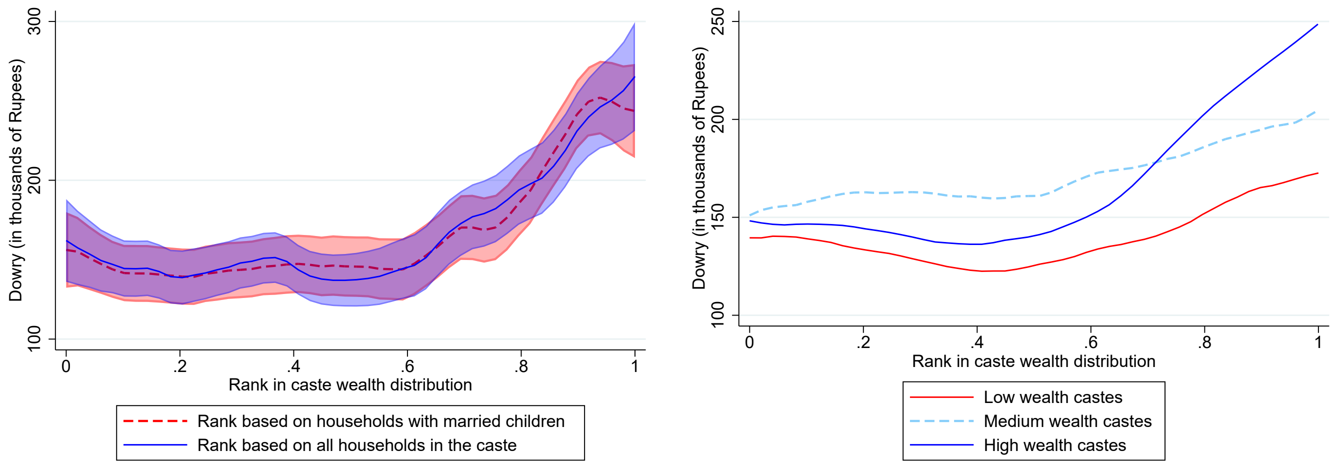
[33]Recall from the model that the level of the dowry is pinned down by the outside option of the last boy to match, which is increasing in his family wealth.

[34]The gender dummy and the caste fixed effects are partialled out using the Robinson (1988) procedure prior to the nonparametric estimation reported in Figure 3a. The same procedure is used in all the nonparametric regressions that follow.

Figure 3: **Dowry**

(a) Dowry and Relative Wealth

(b) Sample Partitioned by Caste Wealth



theoretical results rely on the assumption that dowries are always positive (Lemma 1) and we see in Figure 3a that this is true at each point in the wealth distribution within castes. Figure 3b subjects this result to further scrutiny by partitioning the castes in our sample into terciles, based on average per capita household wealth, where we see that dowries are always positive within each caste group.

Although the model does not have precise implications for the relationship between dowries and wealth, it does tell us that dowries given should be greater than dowries received at each wealth level, as observed in Figure 1b, when there is sex selection (on account of the accompanying hypergamy). This is indeed the case in Figure 4, where we report the association between dowries and relative wealth, separately for boys and girls. The amount given by the girls exceeds the amount received by the boys at each wealth level, with little overlap in the confidence intervals.[35]

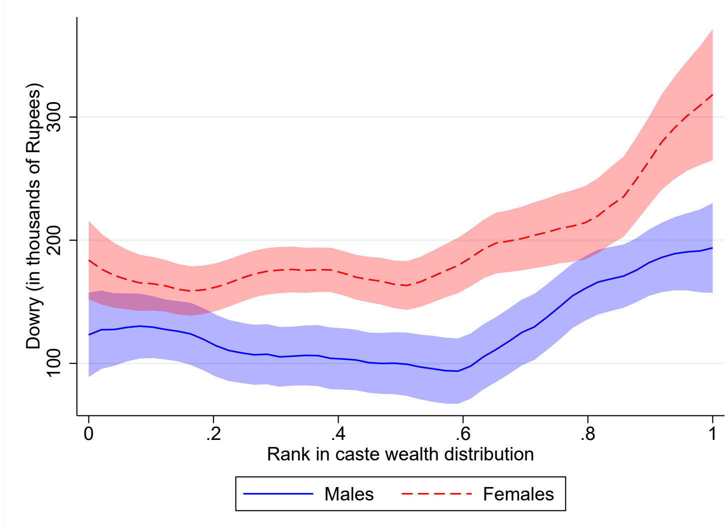
Evidence on Hypergamy. The model predicts that the wealth-gap between matched boys and girls is largest at the bottom of the wealth distribution, with a narrowing of the gap as we move up the distribution until there is convergence at the very top. This theoretical prediction is presented graphically in Figure 1a.

We have information from the household survey on the primary respondent's marriage and the marriages of his children that occurred in the five years prior to the survey. As discussed above, the challenge with the hypergamy analysis is that information on the spouse's household wealth and family size are not directly available. To measure the spouses' household wealth, we take advantage of the fact that the identity of their natal villages was collected in the survey. Based on the estimated relationship between current income and historical village-caste wealth, we can construct a consistent measure of permanent income for the surveyed household and the spouse's household (based on its caste and the historical tax revenue in its natal village).[36] These permanent incomes will typically not be the same because, as observed in Table 1, 80% of women

[35] Recall from Figure 1b that the dowry, as a ratio of household wealth, converges for boys and girls as we move up the wealth distribution. This is not inconsistent with Figure 4 because we are not dividing dowries by wealth

[36] The 1871 tax revenue is available for all villages in the northern Tamil Nadu region that were directly taxed by the colonial government. The estimated equation can thus be used to predict current income even for villages outside the study area.

Figure 4: **Dowry and Relative Wealth (by gender)**



move from their natal village when they marry. While marital migration generates wealth variation on the two sides of the market that we exploit for the hypergamy analysis, the analysis is based on household wealth and not per capita household wealth. We cannot adjust for variation in family size because, as noted, this information is unavailable for the spouses and we will take account of this when interpreting the results that follow.

Table 5: **Marital Matching**

Dependent variable	Relative wealth of groom		Relative education of groom
	(1)	(2)	(3)
Relative wealth of bride	0.541***	0.540***	-0.057
	(0.033)	(0.033)	(0.031)
Relative education of bride	–	0.011	0.479***
		(0.034)	(0.031)
Constant	0.227***	0.222***	0.422***
	(0.018)	(0.024)	(0.022)
Observations	708	708	708

Notes: Sample restricted to primary respondents born after 1980 and children who married in the past 5 years. Relative wealth measured by rank in the caste wealth distribution, from 0 (poorest) to 1 (wealthiest). Education measured relative to all females/males in the SICHS census in the same caste who are no more than 5 years younger or older. Bootstrapped standard errors in parentheses. *** p<0.01. Source: SICHS survey.

Table 5, Column 1 reports estimates from an equation in which the male partner's relative wealth is the dependent variable and the female partner's relative wealth is the independent variable. The primary respondents in the household survey range in age from 25 to 60. The younger respondents are thus born around the same time as the children of the oldest respondents. To increase statistical power, our sample thus includes both generations, with the restriction that the included primary respondents be born after 1980; these birth cohorts would have been subjected to sex selection, even in South India, with subsequent gender imbalance in the marriage market. The estimated constant term, which corresponds to the intercept in Figure 1a, is positive and statistically significant as predicted. The coefficient on the female's relative wealth is also positive and significant and, moreover, significantly smaller than one. This implies that there will be convergence in the wealth of the partners' families as we move up the wealth distribution, once again in line with the model. It is quite striking that variation in wealth across villages can be used to uncover marital hypergamy within castes. To the best of our knowledge, this is the first statistical evidence documenting the presence of a phenomenon that has received much attention in Indian sociology and anthropology.

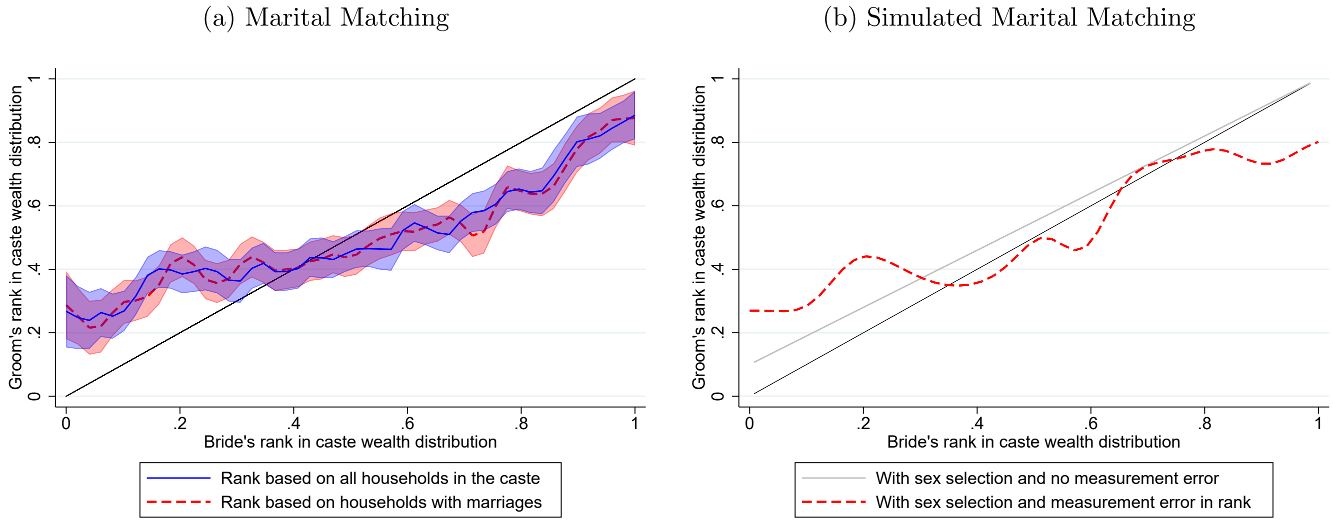
Table 5, Column 2 includes the female's education, measured by her rank among female members of her caste who were born in a 10-year window around her birth year, as an additional regressor.[37] The coefficient on the education variable is small in magnitude and statistically insignificant. A comparison of Columns 1 and 2 indicates that the relative wealth coefficient is hardly affected by the introduction of female education, supporting the assumption in the model that matching on family wealth is independent of individual characteristics. To provide further support for this assumption, we replace the wealth of the boy's family by his relative education as the dependent variable in Table 5, Column 3. The coefficient on the female's family wealth is now small in magnitude and statistically insignificant, whereas the coefficient on her (relative) education is positive and significant. Matching on family wealth appears to be independent of matching on individual characteristics, at least with respect to one important characteristic: education. One potential response to the sex selection problem would be for girls' parents to invest in their education as a substitute for the inefficient dowry mechanism. However, such substitution is unlikely to be observed in practice because female labor force participation in India is exceptionally low, implying that investments in girls' education will have little material benefit. The absence of a tradeoff between the girl's education and her spouse's wealth in Table 5 is consistent with this argument.[38]

Figure 5a subjects the results in Table 5 to closer scrutiny by reporting nonparametric estimates of the relationship between the male's relative wealth and the female's relative wealth. The household's position in its caste's wealth distribution in Table 5 and in the benchmark specification in Figure 5, shown in blue with the accompanying 95% confidence interval, is based on all (surveyed) households in the caste. The alternative construction of its relative wealth, shown in red, is based on the subset of households in the estimation sample. The results are evidently robust to the method that is used to construct relative wealth. However, Figure 5 does not quite match Figure 1a. In particular, although the intercept is positive and there is convergence to the 45 degree line, the estimated intercept is higher than what would be implied by the excess of boys; given

[37]The relative education level is computed with respect to women from the caste in the same age range in the SICHS census. This is because the number of observations in the SICHS survey declines substantially once we go down to the caste-gender-age level.

[38]If there was a tradeoff, then we would also expect educated parents, who are more easily able to invest in their children's education, to take advantage of it (and, hence, have more girls). As reported in Appendix Table A2, the sons and the daughters of more educated parents are more likely to be enrolled in higher secondary school. However, sex selection, measured by the probability that a child aged 0-6 is a girl, is independent of parental education, once again indicating the absence of a tradeoff. These results are also consistent with the assumption (discussed above) that more educated girls do not have greater bargaining positions in their marital homes. If they did, then more educated parents, with more educated daughters, would be more likely to have a girl.

Figure 5: **Hypergamy**



the level of sex selection in our study area, this should be around 0.1. The convergence is also too sharp, with the result that the wealthiest girls do not end up matching with the wealthiest boys.

