Wealth, Marriage, and Sex Selection*

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

February 5, 2018

Abstract

Sex selection continues to be a serious problem in India, despite many decades of economic progress. The first contribution of our research is to document substantial variation in child sex ratios in the cross-section along a new dimension; within castes or *jatis* which are the building blocks of Indian society. Our second contribution is to provide an explanation for this variation, which is based on a model of assortative matching in the marriage market organized within each caste. The root cause of sex selection in the model is specific imperfections in the marriage market, which arise due to the structure of the marriage institution in India and which are shown to affect relatively wealthy households in the caste more severely. We test the predictions of the model for sorting, dowries, and sex selection across the wealth distribution within castes using unique data we have collected, covering a rural population of 1.1 million individuals in South India. Because data on multiple castes with distinct wealth distributions are available, a flexible control function approach can be implemented to estimate the causal relationship between relative wealth and sex selection. We find that the variation in sex ratios within castes in a single district is comparable to the variation across all states in the country. Estimation of the model's structural parameters allows us to quantify the impact of alternative policies that operate through the marriage market to reduce sex selection, highlighting the value of the equilibrium analysis for evaluating interventions.

Keywords. Family Economics. Institutions. Development. Macroeconomic Implications. Sex Selection. Marriage Market. Caste. Assortative Matching. Wealth Distribution. Inequality.
JEL. J12. J16. D31. I3.

---

*Akhil Lohia and Johannes Maywald provided exceptional research assistance. We are grateful to Debraj Ray for detailed notes, which helped us better understand the model, and to 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, ERC Grant 339186, and grant ECO2015-67655-P is gratefully acknowledged. We are responsible for any errors that may remain.

†Brown University
‡University College London and UPF-ICREA-GSE
§Pennsylvania State University
¶Christian Medical College, Vellore
‖University of Cambridge
**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 25 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, if anything, worsened over time.

Sex selection is a relatively recent phenomenon in South India, the setting for our research, dating back to the early 1980's. If we assume that there was no sex selection prior to that date, then sex ratios for South India from the 1961 and 1971 rounds of the population census provide a benchmark level for the natural child (aged 0-6) sex ratio, which is 102.5 boys per 100 girls.[1] Based on this benchmark, child sex ratios continue to be biased in India (see Appendix Table A1). In the worst three states, all of which are in North India, there were 115-120 boys per 100 girls in the most recent, 2011 census, which is indicative of severe sex selection. The all-India statistics are less dramatic at 109, but have, nevertheless, worsened over time. Even South India, where the problem was absent until fairly recently, now reports sex ratios at 108 that are close to the national average.

Much research and policy attention has been devoted to sub-populations in the country with severe sex selection and to explaining changes over time. The first contribution of our research is to use unique data we have collected to document substantial variation in sex ratios in the cross-section along a new dimension; within castes or *jatis* which are the building blocks of Indian society. The basic rule in India, which has been followed for generations, is that individuals must marry within their caste. The Indian population today is divided into 4,000 distinct ethnic groups, each of which is a caste (or its non-Hindu equivalent kinship group) and within which an independent marriage market is organized. Our analysis uncovers a hitherto unknown fact, which is that there is as much variation in sex ratios within castes, in a single district in South India where the average sex ratio is just about the regional average, as there is across states in the country.

The second contribution of our research is to provide an explanation for this variation within castes, which is based on the structure of the marriage institution. We propose a model of positive assortative matching in the marriage market in which wealth-dependent sex selection arises endogenously. It is widely believed that large marriage payments to the groom's family, or dowries, are the main cause of sex selection in India (e.g. Das Gupta (1987), Basu (1999)) and that the wealthy are more likely to practice sex selection because they must pay higher dowries (e.g. Murthi, Guio, and Drèze (1995)). However, wealthy girls match with wealthy boys who provide them with greater resources for consumption during marriage. If girls' parents internalize these benefits, then

[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 natural benchmark for the 0-6 age group.

it is not obvious that having a girl is especially disadvantageous for wealthy parents, despite the fact that they must pay higher dowries. Our theory formalizes the link between wealth, marriage, and sex selection. The root cause of sex selection in our model is specific imperfections in the marriage market, which arise, in turn, due to the structure of the marriage institution in India. These imperfections do not affect all households equally. In particular, sex selection is shown to be increasing with *relative* wealth within the caste.