Our explanation for the preceding discrepancy is based on measurement error in the wealth variables, which will attenuate the association between boys' and girls' wealth, shifting up the intercept and flattening the slope of the nonparametric relationship reported in Figure 5a. We examine this bias in Figure 5b by simulating the association between boys' and girls' relative wealth, with and without measurement error. We use wealth observed in the data for this exercise and specify that sex selection results in 10% of the boys being unmarried. The solid grey line in the figure describes the pattern of matching without measurement error; the intercept is at 0.1 and the matching line converges to the 45° line when relative wealth equals one, as in Figure 1a. The red dashed line describes the estimated relationship between boys' and girls' relative wealth when the underlying matching remains the same, but a mean-zero noise term is added to the wealth measures observed by the econometrician. We observe that the estimated pattern of matching looks similar to Figure 5a. An additional bias arises because we cannot adjust for family size and construct measures of per capita wealth in the hypergamy analysis. Based on the dowry analysis in Table 4, we would expect the resulting bias to reinforce the bias due to measurement error. Note that neither source of bias will undermine the analysis of sex selection that follows; we will be able to adjust for family size and while measurement error will continue to attenuate the association between sex selection and (relative) per capita wealth, the resulting bias will provide a conservative estimate of this association.

Evidence on Sex Selection. The central prediction of the model is that sex selection will increase as we move up the wealth distribution within castes. Large samples are needed to identify sex selection with the requisite level of confidence. For the analysis of sex selection within castes, we thus turn to the SICHS census data; recall from Table 3 that there are nearly 80,000 children aged 0-6 in the study area.

The dependent variable in the equation that we use to test for sex selection is a binary variable indicating whether a child is a girl. The key explanatory variable is the child's family's position in the caste per capita wealth distribution. Per capita wealth in Table 6, Column 1 is constructed as the household's predicted income, based on historical village tax revenue and its caste, divided by family size. We also include caste fixed effects in the estimating equation to allow for the possibility that norms governing the (social) cost of sex selection could vary by caste. The coefficient on relative per capita wealth is negative and statistically significant, as predicted by the model.[39] The analysis that follows will subject this result to further scrutiny by (i) examining its robustness to alternative construction of the relative wealth variable and various partitions of the data, and (ii) by incorporating other independent determinants of sex selection in the estimating equation.

Table 6: **Sex Selection and Relative Wealth**

Dependent variable	Girl dummy		
	(1)	(2)	(3)
Relative wealth	-0.0441***	-0.0528***	-0.0590***
	(0.0086)	(0.0090)	(0.0113)
Household wealth	–	0.0200***	0.0137**
		(0.0054)	(0.0073)
Father's education	–	–	-0.00307
			(0.0030)
Mother's education	–	–	-0.00137
			(0.0017)
Mean of dependent variable	0.480	0.480	0.480
Observations	78877	78877	78877
Caste FE	Yes	Yes	Yes
Village FE	No	No	Yes

Source: SICHS census. Sample restricted to children aged 0-6 years. Relative wealth measured by rank in the caste per capita wealth distribution, from 0 (poorest) to 1 (wealthiest). Father's and mother's education is measured as their highest level of education, in years. Bootstrapped standard errors (in parentheses) are clustered at the *panchayat* level. **p<0.05, *** p<0.01

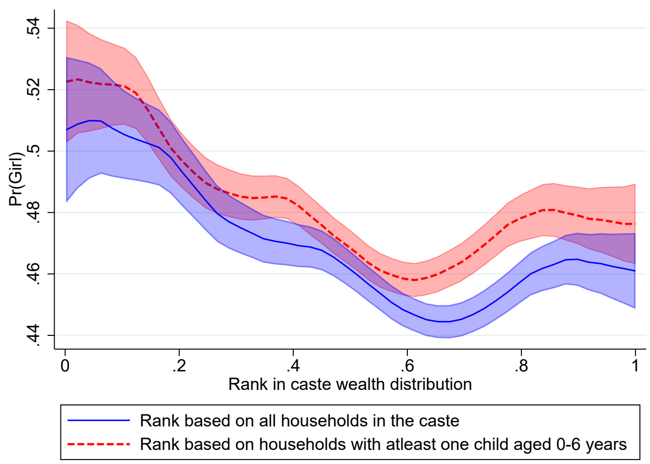
While we consider omitted variable bias more generally and in greater detail in the subsequent section, we introduce some obvious independent determinants of the sex selection decision in the specifications that follow. We begin in Table 6, Column 2 by adding (absolute) household wealth, which is mechanically positively correlated with our relative wealth measure, to the estimating equation. In principle, the net effect of household wealth on sex selection is ambiguous. Landowning households will have a greater demand for a male heir, but wealthy parents are less dependent on their children for old-age support. Household wealth could also be associated with other independent determinants of sex selection; for example, wealthier households might have better access to sex selection technology. In our data, we see that the household wealth (predicted income) coefficient is positive and significant, while the negative coefficient on relative wealth remains significant and increases in absolute magnitude from Column 1 to Column 2. We augment the estimating equation in Column 3 by including village fixed effects and parental education. The village

[39]We use the Linear Probability model to estimate the relationship between the sex of the child and relative wealth for ease of interpretation because the mean of the dependent variable is close to 0.5.

effects will capture spatial variation in access to sex selection technology, while the education variables could potentially determine sex selection through multiple channels. Note that this specification includes the village fixed effects and parental education when predicting household income to construct relative per capita wealth. The relative wealth coefficient continues to be negative and significant in Column 3 and is similar in (absolute) size to the coefficient in Column 2.

Figure 6: **Sex Selection**

(a) Sex Selection and Relative Wealth



(b) Sample Partitioned by Caste Wealth

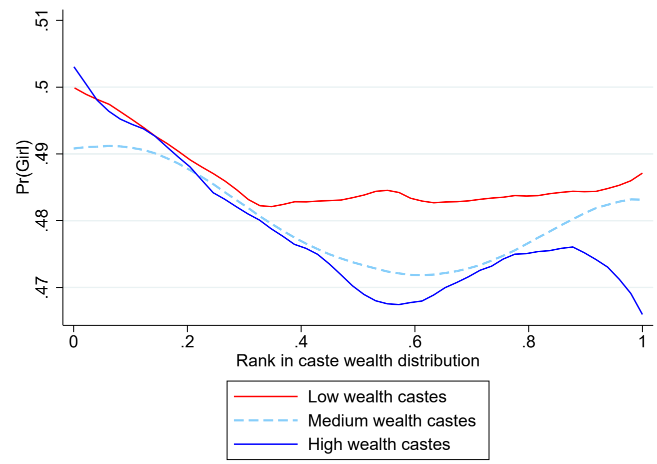


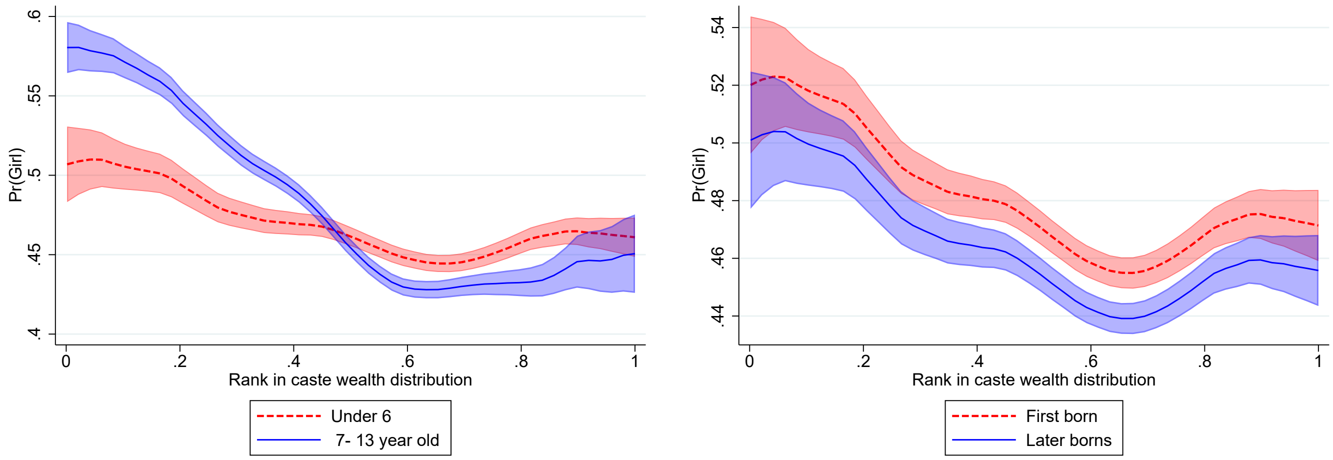
Figure 6a reports nonparametric estimates corresponding to Table 6, Column 1. The family's position in its caste's wealth distribution in the benchmark specification shown in blue, with the accompanying 95% confidence interval, is based on all households in the caste. The alternative construction of relative wealth, shown in red, is based only on households in the caste with at least one child aged 0-6. This is a very different set of households, but the negative relationship between the probability that a child is a girl and relative wealth continues to be obtained. Figure 6b partitions the castes in our sample into terciles, based on average per capita wealth. Wealthier castes have more biased sex ratios, in line with the qualitative studies cited above. However, the negative association between the probability that a child is a girl and relative wealth within the caste is retained in each caste group. The within caste variation in sex ratios is comparable in magnitude to the between caste-group variation in Figure 6b. We will return to this observation in Section 4.1, where we quantify and compare within-caste and between-caste variation in sex ratios.

When constructing per capita wealth, household wealth must be divided by family size. We have an accurate measure of family size in the dowry analysis because the children getting married are adults and even if they have younger siblings, the fertility of their mothers will be complete.[40] This is not the case for the sex selection analysis, which is based on children aged 0-6. We account for this in Figure 7a by separately estimating the association between the probability that a child is a girl and relative wealth for children aged 0-6 and 7-13. Family size will be more accurately measured for the older cohort and while we would not want to compare the two cohorts directly, it is reassuring to observe that the same negative association is

[40]Very few families in the study area have more than three children and birth-spacing rarely exceeds five years.

Figure 7: **Sex Selection across Groups**

(a) Across age groups (b) Across birth order



obtained. Figure 7b partitions the sample by birth order. We observe a negative association between the probability that a child is a girl and relative wealth for first-born children and later-born children. The sex ratio is less biased for first-born children, consistent with the statistics reported in Table 3, on account of the son preference channel.

3.4 Alternative Explanations

Sex selection in our model is determined by intrinsic son preference and the marriage market channel. However, other determinants of sex selection could coexist with these mechanisms. For example, the increasingly biased sex ratios over time in India have been attributed to a number of factors including: (i) improved sex selection technologies (Arnold et al., 2002; Bhalotra and Cochrane, 2010); (ii) changes in the economic returns to having boys versus girls (Rosenzweig and Schultz, 1982; Foster and Rosenzweig, 2001); and (iii) reduced fertility coupled with a desire to have at least one son (Basu, 1999; Jayachandran, 2017).[41] The same factors could generate cross-sectional variation in sex selection.

To understand the biases that could be generated by these independent determinants of sex selection, consider the following estimating equation:

$$P(y_{ij} = 1) = f(w_{ij}, w_{-ij}) + \delta_j + \xi_{ij},$$

where $y_{ij} = 1$ if child i in caste j is a girl and 0 if a boy, w_{ij} is the per capita wealth of the child's family, w_{-ij} is a vector representing the per capita wealth of all other families in the caste-based marriage market, and δ_j are caste fixed effects. The $f(w_{ij}, w_{-ij})$ function determines the family's relative wealth and the ξ_{ij} term

[41] As noted, sex ratios will worsen at higher birth-orders when there is a demand for a male heir. Exogenous fertility decline makes the sex ratios worsen earlier, resulting in an overall increase in the bias. This argument has been used to explain why programs that couple incentives to reduce fertility and to have daughters could result in a worsening of the sex ratio (Anukriti, 2018).

collects the independent determinants of sex selection. Access to sex selection technology and the differential economic returns to boys versus girls could potentially vary with household wealth, W_{ij}. A reduction in family size, N_{ij}, worsens the sex ratio when there is a demand for a single male heir. In addition, W_{ij} and N_{ij} could be correlated with other independent determinants of sex selection. Per capita wealth, $w_{ij} = \frac{W_{ij}}{N_{ij}}$ and, hence, $E(\xi_{ij} \mid w_{ij}, w_{-ij}, \delta_j) \neq 0$. The estimated relationship between the sex of the child and relative per capita wealth will thus be biased.

Our solution to this identification problem is to include a flexible control function, $g(W_{ij}, N_{ij})$, in the estimating equation. This solution differs from the approach of including a limited number of control variables, which are often imperfect proxies for the omitted variables, in the estimating equation. The key to our strategy is the observation that the determinants of sex selection listed above and, more generally, any independent household-specific determinant, can only bias our estimates of the relative wealth effect if they are correlated with ω_{ij}, through W_{ij} or N_{ij}.

It is not standard practice to use the control function approach, first proposed by Heckman (1976) in the context of selection correction, to deal with omitted variable bias. This is because the objective with this approach is to subsume the entire component of the error term, ξ_{ij}, that is correlated with the running variable, ω_{ij}, in a single control function, and this is typically infeasible when there are multiple omitted variables. The control function approach is especially effective in dealing with selection bias because there is a single decision that determines this bias. A polynomial function of a single propensity score (estimated in a first-stage selection equation) can thus be used to extract the component of the error term that is associated with the selection bias (Heckman and Navarro-Lozano, 2004). The control function approach can be used to correct for omitted variable bias in our application because: (i) any household-specific determinant of sex selection can only bias our estimates through W_{ij}, N_{ij}, (ii) there are multiple marriage markets organized at the level of the caste, and (iii) the wealth distribution varies across castes. Once the $g(W_{ij}, N_{ij})$ function is included in the estimating equation, we are effectively comparing the sex selection decisions of families with the same wealth, W_{ij}, and size, N_{ij}, but who have different relative per capita wealth because they belong to castes with different wealth distributions.

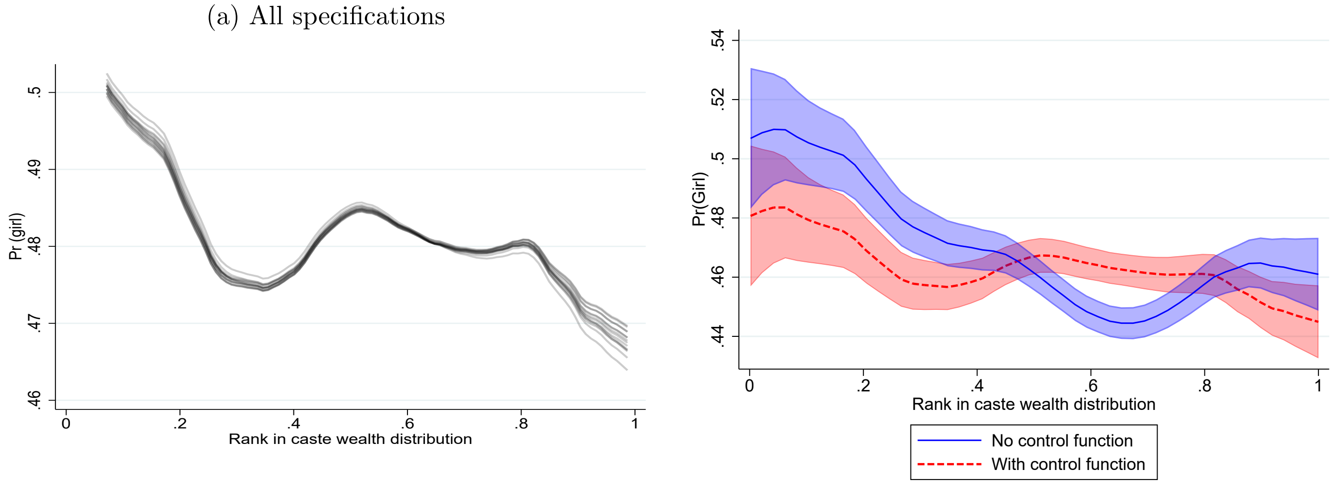
When using instrumental variables to deal with omitted variable bias, the objective is to find a variable that is (mean) independent of the error term. This rules out the use of choice variables as instruments because they will typically be jointly determined with other unobserved determinants of the outcome of interest. With the control function approach, the objective is to capture the component of the error term that is correlated with the running variable in its entirety. Independence is not required with this approach (Heckman and Navarro-Lozano, 2004), and choice variables – family size in our application – and outcomes – household wealth in our application – can be included in the control function without undermining the identification strategy. The challenge in this case is to correctly specify the control function: it should not be too parsimonious (under-fitting the data) nor should it be too flexible (over-fitting the data).

Very few families in the study area have more than three children and, hence, family size, N_{ij}, is specified as a vector of binary variables indicating whether family i in caste j has 3, 4, or 5 (and above) members. We allow the control function $g(W_{ij}, N_{ij})$ to be linear, quadratic, cubic, or quartic in household wealth, W_{ij}. In addition, we allow for a wide range of interactions between W_{ij} and N_{ij} in the control function. For example, when $g(W_{ij}, N_{ij})$ is specified to be a quadratic function of W_{ij}, we allow for (i) no interactions between the family size dummies and the wealth terms, (ii) interactions with the linear wealth term, and (iii) interactions with both the linear and quadratic wealth terms. This leaves us with 14 possible specifications of the control function. Figure 8a reports the relationship between the probability that a child is a girl and relative wealth with each of these specifications, using the Robinson procedure to partial out the terms in the control function and the caste fixed effects as usual. There is little variation in the estimated relationship across the alternative specifications of the control function and the probability that the child is a girl continues to be declining in relative wealth.

Figure 8: **Sex Selection and Relative Wealth - Control Function**

(a) All specifications

(b) Optimal specification



Das et al. (2003) recommend cross validation to determine the optimal specification of the control function when correcting nonparametrically for sample selection and endogeneity. We use k-fold cross validation, with $k = 10$, to partition the sample of children aged 0-6 into 10 randomly selected groups (stratified by caste and by relative wealth within castes to maintain balance). Of the 10 subsamples, a single subsample is retained for testing; i.e. to compute the forecast error, while the model is estimated with the remaining subsamples. This process is repeated 10 times, with each testing subsample, to compute the average forecast error with each specification of the control function. These average forecast errors are reported in Appendix Table A3, where we see that the optimal specification of the control function, which minimizes the forecast error, is linear in wealth and without interactions. Figure 8b reports the relationship between the probability that a child is a girl and relative wealth, with the optimal control function (partialled out) and without the control function. Inclusion of the control function does affect the estimated relationship, judging from the fact that there is little overlap in the 95% confidence intervals for the two specifications. Consistent with this result, we see in Appendix Table A4 that the underlying coefficients of the control function are jointly highly significant (with an F statistic of 17.8) and that the coefficient on household wealth, as in Table 6, is positive and significant.

However, the probability that the child is a girl continues to be decreasing in relative wealth.[42]

3.5 Caste-Specific Evidence

The advantage of including multiple castes in the analysis is that a control function can be included in the estimating equation, which allows us to isolate the association between sex selection and relative wealth within the caste. To independently assess the robustness of our results, we nonparametrically estimate the association between sex selection and relative wealth, caste by caste. The marriage market is organized the same way in all castes, and thus we would expect the predictions of the model to apply to all castes. Figure 9a reports this test for the 0-6 age group for the 12 largest castes in our study area, which account for 82% of the children in this age group. The probability that a child is a girl is decreasing with relative wealth for 9 of the 12 castes. For the three castes that it is not – Balija, Boya, and Naikar – the number of observations is relatively small (less than 2,000 children per caste). It is possible that the anomalous pattern for these castes is simply a consequence of the small sample size, which makes the estimated relationship unstable. To examine this possibility, we report the association between relative wealth and the probability that the child is a girl for the 7-13 year olds in Figure 9b. Reassuringly, the relationship is negative at each point in the wealth distribution for the three castes for which anomalous patterns were observed in Figure 9a.

Two castes, the Vanniyas and the Adi Dravidars (Paraiyars), dominate the population in the study area. The Vanniyas are a relatively wealthy landowning caste. In contrast, the Adi Dravidars lie at the very bottom of the social hierarchy (even among the SC's). Despite their social differences, the probability that a child is a girl is decreasing with wealth within each of these castes. As a final robustness test, we estimate the relationship between sex selection and relative wealth, (i) without the Vanniyas, (ii) without the Vanniyas and the Adi Dravidars, and (iii) with just the 12 largest castes. The estimates with these different samples, reported in Appendix Table A5, are very similar to what we obtain with the full sample of 0-6 year olds in Table 6, Column 1. There is a robust negative association between the probability that a child is a girl and the family's position in its caste's per capita wealth distribution.

4 Structural Estimation and Quantification

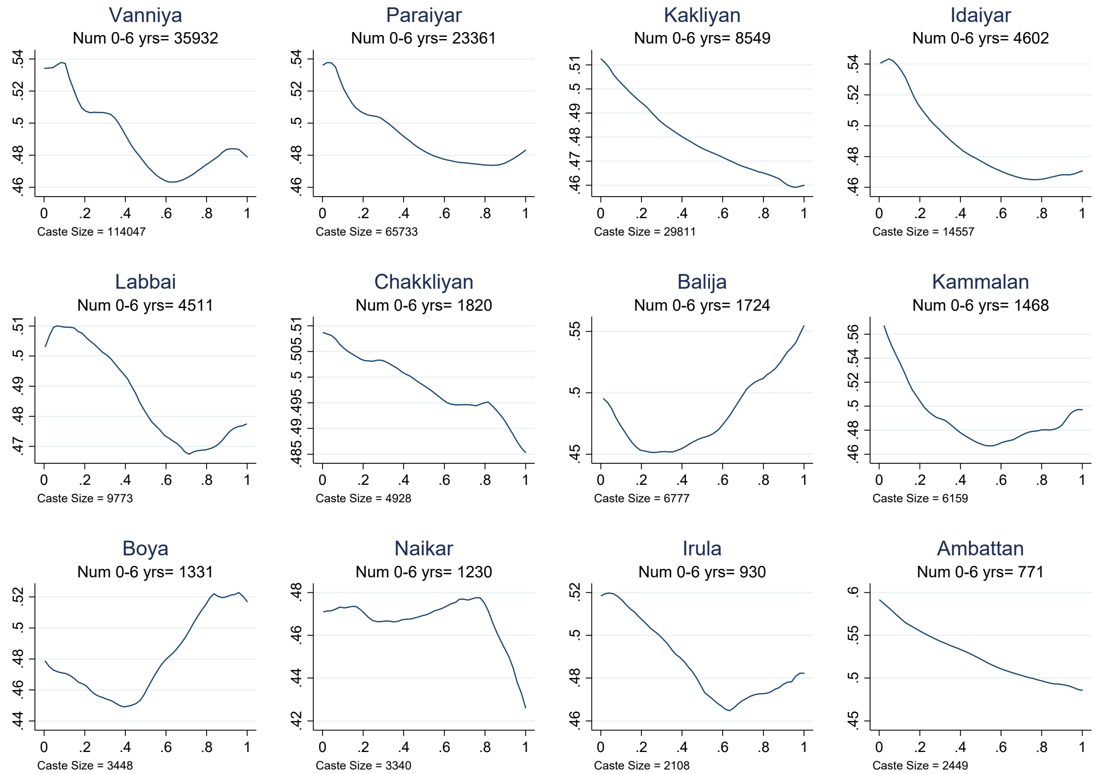
4.1 Magnitude of Within-Caste Variation

To quantify the magnitude of the variation within castes that we have uncovered, we partition the households with children aged 0-6 in each caste into ten equally sized wealth classes. The number of classes is chosen by weighting two competing considerations: The larger the number of wealth classes, the closer we can approximate the corresponding nonparametric plots which describe the sex ratio at each point in the wealth distribution. However, this comes at the cost of less precise estimates of the sex ratio within wealth classes,