Our analysis, linking wealth to sex selection through the marriage market advances the family economics literature in a new direction. A central objective of this literature, reviewed in Greenwood, Guner, and Vandenbroucke (2017), is to link exogenous changes in the macroeconomy to changes in marriage, fertility, investment in children, and female labor force participation. Within this literature, there is a growing body of work that examines the role of the marriage market in mediating this relationship (see Browning, Chiappori, and Weiss (2014) for a comprehensive overview). In these models, family decisions determine the sorting equilibrium in the marriage market, which, in turn, feeds back into these decisions. Applications of such equilibrium analyses include women's choice of their own or their children's education (Chiappori, Iyigun, and Weiss (2009), Chiappori, Salanié, and Weiss (2017)), birth control and fertility (Chiappori and Oreffice (2008), Caucutt, Guner, and Knowles (2002)), and parenting style (Doepke and Zilibotti (2017)). In our model, the two-way interaction is between the assortative matching equilibrium in the marriage market and a previously unexplored family decision: sex selection. The novelty of our approach is that sex selection decisions distort the distribution of males and females and, hence, the distribution of wealth on the two sides of the market. This changes the equilibrium marriage outcome, which, in turn, affects the incentives for sex selection.

In the model, each family consists of a single parent and a single child. Each parent is endowed with a particular level of wealth. When two families match, the total wealth endowment of the parents must be allocated for consumption across the four members of the two families. 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. Nonetheless, given the structure of the marriage institution, this is not the allocation of resources that emerges in equilibrium. Marriage in India is patrilocal, which means that the the girl leaves her natal home when she marries. Her altruistic parent must thus use the dowry as a mechanism to share wealth with his daughter. However, the dowry is given to the in-laws, who will siphon off some of it. This is the first source of inefficiency in our model. The second inefficiency is associated with the bequest that the boy's parent makes to the son. Because the daughter-in-law lives with the son, she cannot be stopped from consuming some of the resources that are transferred. This implies that the bequest will be set so that the boy ends up consuming less than his parent. The allocation of resources for consumption is thus sub-optimal from the perspective of both parents.

To link these distortions in consumption to marriage matching and sex selection, we solve the model in three steps. In the first step, we establish that there is positive assortative matching on wealth in this marriage market. All girls' parents want their daughters to match with wealthy boys, who will receive larger transfers from their parents, thus allowing the daughter-in-law to consume at a higher level. The dowry must, therefore, 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 either 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 and a positive dowry at every wealth level. This result is driven by the two inefficiencies in the marriage institution described above, which distort consumption allocations, and the social norm in India that all girls must marry. Although the social cost of remaining single for the girl and her parent is prohibitive, the option outside marriage for the boy is to remain single. This difference in outside options, coupled with the distortion in consumption allocations, shifts the distribution of the surplus from marriage in favor of the boy's family; the boy and his parent end up consuming more than their wealth endowment, which implies that dowries must flow to the boy's side and that parents are better off with a girl than with a boy (with resulting sex selection) at every wealth level.

The third and final step in solving the model is to establish that sex selection is increasing as we move up the wealth distribution within castes. 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).