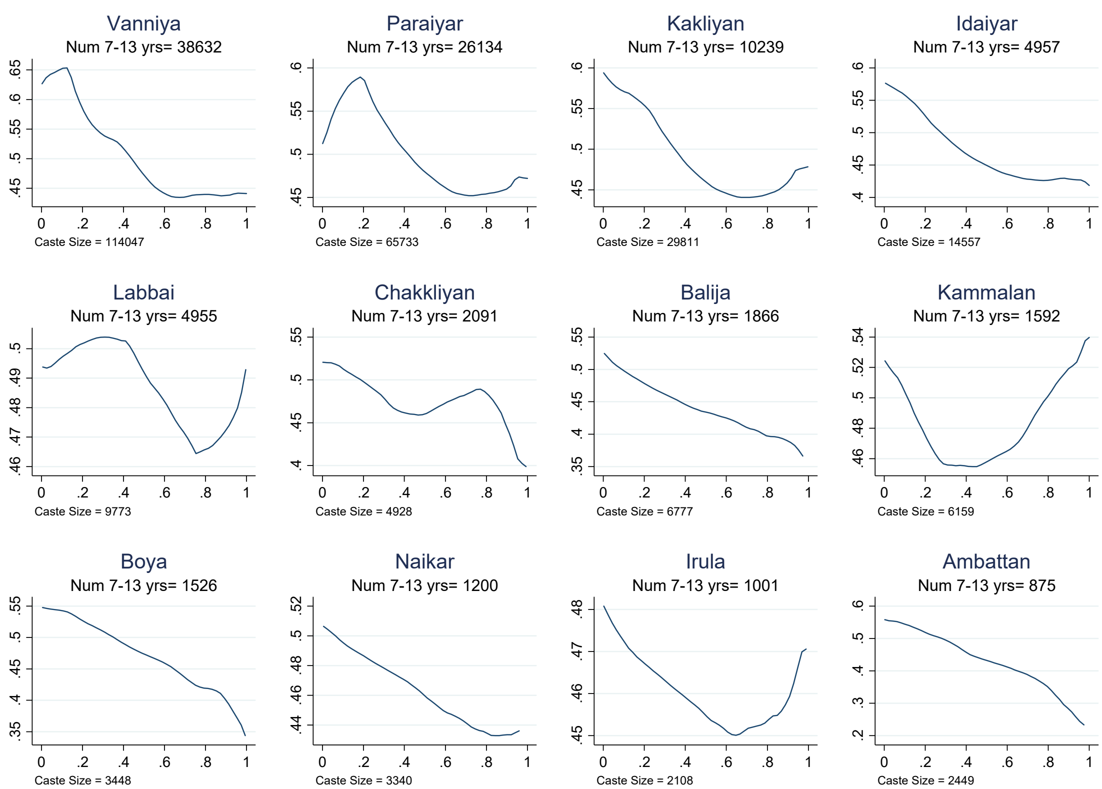
[42] In our model, relatively less wealthy girls are advantaged because they marry up and are on the short side of the marriage market (which improves their family's *ex ante* bargaining position). An alternative interpretation of the relative wealth effect is based on these girls having stronger *ex post* bargaining positions because relatively less wealthy girls marry at a later age or have a smaller age-gap with their spouse. As observed in Appendix Figures A3 and A4, there is no support for this alternative interpretation.

Figure 9: **Sex Selection and Relative Wealth (12 largest castes)**

(a) Ages 0-6



(b) Ages 7-13



especially for castes with just a couple thousand children aged 0-6.

The benchmark equation that we use to quantify the magnitude of the within-caste variation in sex ratios has the fraction of girls in each wealth class in the caste as the dependent variable and a full set of wealth-class dummies as regressors. The R^2 in this regression, which indicates how much of the overall variation in sex ratios can be explained by relative wealth within the caste is 0.11 when the sample is restricted to the 30 largest castes and 0.22 when the sample is restricted even further to the 12 largest castes. To compare the magnitude of the within-caste and between-caste variation in sex ratios, we estimate an augmented equation that incorporates both sources of variation by including caste fixed effects. The R^2 with this specification increases to 0.21 for the sample with 30 castes and 0.30 for the sample with 12 castes (the coefficient estimates for all specifications are reported in Appendix Table A6). This implies that within-caste variation accounts for 52% of the explained variation with the 30-caste sample and as much as 74% of the explained variation with the 12-caste sample. No caste that is known to be traditionally associated with severely biased sex ratios is present in the study area. It is possible that in other districts where such castes are present, the between-caste variation will be more substantial. Nevertheless, these results highlight the importance of the within-caste variation that is the focus of our analysis.

A second approach to quantify the magnitude of the within-caste variation would be to measure the range of sex ratios across the ten wealth classes. Converting the fraction of girls to the number of boys per 100 girls, the sex ratio ranges from 101 to 118. To put these statistics in perspective, the sex ratio derived from the 2011 population census ranges from 115 to 120 in the three worst states in the country. Our estimates thus indicate that the variation in sex ratios within castes is as large as the variation across states in the country. There is nothing unusual about the study area with respect to demographic and socioeconomic characteristics. Sex selection may thus be more serious and more pervasive than is commonly believed, affecting relatively wealthy households (within their caste) throughout the country.

4.2 Structural Estimation and Counter-Factual Simulations

We complete the analysis by estimating the structural parameters of the model. The model can subsequently be used to (i) quantify the contribution of the alternative mechanisms to sex selection, and (ii) to evaluate the effect of existing and counterfactual policies targeting sex selection. The algorithm we used to solve the model numerically for given parameter values was described above. To estimate the parameters, we search over all combinations of the parameters to find the combination for which the predicted fraction of girls across the ten equal-sized wealth classes matches most closely with the actual fractions; i.e. for which the sum of squared errors is minimized. We use the 12 largest castes for the structural estimation. (Block) bootstrapped means and confidence intervals for the parameters are reported in Table 7 by drawing repeated samples (of castes) with replacement.

The parameter associated with boys' bargaining position, β, is equal to 0.77, satisfying the condition in the model that its value should exceed 2/3. The son preference parameter, u_b, is also positive and statistically significant, indicating that both sources of sex selection are present. The parameters associated with the cost to boys from being single, m_b, and the cost of sex selection, a, both have the expected (positive) sign and are

Table 7: **Structural Parameter Estimates**

	Mean [95% Confidence Interval]
Boys' bargaining position: β	0.767 [0.678 , 0.862]
Son Preference: u_0	0.637 [0.082, 1.092]
Boys' cost of being single: m_b	0.183 [0.102, 0.812]
Cost of sex selection: a	24.48 [16.859, 29.235]

Source: SICHS census. Sample restricted to children aged 0-6 in 12 largest castes. Each caste is partitioned into 10 equal sized income classes and the sex ratio is computed within each class. Bootstrapped confidence intervals in brackets.

precisely estimated.

The estimated structural parameters can be used to decompose the contributions of the marriage market channel and the son preference channel to sex selection. The lower surface of the blue area in Figure 10a traces out the predicted fraction of girls, based on the estimated model, for each wealth class. The line dividing the red and blue areas traces out the corresponding fraction of girls when the son preference channel is shut down; i.e. u_b is set to zero. The red area thus measures the contribution of the marriage market channel, while the blue area measures the contribution of the son preference channel, under the assumption that the fraction of girls equals 0.5 without sex selection. Under this decomposition, 52% of sex selection in the study area is due to the marriage market channel, which is close to the figure of 55% that we arrived at above, based on the sex ratios of first-born and all children in Table 3.

Although our analysis provides new evidence on the extent of sex selection and the mechanisms underlying this phenomenon, the problem itself is well known and widely discussed in academic and policy circles. Many states and the central government have responded to the problem by introducing Conditional Cash Transfer schemes with the stated objective of improving the survival and the welfare of girls and reversing the bias in the child sex ratio. Sekher (2010) evaluates 15 such schemes. These schemes have a number of common features. Parents receive a cash transfer when (i) the birth of a female child is registered, (ii) she receives the requisite immunizations, and (iii) she achieves specific educational milestones. In addition, an insurance cover is provided, which matures when the girl turns 18. Although governmental transfers when she is young go directly to her parents, the insurance payment, when it matures, goes into a bank account that is set up for the girl. Most girls in India are unmarried at 18 and are thus dependents in their parents' home when the (large) final transfer is received by the family. It thus seems reasonable to assume that the transfer is appropriated by the girl's parents; this transfer coincides with the time when many girls enter the marriage market and, thus, it can be conveniently used to pay the dowry.

Our model is well suited to examine the impact of these policies since it incorporates the son preference and the marriage market mechanism, and both mechanisms have been seen to contribute substantially to sex selection. The dowry is also determined endogenously, allowing for general equilibrium effects. In the framework of our model, a transfer to the girl's parents is equivalent to an exogenous increase in the wealth of the girl's family. As an additional counter-factual policy exercise, we examine an alternative scheme that provides a transfer to the daughter *after* she is married. We have argued above that the entire dowry payment, including the bequest component, is appropriated by the in-laws and then redistributed. One concern is that government transfers to married women would be similarly appropriated. However, recent evidence from rural India indicates that direct payments into married women's bank accounts can, in fact, be controlled by them (Field et al., 2021). The alternative transfer scheme targeting married women may thus be feasible.

The schemes described above will change the maximized utility of the girl's family in the following ways:

(a) If the wealth transfer, w, is to the girl's parents,

$$v = \log\left(x + y + w - \frac{2e^{\frac{u-u_b}{2}}}{\beta}\right) + \log\left((1-\beta)\frac{2e^{\frac{u-u_b}{2}}}{\beta}\right). \tag{12}$$

(b) If the wealth transfer is directly to the girl,

$$v = \log\left(x + y - \frac{2e^{\frac{u-u_b}{2}}}{\beta}\right) + \log\left((1-\beta)\frac{2e^{\frac{u-u_b}{2}}}{\beta} + w\right). \tag{13}$$

If u is fixed, then the most effective scheme will target the family member; i.e. the girl or her parent, who has a lower level of consumption in equilibrium. However, the effect of the wealth transfer is more complex than that because it will change the equilibrium marriage price and, hence, matching and sex selection over the entire wealth distribution. This is especially important because while some of the welfare schemes are available to all families with girls, many are targeted to families below the poverty line. While the targeted families may be induced to have more girls, there will be spillover effects through the equilibrium marriage market price that could *increase* sex selection at other points in the wealth distribution.

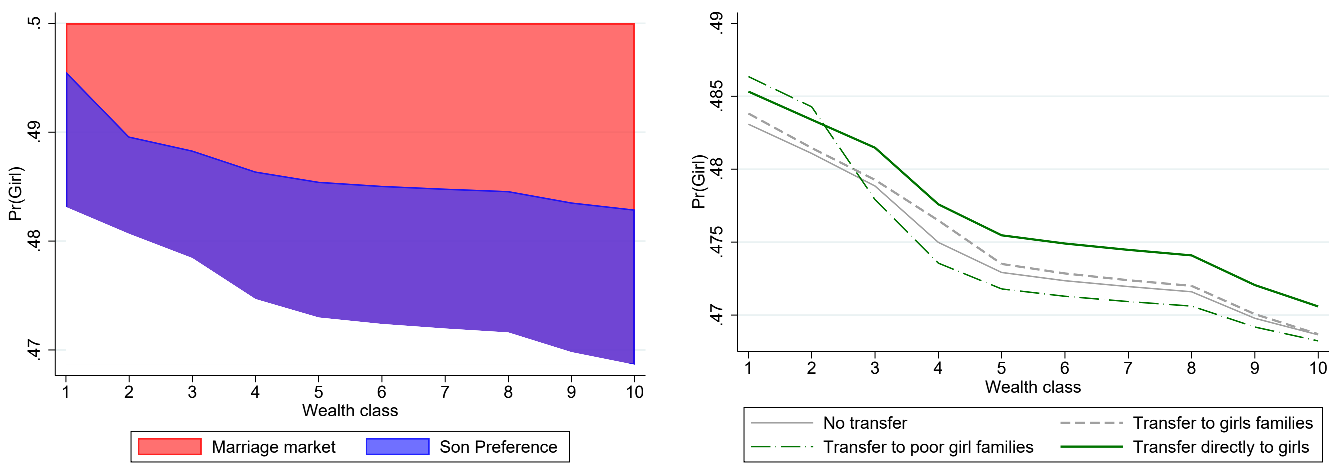
The grey solid line in Figure 10b is the benchmark fraction of girls predicted by the model in each wealth class. The first counter-factual policy experiment that we consider, whose effect is described by the green dashed line, is a 20% wealth transfer to families in the bottom two classes with girls. This experiment is designed to reflect the wealth eligibility requirement in many existing schemes. The fraction of girls increases substantially in each of the two treated wealth classes. This increase in the number of girls at the bottom of the wealth distribution will shift the entire equilibrium price (dowry) schedule and we see in the figure that this results in an increasingly biased sex ratio in the upper eight wealth classes.[43] Combining all wealth classes, the net effect of this scheme is to *worsen* the overall sex ratio. This result would not be obtained with models of sex selection that ignore the pecuniary spillovers within castes, highlighting the value of the market equilibrium analysis for policy evaluation.