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, 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 is increasing with wealth at two focal points in the wealth distribution: (i) at the top of the distribution, where boys and girls of equal wealth match with each other, and (ii) at the lowest wealth level where boys match, where the dowry is determined by the boy's outside option (remaining single). To characterize sex selection across the entire wealth distribution, we solve the model numerically. Validating the intuition from our simple example, the wealth-gap between boys and girls is increasing, while sex selection is decreasing, as we move down the wealth distribution. In addition, the dowry is positive at every wealth level (and increasing in wealth). This last result distinguishes our model from earlier work based on exogenous son preference; e.g. Edlund (1999), Bhaskar (2011), which also predicts that the sex ratio will be less biased moving down the wealth distribution. Because girls are on the short side of the market with sex selection, these models predict bride-price in equilibrium (as in China). Dowries co-exist with sex selection in our model, as they do in India, because particular features of the marriage institution are the root cause of the problem.[2]

We test the predictions of the model with unique data we have recently collected as part of the South India Community Health Study (SICHS), which 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, and labor force participation. 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; i.e. the household head, and the marriages of his children in the preceding five years (regardless of whether they continued to reside in their natal home). This information is used to establish that the wealth-gap between the male and the female partner is largest at the bottom of the caste's wealth distribution, with a closing of this gap moving up the wealth distribution. We also find that the dowry is positive at each point in the caste's *per capita* wealth distribution, and increasing with wealth.[3]

To test the predictions of the model for sex selection, we use the 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. We first document a negative and significant relationship between the probability that a child is a girl and the family's rank in its caste's *per capita* wealth distribution. We next

[2]Although we allow for exogenous son preference in the empirical analysis, this alternative mechanism cannot by itself explain the co-existence of positive dowries at every wealth level and sex selection.

[3]Family size determines the wealth available per member, which is the relevant variable for matching.

subject this result to greater empirical scrutiny. As discussed below, a number of explanations have been proposed for the worsening sex ratio in India. Some of these explanations could also generate variation in the cross-section. We take advantage of two features of the theory and the data to identify the marriage market mechanism: First, sex selection is determined by the family's *relative* position in its caste's *per capita* wealth distribution. Second, data from multiple castes, with distinct wealth distributions and within which independent marriage markets are organized, are available. Once caste fixed effects are included in the estimating equation, the only threat to identification is that wealth or family size, which are used to construct *per capita* wealth, are correlated with other direct determinants of sex selection. We thus include a flexible control function, with family 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 which are located at different positions in their caste's wealth distribution. The conditional estimates (with the control function) align more closely with the model's predictions than the unconditional estimates; the probability that a child is a girl is decreasing with the family's rank at each point in the wealth distribution.

The current child sex ratio in the study area, obtained from the SICHS census, is 108, matching the corresponding statistic for South India from the 2011 census. Although the sex ratio is biased, it is not exceptional. Nevertheless, within castes, our estimates indicate that the sex ratio varies from just below 100; i.e. parity, at the bottom of the wealth distribution, to as high as 117 at the top. Recall that the worst states in the country have sex ratios ranging from 115-120. Thus, the magnitude of the variation in sex ratios within castes in a single (unexceptional) district is comparable to the variation across all states in the country. The marriage market is organized in the same way in all castes; the positive relationship between relative wealth and sex selection is obtained caste by caste across the social spectrum in our study area and we would expect this relationship to hold more widely.

Given these new findings on the severity and the extent of the sex selection problem, the design of policies to address it becomes especially important. In order to evaluate the impact of alternative policy interventions, we structurally estimate the marital sorting model that we laid out above using the SICHS data. The exogenous determinants of the equilibrium allocation are (i) the parameter determining the fraction of the transfer from the boy's parent that reaches the son, and (ii) the parameter governing the cost of sex selection. For a given pattern of marital matching, which is defined by the wealth-gap between girls at every wealth level and their partners, the first parameter (governing the distortion in the allocation of consumption) will determine the utility differential between having a girl and a boy. The second parameter (governing the cost of sex selection) maps this utility differential into the sex ratio, which in turn determines the distribution of wealth for boys and girls and the associated pattern of marital matching, to close the model. The two structural parameters are estimated by matching the sex ratios predicted by the model at each wealth level in each caste to the actual sex ratios in the data.