[43]The wealth increase of the girls at the bottom pushes up the dowry for them, but also for those who do not receive the subsidy, since they compete for the same boys. The higher equilibrium dowry leads to more sex selection further up the wealth distribution.

The next policy experiment that we consider, whose effect is described by the grey dashed line in Figure 10b, provides the wealth transfer to all girls' parents. To be comparable with the first experiment, the amount of the per family transfer is divided by five (because the beneficiaries are now in 10 rather than 2 of the equal-sized wealth classes). Although the transfer now increases the fraction of girls in each wealth class, the magnitude of the effect is small. This is because part of the subsidy is transferred to boys' families via higher dowries.

Figure 10: **Quantification**

(a) Sex Selection Mechanisms (b) Counterfactual Policy Experiments



The final policy experiment that we consider, whose effect is described by the green solid line, has the most promise. It is the same as the preceding experiment, except that the subsidy goes directly to the adult girls rather than their parents. Note that this transfer should not be given until the girl is married and it cannot be used as a dowry payment. As we can see in Figure 10b, there is now a substantial increase in the fraction of girls in each wealth class.[44] This is because the bequest that must be transferred to the girl through the inefficient dowry mechanism will decline. With the resulting decline in the mismatch between the girl's actual consumption and the preferred level of consumption from her parent's perspective, it is less costly to have a girl. Policies that give resources directly to girls when they are adults, as opposed to their (altruistic) parents when they are children, may thus be especially effective in reducing the bias in child sex ratios in India.

5 Conclusion

Two mechanisms have been proposed to explain the persistently high rates of sex selection in India: *son preference* in which parents desire a male heir and *daughter aversion* in which dowry payments make parents worse off with girls than with boys. The first mechanism has been examined theoretically and empirically.

[44] This policy experiment is conducted holding constant the β parameter. It is possible that the girl's bargaining position will increase when she has direct control of the resources she brings into the marriage. The resulting decline in β will further increase the fraction of girls from Figure 2a.

However, while recent evidence indicates that exogenous increases in dowries worsen the sex ratio, it is not obvious why girls' families who are on one side of the marriage market – buying grooms for a price – should be disadvantaged in equilibrium. The model that we develop in this paper (incorporating both mechanisms) provides micro-foundations, based on the organization of the marriage institution in India, for daughter aversion. We show that sex selection due to daughter aversion will arise if the following conditions are satisfied: (i) The in-laws must be unable to commit to transferring the bequest component of the marriage payment to the girl; (ii) The daughter-in-law must receive a less than equal share of the resources available for consumption in her marital home; (iii) The social norm that all girls marry must be binding, or else girls' parents could avoid the disutility associated with the marriage market channel by leaving their daughters single.

While the preceding conditions may hold in the Indian context, they may not elsewhere, providing an explanation for why daughter aversion is not observed in other countries. Take Bangladesh, for example. Marriage in Bangladesh is patrilocal, dowries have emerged in recent years (Ambrus et al., 2010), and there is pressure on girls to marry (Field and Ambrus, 2008). What makes Bangladesh different is an institution – the *mehr* – that exists in all Islamic societies and which guarantees the woman a share of her husband's family's wealth in the event of a divorce. In the context of our model, the *mehr* improves the outside option for married women, increasing their bargaining position. This might explain why the emergence of dowries in Bangladesh did not increase sex selection (Kabeer et al., 2014), in contrast with what was observed in South India. Next, consider China, where there is a long tradition of son preference and the accompanying deficit of girls goes back many centuries, but where the daughter aversion mechanism appears to be absent. Indeed, increasingly biased sex ratios have been accompanied by a substantial increase in bride prices. Based on our model, we argue that this outcome has emerged in China because (i) the married couple rarely cohabits with the in-laws and, hence, the bequest to the girl and the marriage payment from the boy's parents to the girl's parents can be decoupled (Zhang and Chan, 1999); (ii) most Chinese women work and, hence, the bride is in a position to ensure that she receives the bequest.[45]

Marital matching, sex selection and dowries are jointly determined in our model. We show that under reasonable parametric restrictions, the two co-existing channels generate sex selection at every wealth level within the marriage market, which is defined by the caste in India. In addition, hypergamy (the wealth-gap between grooms and brides) is increasing, and sex selection is decreasing, as we move down the wealth distribution. These implications are verified with unique data from the South India Community Health Study (SICHS) covering a representative sample of rural households. A novel nonparametric (cross validated) control function approach is used to estimate the relative wealth effect, independently of other determinants of sex selection, effectively comparing outcomes for households with the same (absolute) wealth and family size who happen to be located at different positions in their caste's (per capita) wealth distribution.

We complete the analysis by estimating the structural parameters of the model and then conducting counter-factual simulations. Although the overall child sex ratio in the study area is not unusually biased and is comparable to the corresponding statistic for rural South India, we document wide variation in the sex

[45]Based on official ILO statistics, female labor force participation in 1990 and 2017, respectively, was 73% and 61% in China and 35% and 27% in India.

ratio across the wealth distribution within castes (comparable in magnitude to the variation across states in the country). Our first counter-factual analysis indicates that about 52% of this variation in sex ratios can be attributed to the marriage market mechanism, with the remainder accounted for by the son preference mechanism. Our second counter-factual analysis examines the impact of alternative programs targeting sex selection. In recent years, the government has implemented a number of conditional cash transfer programs, rewarding parents if they have a girl, with the stated objective of reducing sex selection. Our analysis indicates that these programs, particularly those restricted to less wealthy households, could actually worsen the problem on net through the general equilibrium effect on marriage prices (dowries). In addition, our analysis indicates that a small, easily implementable, change in current programs – targeting the transfer to married women directly rather than to their (altruistic) parents earlier in life – could substantially increase their impact.

Appendix A Tables and Figures

Table A1: **Comparison of Demographic and Socioeconomic Characteristics**

	Men			Women		
Region	South India	Tamil Nadu	Study Area	South India	Tamil Nadu	Study Area
Panel A						
Age distribution						
married (%) <10Yrs	0.0	0.0	0.0	0.0	0.0	0.0
10-14	0.0	0.0	0.0	0.0	0.0	0.0
15-19	0.2	0.0	0.0	1.5	0.5	0.8
20-24	2.2	1.0	1.0	6.9	5.2	5.9
25-29	5.4	4.4	4.8	8.0	7.9	8.1
30-34	6.7	6.8	6.5	7.2	7.3	6.6
35-39	6.8	6.6	7.1	6.7	6.6	7.7
40-44	6.2	6.3	3.7	5.7	5.9	3.1
45-49	5.5	5.9	6.8	5.0	5.2	5.8
50-54	4.6	5.0	4.8	3.4	3.6	3.5
55-59	3.6	4.1	4.1	3.4	3.8	3.3
60-64	2.9	3.4	3.8	2.1	2.0	1.9
65-69	2.2	2.5	2.4	1.3	1.2	1.1
70-74	1.4	1.5	1.6	0.7	0.5	0.3
75-79	0.8	0.9	0.8	0.3	0.2	0.1
80-84	0.4	0.4	0.4	0.1	0.1	0.0
85>Yrs	0.2	0.2	0.2	0.1	0.1	0.0
Kolmogorov-Smirnov test of equality (p-value)	1.00	1.00	–	1.00	0.96	–
Literacy rate (%)	78.5	82.0	76.9	61.1	65.1	62.4
Labor force participation rate (%, 15-59 years)	79.8	81.1	81.0	44.9	42.6	40.0
Panel B						
Hindu (%)	91.0	92.8	93.8	–	–	–
Muslim (%)	5.3	2.7	1.7	–	–	–
Christian (%)	1.1	3.5	4.5	–	–	–
Other (%)	2.6	1.0	0.0	–	–	–
Sex Ratio	102	101	100	–	–	–
Child Sex Ratio	108	107	109	–	–	–
Population	97,390,694	18,679,065	474,384	95,226,007	18,550,525	475,022

Notes: South India includes rural Maharashtra, Karnataka, Andra Pradesh, and Tamil Nadu. Tamil Nadu refers to rural Tamil Nadu. % married measures the number of married individuals in each age category as a fraction of the total population, seperately for men and women. Literacy rate is defined by the Goverment of India as the percentage of those aged 7+ who can, with understanding, read and write a short, simple statement on their everyday life; SICHS Census definition is those aged 7+ with $\geq$ 1 year of education (figures for $\geq$ 3 years of education are similar, 73.8% for men and 59.5% for women). Labor force participation is defined as the proportion of 15-59 year old persons of the total 15-59 years population who are either employed or seeking or available for employment. Sex Ratio refers to the number of males per 100 females in the population. Child sex ratio is the number of males per 100 females for those aged between 0-6 years.
Sources: For Tamil Nadu and South India, the data on age distribution, literacy rate, religious composition and sex ratios are from the 2011 Census of India. The data on labor force participation is from the Ministry of Labor and Employment, Government of India, 2009-10. For Study Area, all statistics based on SICHS Census.

Table A2: **School Enrollment and Sex Selection**

Dependent variable	Boys higher secondary enrollment	Girls higher secondary enrollment	Girl dummy
Age range	14-17	14-17	0-6
	(1)	(2)	(3)
Mother's education	0.0126***	0.0106***	-0.000422
	(0.000454)	(0.000440)	(0.000382)
Father's education	0.0109***	0.0112***	0.000167
	(0.00049)	(0.000445)	(0.000387)
Mean of dependent variable	0.842	0.838	0.479
Observations	25,303	27,109	91,405
Caste FE	Yes	Yes	Yes

Notes: Higher secondary enrollment indicates whether the child is enrolled in school. The lower bound for the age range is set at 14 because most children in rural Tamil Nadu study till the 8th grade (age 13). The upper bound is set at 17 because girls start to marry (and leave their parental homes) by the age of 18. Sex selection is measured by the probability that the child (aged 0-6) is a girl. *** $p<0.01$. Source: SICHS census.

Table A3: **Mean Forecast Sum of Square Errors**

	Linear (1)	Quadratic (2)	Cubic (3)	Quartic (4)
No interaction	0.249342	0.249352	0.249351	0.249357
Linear interaction	0.249347	0.249356	0.249355	0.249362
Linear + quadratic interaction		0.249366	0.249365	0.249371
Linear + quadratic + cubic interaction			0.249377	0.249384
all interactions				0.249395

Notes: Mean forecast error is based on k-fold cross validation, with k=10. We consider 14 specifications of the control function: linear, quadratic, cubic and quartic functions of wealth, with varying degrees of interaction between the wealth terms and the family size dummies.

Table A4: **Control Function Parameter Estimates**

Dependent variable:	Girl dummy
Family size = 3	-0.014
	(0.019)
Family size = 4	-0.044***
	(0.009)
Household wealth	0.015**
	(0.007)
Joint $F-$statistic	17.83
	[0.000]
Observations	79,027
Caste FE	Yes

Notes: The Robinson (1988) procedure is implemented in three steps. (1) The dependent variable (girl dummy), each argument in the control function and the caste dummies are regressed nonparametrically on relative wealth. (2) The residual from the first regression with the girl dummy as the outcome is regressed on the other residuals, without a constant term, to obtain consistent estimates of the control function parameters (reported above) and the caste fixed effects. (3) The girl dummy minus the control function and caste dummies (adding back their means) is regressed nonparametrically on relative wealth, as reported in Figure 8.

Table A5: **Alternative Samples**

Dependent variable	Girl dummy (0-6 years)			
Sample	Benchmark	Dropping biggest caste	Dropping 2 biggest castes	Dropping smaller castes
	(1)	(2)	(3)	(4)
Rank in caste per capita wealth distribution	-0.0457***	-0.0408***	-0.0386***	-0.0541***
	(0.0068)	(0.0068)	(0.0104)	(0.0071)
Sample mean of dependent variable	0.480	0.482	0.481	0.480
Observations	79,027	49,522	29,883	69,233
Caste FE	Yes	Yes	Yes	Yes

Source: SICHS census. Bootstrapped standard errors (in parentheses) are clustered at the *panchayat* level.
*** p<0.01

Table A6: **Within Caste and Between Caste Variation in Sex Ratios**

Dependent variable	Fraction of girls			
Sample	12 castes		30 castes	
	(1)	(2)	(3)	(4)
Wealth class				
1	0.00979	0.00965	0.0139	0.0136
	(0.00840)	(0.00790)	(0.00946)	(0.00893)
2	0.0157*	0.0160**	0.0141*	0.0143**
	(0.00814)	(0.00745)	(0.00776)	(0.00721)
3	0.0104	0.0104	0.00732	0.00727
	(0.00820)	(0.00730)	(0.00770)	(0.00687)
4	0.00732	0.00732	0.00353	0.00362
	(0.00833)	(0.00781)	(0.00828)	(0.00790)
5	-0.00348	-0.00340	-0.00601	-0.00585
	(0.00912)	(0.00806)	(0.00866)	(0.00775)
6	-0.00599	-0.00611	-0.00695	-0.00713
	(0.00872)	(0.00780)	(0.00835)	(0.00747)
7	-0.0223**	-0.0223**	-0.0237**	-0.0235**
	(0.00998)	(0.0106)	(0.00935)	(0.00986)
8	-0.00903	-0.00914	-0.00970	-0.00983
	(0.00856)	(0.00729)	(0.00818)	(0.00699)
9	0.00528	0.00521	0.00122	0.00126
	(0.00711)	(0.00712)	(0.00716)	(0.00724)
Constant (Wealth class 10)	0.481***	0.481***	0.483***	0.483***
	(0.00654)	(0.00566)	(0.00626)	(0.00545)
Observations	118	118	298	298
R-squared	0.219	0.297	0.108	0.208
Caste FE	No	Yes	No	Yes

Source: SICHS census. Sample restricted to children aged 0-6 years. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05

Figure A1: **Age Difference Between Husband and Wife by Birth Cohorts**

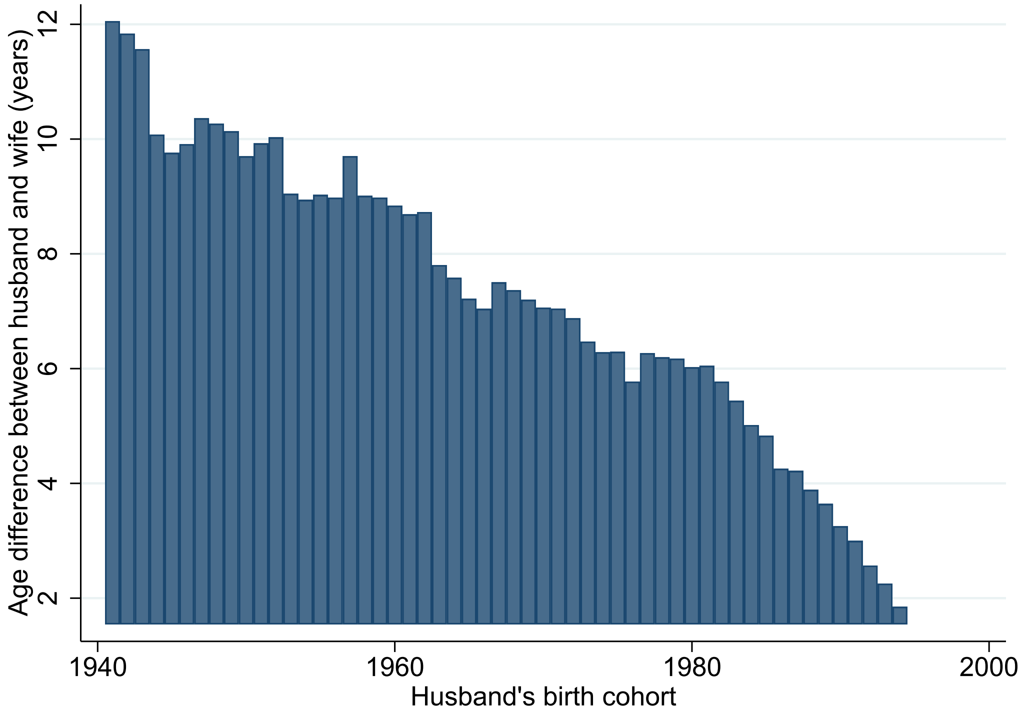


Figure A2: **Proportion Married by Age**

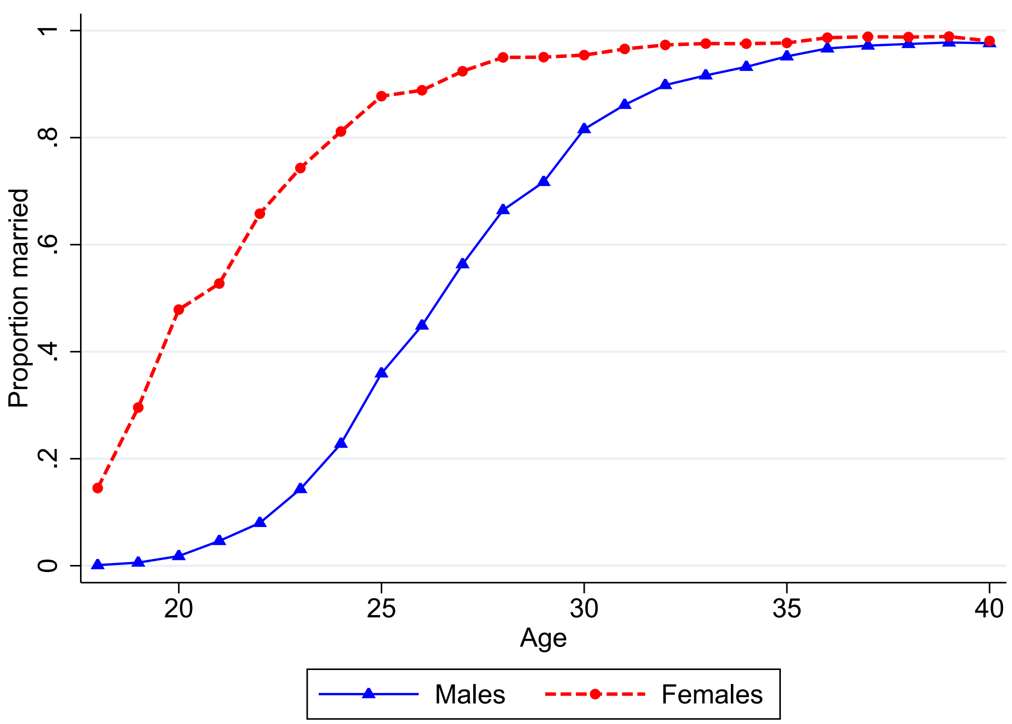


Figure A3: **Variation in Age at Marriage by Relative Wealth**

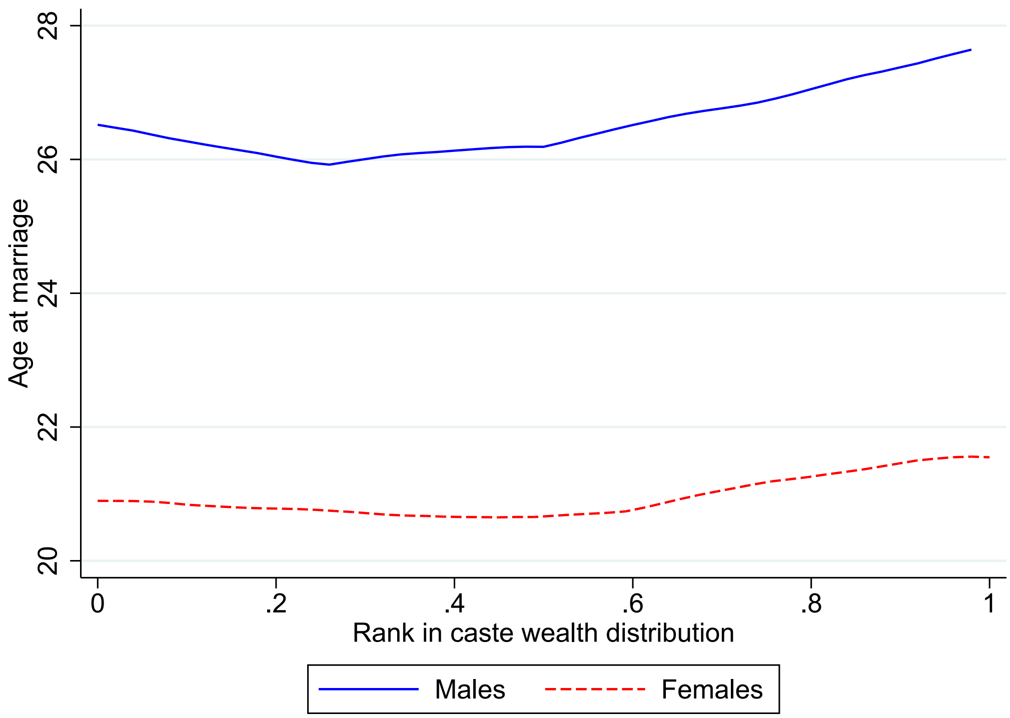
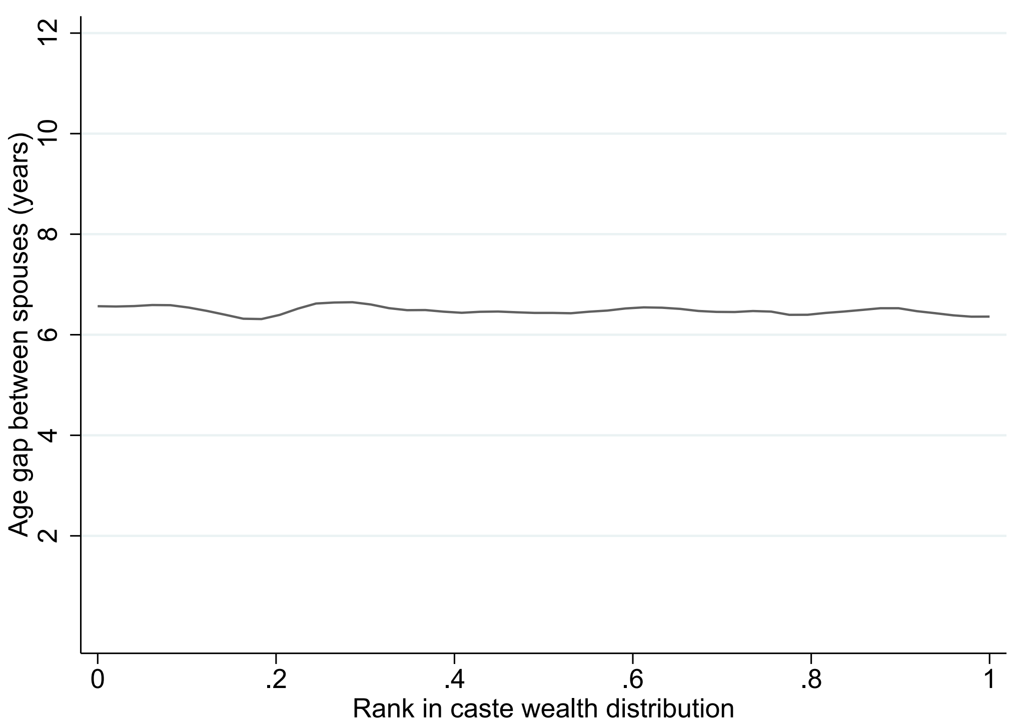


Figure A4: **Variation in Spousal Age Gap by Relative Wealth**



Appendix B Omitted Proofs

B.1 Proof of Proposition 1

Proof. Equation (6) describes the first order condition associated with the girl's family's utility maximization problem:

$$v_x + v_u u' = 0 \quad \Rightarrow \quad u' = -\frac{v_x}{v_u}. \tag{14}$$

Then the surplus is supermodular and the allocation will be PAM (see Legros and Newman (2007) and Chade et al. (2017)) provided:

$$\frac{\partial^2 v(x, y, u)}{\partial x \partial y} = v_{xy} + v_{uy} u' > 0. \tag{15}$$

From equation (5), we can write the equilibrium utility of the girl's family as:

$$v = \log\left(x + y - \frac{2e^{\frac{u-u_b}{2}}}{\beta}\right) + \frac{u - u_b}{2} + \log\left(\frac{2(1-\beta)}{\beta}\right). \tag{16}$$

Next, to derive the condition for PAM in inequality (15), we derive the following terms:

$$v_x = \frac{1}{x + y - \frac{2e^{\frac{u-u_b}{2}}}{\beta}} \tag{17}$$

$$v_{xy} = -\left(x + y - \frac{2e^{\frac{u-u_b}{2}}}{\beta}\right)^{-2} \tag{18}$$

$$v_u = \frac{-\frac{e^{\frac{u-u_b}{2}}}{\beta}}{x + y - \frac{2e^{\frac{u-u_b}{2}}}{\beta}} + \frac{1}{2} \tag{19}$$

$$v_{uy} = -\left(x + y - \frac{2e^{\frac{u-u_b}{2}}}{\beta}\right)^{-2}\left(-\frac{e^{\frac{u-u_b}{2}}}{\beta}\right). \tag{20}$$

Inequality (15) is equivalent to:

$$v_{xy} > \frac{v_x}{v_u} v_{uy} \tag{21}$$

$$-\left(x+y-\frac{2e^{\frac{u-u_b}{2}}}{\beta}\right)^{-2} > \frac{\frac{1}{x+y-\frac{2e^{\frac{u-u_b}{2}}}{\beta}}}{\frac{-\frac{e^{\frac{u-u_b}{2}}}{\beta}}{x+y-\frac{2e^{\frac{u-u_b}{2}}}{\beta}}+\frac{1}{2}}\left[-\left(x+y-\frac{2e^{\frac{u-u_b}{2}}}{\beta}\right)^{-2}\left(-\frac{e^{\frac{u-u_b}{2}}}{\beta}\right)\right] \tag{22}$$

$$-1 > \frac{1}{\frac{-e^{\frac{u-u_b}{2}}}{\beta}+\frac{1}{2}\left(x+y-\frac{2e^{\frac{u-u_b}{2}}}{\beta}\right)}\left(\frac{e^{\frac{u-u_b}{2}}}{\beta}\right) \tag{23}$$

$$-1 > \frac{\frac{e^{\frac{u-u_b}{2}}}{\beta}}{\frac{1}{2}\left(x+y-\frac{4e^{\frac{u-u_b}{2}}}{\beta}\right)}. \tag{24}$$

If $x+y-\frac{4e^{\frac{u-u_b}{2}}}{\beta} < 0$ then condition (24) implies that:

$$x+y-\frac{2e^{\frac{u-u_b}{2}}}{\beta} > 0, \tag{25}$$

which is always satisfied since the LHS is equal to $y-d$, the consumption of the girl's parent, which is constrained to be positive.

$$x+y-\frac{4e^{\frac{u-u_b}{2}}}{\beta} < 0 \quad \iff \quad \frac{u-u_b}{2} > \log\frac{\beta}{4}(x+y). \tag{26}$$

The utility from marriage must exceed the outside option of remaining single:

$$u \geq u_b + 2\log\frac{x}{2} - m_b \quad \iff \quad \frac{u-u_b}{2} \geq \log\frac{x}{2} - \frac{m_b}{2}. \tag{27}$$

When $y=x$, given (27) holds, equation (26) is satisfied provided $\log\left(\frac{\beta}{4}(2x)\right) < \log\frac{x}{2} - \frac{m_b}{2}$ or equivalently, $2\log\beta < -m_b$. Whenever $y<x$, this sufficient condition is satisfied as well. This establishes the proof. ■

B.2 Proof of Proposition 2

Proof. 1. At the top of the wealth distribution, $\overline{y} = \overline{x}$. When the dowry $d = 0$,

$$u(\overline{y}) - v(\overline{y}) = u_b + 2\log\left(\frac{\beta}{2}\overline{y}\right) - \log\left((1-\beta)\overline{y}\right) - \log(\overline{y}) \tag{28}$$

$$= u_b + 2\log\left(\frac{\beta}{2}\right) - \log(1-\beta). \tag{29}$$

It follows that:

$$u(\overline{y}) - v(\overline{y}) > 0 \quad \Longleftrightarrow \quad u_b > \log(1-\beta) - 2\log\left(\frac{\beta}{2}\right). \tag{30}$$

The right hand side of the second inequality is decreasing in β. Thus, there exists $\underline{u_b}(\beta)$, $\underline{u_b}'(\beta) < 0$, such that $u(\overline{y}) - v(\overline{y}) > 0$ if $u_b > \underline{u_b}(\beta)$. $u_b > 0$, $\beta \in (2/3, 1)$. For $\beta = 2/3$, $\underline{u_b} = -\log(1/3)$. For $\beta \to 1$, $\underline{u_b} \to 0$.

Next, we show that $u(\overline{y}) - v(\overline{y})$ is increasing in the dowry d:

$$u(\overline{y}) - v(\overline{y}) = u_b + 2\log\left(\frac{\beta}{2}(\overline{y}+d)\right) - \log\left((1-\beta)(\overline{y}+d)\right) - \log(\overline{y}-d) \tag{31}$$

$$\frac{\partial}{\partial d}\Big(u(\overline{y}) - v(\overline{y})\Big) = \frac{2}{\overline{y}+d} - \frac{1}{\overline{y}+d} + \frac{1}{\overline{y}-d} = \frac{1}{\overline{y}+d} + \frac{1}{\overline{y}-d} > 0. \tag{32}$$

Assuming that the condition in Lemma 1 is satisfied, $d > 0$ and, hence, $u(\overline{y}) - v(\overline{y}) > 0$ for any value of the equilibrium dowry. There is sex selection at the top of the wealth distribution.

2. At the bottom of the wealth distribution, girls' families with wealth $\underline{y}$ match with boys' families with wealth x^*. The last boy to match is indifferent between marrying and staying single:

$$u_b + 2\log\left(\frac{\beta}{2}(x^*+d)\right) = u_b + 2\log\left(\frac{x^*}{2}\right) - m_b. \tag{33}$$

Denote $M_b \equiv \exp(m_b/2) > 1$.

Then we can solve for the equilibrium dowry $d(x^*)$ received by the last boy to marry:

$$d(x^*) = \left(\frac{1 - M_b\beta}{M_b\beta}\right)x^* \tag{34}$$

$M_b\beta < 1$ from Lemma 1 and, hence, $d(x^*)$ is increasing in x^*.

Next, we show that $u(\underline{y}) - v(\underline{y}) > 0$ for $x^* = \underline{x} = \underline{y}$. If the family with wealth $y = \underline{y}$ chooses to have a boy instead of a girl, he will either be unmatched or the last boy to match. Either way, its utility will be

$u(\underline{y}) = u_b + 2\log\frac{\underline{y}}{2} - m_b$.

$$u(\underline{y}) - v(\underline{y}) = u_b + 2\log\frac{\underline{y}}{2} - m_b - \log\left((1-\beta)(\underline{x} + d(\underline{x}))\right) - \log\left(\underline{y} - d(\underline{x})\right) \tag{35}$$

$$= u_b + 2\log\left(\frac{\underline{y}}{2}\right) - m_b - \log\left(\frac{(1-\beta)\underline{y}}{M_b\beta}\right) - \log\left(\frac{(2M_b\beta - 1)\underline{y}}{M_b\beta}\right) \tag{36}$$

$$= u_b - 2\log 2 - \log\left(\frac{(1-\beta)(2M_b\beta - 1)}{\beta^2}\right) \tag{37}$$

It follows that:

$$u(\underline{y}) - v(\underline{y}) > 0 \quad \Longleftrightarrow \quad u_b > \log(1-\beta) - 2\log\left(\frac{\beta}{2}\right) + \log(2M_b\beta - 1). \tag{38}$$

This is the same condition as at the top of the wealth distribution, except for the $\log(2M_b\beta - 1)$ term. From Lemma 1, $M_b\beta < 1$ and, hence, this term is negative. If $u(\overline{y}) - v(\overline{y}) > 0$, then $u(\underline{y}) - v(\underline{y}) > 0$ for $x^* = \underline{x}$.

Next, we show that $u(\underline{y}) - v(\underline{y})$ is decreasing in $x^\star$ for $x^* \leq \underline{x}^*$, such that $d(\underline{x}^*) = \underline{y}/2$, and increasing in $x^\star$ for $x^* > \underline{x}^*$

$$u(\underline{y}) - v(\underline{y}) = u_b + 2\log\left(\frac{\underline{y}}{2}\right) - m_b - \log\left((1-\beta)\left(x^* + d(x^*)\right)\right) - \log\left(\underline{y} - d(x^*)\right) \tag{39}$$

Substituting the expression for $d(x^*)$ from equation (34),

$$u(\underline{y}) - v(\underline{y}) = u_b + 2\log\left(\frac{\underline{y}}{2}\right) - m_b - \log\left(\frac{(1-\beta)x^*}{M_b\beta}\right) - \log\left(\underline{y} - \left(\frac{1 - M_b\beta}{M_b\beta}\right)x^*\right). \tag{40}$$

Therefore:

$$\frac{\partial\left(u(\underline{y}) - v(\underline{y})\right)}{\partial x^\star} = \frac{-1}{x^*} + \frac{1}{\underline{y} - \left(\frac{1-M_b\beta}{M_b\beta}\right)x^*}\left(\frac{1 - M_b\beta}{M_b\beta}\right) \tag{41}$$

$$= \frac{-\underline{y} + 2\left(\frac{1-M_b\beta}{M_b\beta}\right)x^*}{x^*\left(\underline{y} - \left(\frac{1-M_b\beta}{M_b\beta}\right)x^*\right)} \tag{42}$$

$$= \frac{2d(x^*) - \underline{y}}{x^*\left(\underline{y} - d(x^*)\right)}. \tag{43}$$

The term in the denominator is positive because $\underline{y} - d(x^*)$, the consumption of the girl's parent, is constrained to be positive. The term in the numerator is negative if $d(x^*) \leq \underline{y}/2$ and positive if $d(x^*) > \underline{y}/2$. To complete the proof, we thus need to establish that $u(\underline{y}) - v(\underline{y}) > 0$ at x^* such that $d(x^*) = \underline{y}/2$. From the expression for $d(x^*)$, the corresponding value of x^* is $\frac{\underline{y}}{2}\frac{M_b\beta}{1-M_b\beta}$.

$$u(\underline{y}) - v(\underline{y}) = u_b + 2\log\left(\frac{\underline{y}}{2}\right) - m_b - \log\left((1-\beta)\left(\frac{\underline{y}}{2}\frac{M_b\beta}{1-M_b\beta} + \frac{\underline{y}}{2}\right)\right) - \log\left(\underline{y} - \frac{\underline{y}}{2}\right) \tag{44}$$

$$= u_b + 2\log\left(\frac{\underline{y}}{2}\right) - m_b - \log\left(\frac{1-\beta}{1-M_b\beta}\right) - 2\log\left(\frac{\underline{y}}{2}\right) \tag{45}$$

$$= u_b - 2\log M_b - \log\left(\frac{1-\beta}{1-M_b\beta}\right). \tag{46}$$

$$u(\underline{y}) - v(\underline{y}) > 0 \quad \iff \quad u_b > \log(1-\beta) - \log(1-M_b\beta) + 2\log M_b. \tag{47}$$

This condition will be satisfied for $M_b \to 1$ since $u_b > 0$ and, hence, for some $\overline{m}_b$ such that $m_b < \overline{m}_b$.

■

B.3 Proof of Proposition 3

Proof. The extent of sex selection is given by $k^\star(y)$:

$$k^\star(y) = u(y) - v(\mu(y), y, u(\mu(y))). \tag{48}$$

We need to show that $k^\star(y)$ is increasing in y or

$$k^{\star\prime}(y) = u'(y) - \left(v_x\mu' + v_y + v_u u'\mu'\right) > 0. \tag{49}$$

From the first order condition (6), along the equilibrium matching $\mu(y)$, it must be that $v_x + v_u u' = 0$, so the derivative can be written as:

$$k^{\star\prime}(y) = u'(y) - \left((v_x + v_u u')\mu' + v_y\right) \tag{50}$$

$$= u'(y) - v_y(\mu, y, u(\mu)). \tag{51}$$

This is increasing provided:

$$\frac{-2}{y + \mu(y) - \frac{4e^{\frac{u(\mu(y))-u_b}{2}}}{\beta}} - \frac{1}{y + \mu(y) - \frac{2e^{\frac{u(\mu(y))-u_b}{2}}}{\beta}} > 0. \tag{52}$$

To derive the preceding inequality, we note that $u' = -\frac{v_x}{v_u}$ from the First-Order Condition (14) and that expressions for v_x and v_u can be obtained from equations (17) and (19). The expression for v_y is obtained by partially differentiating expression (5).

1. At the top of the wealth distribution. At $y = \overline{y}$, under positive sorting we have $\overline{y} = \mu(\overline{y}) = \overline{x}$. Then condition (52) can be written as:

$$\frac{-2}{2\overline{y} - \frac{4e^{\frac{u(\overline{y})-u_b}{2}}}{\beta}} - \frac{1}{2\overline{y} - \frac{2e^{\frac{u(\overline{y})-u_b}{2}}}{\beta}} > 0 \tag{53}$$

$u(\overline{y}) = u_b + 2\log\left(\frac{\beta}{2}(\overline{y} + d)\right)$. Substituting the expression for $u(\overline{y})$ in equation (53) and rearranging, we obtain:

$$\frac{1}{d} - \frac{1}{\overline{y} - d} > 0. \tag{54}$$

$\overline{y} - d > 0$ to satisfy the constraint that the girl's parent's consumption must be positive and, hence, the preceding condition will be satisfied if $d < \overline{y}/2$.

2. At the bottom of the wealth distribution. At $y = \underline{y}$, $\mu(\underline{y}) = x^\star$. As noted above, if the family chooses to have a boy instead of a girl, he will either remain unmatched or be the last boy to match and, hence, its utility will be $u(\underline{y}) = u_b + 2\log\frac{\underline{y}}{2} - m_b$. Therefore $u'(\underline{y}) = \frac{2}{\underline{y}}$. At an income level $\underline{y}$, we can then write

condition (52), noting that the first term is $u'(y)$ from (51) as:

$$\frac{2}{\underline{y}} - \frac{1}{x^\star + \underline{y} - \frac{2e^{\frac{u(x^*)-u_b}{2}}}{\beta}} > 0. \tag{55}$$

Substituting the expression for $u(x^*)$ in equation (55) and rearranging, we obtain:

$$\frac{2}{\underline{y}} - \frac{1}{\underline{y} - d(x^*)} > 0. \tag{56}$$

$\underline{y} - d(x^*) > 0$ to satisfy the constraint that the girl's parent's consumption must be positive and, hence, it is straightforward to verify that the preceding condition will be satisfied if $d(x^*) < \underline{y}/2$. ■

References

Alfano, M. (2017): "Daughters, dowries, deliveries: The effect of marital payments on fertility choices in India," *Journal of Development Economics*, 125, 89–104.

Ambrus, A., E. Field, and M. Torero (2010): "Muslim family law, prenuptial agreements, and the emergence of dowry in Bangladesh," *The Quarterly Journal of Economics*, 125, 1349–1397.

Anderson, S. (2003): "Why Dowry Payments Declined with Modernization in Europe but are Rising in India," *Journal of Political Economy*, 111, 269–310.

——— (2007): "The Economics of Dowry and Brideprice," *The Journal of Economic Perspectives*, 21, 151–174.

Anderson, S. and C. Bidner (2015): "Property Rights over Marital Transfers," *The Quarterly Journal of Economics*, 130, 1421–1484.

Anukriti, S. (2018): "Financial incentives and the fertility-sex ratio trade-off," *American Economic Journal: Applied Economics*, 10, 27–57.

Anukriti, S., S. Kwon, and N. Prakash (2022): "Saving for dowry: Evidence from rural India," *Journal of Development Economics*, 154, 102750.

Arnold, F., M. K. Choe, and T. K. Roy (1998): "Son Preference, the Family-Building Process and Child Mortality in India," *Population Studies*, 52, 301–315.

Arnold, F., S. Kishor, and T. Roy (2002): "Sex-Selective Abortions in India," *Population and Development Review*, 28, 759–785.

Ashraf, N., N. Bau, N. Nunn, and A. Voena (2020): "Bride price and female education," *Journal of Political Economy*, 128, 591–641.

Attanasio, O. P. and G. Weber (1993): "Consumption Growth, the Interest Rate and Aggregation," *The Review of Economic Studies*, 60, 631–649.

Basu, A. M. (1999): "Fertility Decline and Increasing Gender Imbalance in India, Including a Possible South Indian Turnaround," *Development and Change*, 30, 237–263.

Becker, G. S. (1973): "A theory of Marriage: Part I," *Journal of Political Economy*, 81, 813–846.

Bhalotra, S., A. Chakravarty, and S. Gulesci (2020): "The price of gold: Dowry and death in India," *Journal of Development Economics*, 143, 102413.

Bhalotra, S., A. Chakravarty, D. Mookherjee, and F. J. Pino (2019): "Property rights and gender bias: Evidence from land reform in West Bengal," *American Economic Journal: Applied Economics*, 11, 205–37.

Bhalotra, S. and T. Cochrane (2010): "Where Have All the Young Girls Gone? Identification of Sex Selection in India," *IZA Discussion Paper No. 5381*.

Bhaskar, V. (2011): "Sex Selection and Gender Balance," *American Economic Journal: Microeconomics*, 3, 252–253.

Bhat, P. M. and S. S. Halli (1999): "Demography of Brideprice and Dowry: Causes and Consequences of the Indian Marriage Squeeze," *Population Studies*, 53, 129–148.

Bloch, F. and V. Rao (2002): "Terror as a bargaining instrument: A case study of dowry violence in rural India," *American Economic Review*, 92, 1029–1043.

Blundell, R., M. Browning, and C. Meghir (1994): "Consumer Demand and the Life-Cycle Allocation of Household Expenditures," *The Review of Economic Studies*, 61, 57–80.

Botticini, M. (1999): "A Loveless Economy? Intergenerational Altruism and the Marriage Market in a Tuscan Town, 1415–1436," *The Journal of Economic History*, 59, 104–121.

Botticini, M. and A. Siow (2003): "Why Dowries?" *American Economic Review*, 93, 1385–1398.

Caldwell, J. C., P. H. Reddy, and P. Caldwell (1983): "The Causes of Marriage Change in South India," *Population Studies*, 37, 343–361.

Calvi, R., A. Keskar, et al. (2021): "'Til Dowry Do Us Part: Bargaining and Violence in Indian Families," Tech. rep., CEPR Discussion Papers.

Chade, H., J. Eeckhout, and L. Smith (2017): "Sorting through Search and Matching Models in Economics," *Journal of Economic Literature*, 55, 493–544.

Chiplunkar, G. and J. Weaver (2021): "Marriage markets and the rise of dowry in India," *Available at SSRN 3590730*.

Choo, E. and A. Siow (2006): "Who Marries Whom and Why," *Journal of Political Economy*, 114, 175–201.

Corno, L., N. Hildebrandt, and A. Voena (2020): "Age of marriage, weather shocks, and the direction of marriage payments," *Econometrica*, 88, 879–915.

Das, M., W. K. Newey, and F. Vella (2003): "Nonparametric estimation of sample selection models," *The Review of Economic Studies*, 70, 33–58.

Desai, S. and L. Andrist (2010): "Gender scripts and age at marriage in India," *Demography*, 47, 667–687.

Doepke, M. and M. Tertilt (2009): "Women's Liberation: What's in it for Men?" *The Quarterly Journal of Economics*, 124, 1541–1591.

Dyson, T. and M. Moore (1983): "On Kinship Structure, Female Autonomy, and Demographic Behavior in India," *Population and Development Review*, 35–60.

Edlund, L. (1999): "Son Preference, Sex Ratios, and Marriage Patterns," *Journal of Political Economy*, 107, 1275–1304.

Field, E. and A. Ambrus (2008): "Early Marriage, Age of Menarche, and Female Schooling Attainment in Bangladesh," *Journal of Political Economy*, 116, 881–930.

Field, E., R. Pande, N. Rigol, S. Schaner, and C. Troyer Moore (2021): "On Her Own Account: How Strengthening Women's Financial Control Impacts Labor Supply and Gender Norms," *American Economic Review*, 111, 2342–75.

Foster, A. and M. Rosenzweig (2001): "Missing Women, the Marriage Market and Economic Growth," Brown mimeo.

Geddes, R. and D. Lueck (2002): "The Gains from Self-Ownership and the Expansion of Women's Rights," *The American Economic Review*, 92, 1079–1092.

George, S., R. Abel, and B. D. Miller (1992): "Female Infanticide in Rural South India," *Economic and Political Weekly*, 1153–1156.

Greenwood, J., N. Guner, and G. Vandenbroucke (2017): "Family Economics Writ Large," *Journal of Economic Literature*, 55, 1346–1434.

Guilmoto, C. Z. (2009): "The Sex Ratio Transition in Asia," *Population and Development Review*, 35, 519–549.

Gupta, B. (2014): "Where have all the brides gone? Son preference and marriage in India over the twentieth century," *The Economic History Review*, 67, 1–24.

Heckman, J. and S. Navarro-Lozano (2004): "Using matching, instrumental variables, and control functions to estimate economic choice models," *Review of Economics and Statistics*, 86, 30–57.

Heckman, J. J. (1976): "Simultaneous equation models with both continuous and discrete endogenous variables with and without structural shift in the equations," *Studies in Nonlinear Estimation, Ballinger.*

Hussam, R. N. (2021): "Sex Selection and the Indian Marriage Market," .

Jayachandran, S. (2017): "Fertility Decline and Missing Women," *American Economic Journal: Applied Economics*, 9, 118–139.

Jayachandran, S. and I. Kuziemko (2011): "Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India," *The Quarterly Journal of Economics*, 126, 1485–1538.

Jayachandran, S. and R. Pande (2017): "Why are Indian children so short? The role of birth order and son preference," *American Economic Review*, 107, 2600–2629.

Jeffery, R., P. Jeffery, and A. Lyon (1984): "Female Infanticide and Amniocentesis," *Social Science & Medicine*, 19, 1207–1212.

Jejeebhoy, S. J. and Z. A. Sathar (2001): "Women's Autonomy in India and Pakistan: the Influence of Religion and Region," *Population and Development Review*, 27, 687–712.

Kabeer, N., L. Huq, and S. Mahmud (2014): "Diverging stories of "missing women" in South Asia: Is son preference weakening in Bangladesh?" *Feminist Economics*, 20, 138–163.

Kapadia, K. (1993): "Marrying Money: Changing Preference and Practice in Tamil Marriage," *Contributions to Indian Sociology*, 27, 25–51.

——— (1995): *Siva and her Sisters: Gender, Caste, and Class in Rural South India.*, Westview Press, Inc.

Legros, P. and A. F. Newman (2007): "Beauty is a Beast, Frog is a Prince: Assortative Matching with Nontransferabilities," *Econometrica*, 75, 1073–1102.

Lindenlaub, I. (2017): "Sorting multidimensional types: Theory and application," *Review of Economic Studies*, 84, 718–789.

Moorjani, P., K. Thangaraj, N. Patterson, M. Lipson, P.-R. Loh, P. Govindaraj, B. Berger, D. Reich, and L. Singh (2013): "Genetic Evidence for Recent Population Mixture in India," *The American Journal of Human Genetics*, 93, 422–438.

Munshi, K. and M. Rosenzweig (2016): "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *The American Economic Review*, 106, 46–98.

Murthi, M., A.-C. Guio, and J. Drèze (1995): "Mortality, fertility, and gender bias in India: A district-level analysis," *Population and Development Review*, 745–782.

Neelakantan, U. and M. Tertilt (2008): "A Note on Marriage Market Clearing," *Economics Letters*, 101, 103–105.

Prasad, B. D. (1994): "Dowry-related violence: A content analysis of news in selected newspapers," *Journal of Comparative Family Studies*, 71–89.

Rahman, L. and V. Rao (2004): "The Determinants of Gender Equity in India: Examining Dyson and Moore's Thesis with New Data," *Population and Development Review*, 30, 239–268.

Rao, V. (1993a): "Dowry 'Inflation'in Rural India: A Statistical Investigation," *Population Studies*, 47, 283–293.

——— (1993b): "The Rising Price of Husbands: a Hedonic Analysis of Dowry Increases in Rural India," *Journal of Political Economy*, 101, 666–677.

Robinson, P. M. (1988): "Root-N-consistent semiparametric regression," *Econometrica*, 931–954.

Rosenzweig, M. R. and T. P. Schultz (1982): "Market Opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in Rural India," *The American Economic Review*, 72, 803–815.

Sekher, T. (2010): "Special Financial Incentive Schemes for the Girl Child in India: a Review of Select Schemes," *United Nations Population Fund.*

Sen, A. (1990): "More than 100 Million Women are Missing," *The New York Review of Books.*

Sharma, U. (1980): *Women, Work and Property in North-West India*, Routledge Kegan & Paul.

Srinivas, M. N. (1984): *Some Reflections on Dowry*, Published for the Centre for Women's Development Studies, New Delhi, by Oxford University Press.

——— (1989): *The Cohesive Role of Sanskritization and other essays*, Oxford University Press, USA.

Tertilt, M. (2005): "Polygyny, Fertility, and Savings," *Journal of Political Economy*, 113, 1341–1371.

Zhang, J. and W. Chan (1999): "Dowry and Wife's Welfare: A Theotrical and Empirical Analysis," *Journal of Political Economy*, 107, 786–808.