Given these parameter estimates, we can use the model to evaluate and interpret the impact of different policy interventions, which will work through the marriage market equilibrium to change patterns of sex selection. The dowry is effectively the price for a boy, and so one policy lever to reduce the (excess) demand for boys would be to tax dowries.[4] Although such a policy has not been implemented, the central government as well as many state governments have introduced welfare schemes with the stated objective of reducing sex selection. These schemes typically consist of a cash transfer to parents, conditional on having a girl, or a direct transfer to the girl (when she becomes an adult). In some cases, the schemes are restricted to households below an income ceiling. The main findings of the counter-factual policy simulations are: (i) A tax on the dowry will have both positive and negative effects at different points in the wealth distribution. (ii) Interventions targeted at specific (low income) households will have unintended pecuniary externalities on other caste members by changing the equilibrium marriage price, possibly *increasing* the overall bias in sex ratios. These spillover effects would not be captured by models which assume an exogenous preference for sons, emphasizing the value of our equilibrium analysis for the evaluation of policy. (iii) Direct transfers to girls when they are adults are much more effective than transfers to their parents when they are children. This result suggests a promising way forward, which we examine in greater detail in the concluding section.

2 Institutional Setting

The emergence of sex selection in South India in the 1980's is especially useful for understand why this phenomenon is linked to marriage in India. Although Indians have married within their castes or *jatis* for centuries, marriages in South India were, in addition, 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)). Marriage in India is patrilocal, with girls moving into their husbands' homes, which implies that the girl typically moved into her maternal grandfather's or a maternal uncle's home. 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 (Anderson (2007), Srinivas (1989)). 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, Reddy, and Caldwell (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 de-

[4]Although dowries are illegal in India, parents circumvent the law by claiming that the dowry is a gift to the newly married couple. A tax on the dowry could thus be implemented as a gift tax.

clined (Caldwell, Reddy, and Caldwell (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.[5] A practice that was formerly confined to the upper castes has now spread across the caste distribution (Bhat and Halli (1999)).[6] The widespread emergence of dowries in South India in the early 1980's coincided with the onset of sex selection. Although it is tempting to conclude from the temporal correlation that dowries caused the sex selection, the root cause of sex selection in our model is frictions in the marriage market that emerged with the dowry, which are, in turn, determined by particular features of the marriage institution in India.

The following features of the marriage institution are relevant for our analysis: (i) Marriages are endogamous, matching individuals almost exclusively within their caste or *jati*.[7] (ii) Marriages are patrilocal, with women moving into their husbands' homes, which are often outside their natal village. (iii) Marriages are arranged by the parents and relatives of the groom and bride, with family wealth being a major consideration when forming a match.[8] (iv) Marriages involve a transfer payment or dowry from the bride's family to the groom's family.[9] (v) Despite the cost (largely associated with the dowry) of marrying a daughter, the social norm is that all girls must marry.[10]

3 A Model of Wealth, Marriage, and Sex Selection

3.1 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

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

[6]Dowries were first observed among Brahmins in Madras in the 1930's. They had spread to provincial towns, but continued to be restricted to the Brahmin caste, by the 1960's (Srinivas (1984)). By the 1980's, the practice of dowry was observed even among the lower castes in South India (Caldwell, Reddy, and Caldwell (1983)).

[7]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, Thangaraj, Patterson, Lipson, Loh, Govindaraj, Berger, Reich, and Singh (2013)).

[8]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 individual wealth.

[9]Marriage markets have always existed within castes in North India, with dowries being used to clear these markets. Although this is a relatively recent phenomenon in South India, dowries in that region are now as high as they are in the North (Caldwell, Reddy, and Caldwell (1983), Srinivas (1984), Rahman and Rao (2004), Anderson (2007)).

[10]Early and universal marriage remains the norm for females, especially in rural India (Caldwell, Reddy, and Caldwell (1983), Arnold, Choe, and Roy (1998), Bhat and Halli (1999), Basu (1999)). According to the 2005-06 DHS, less than 2% of women in rural Tamil Nadu, the setting for our research, are never married by age 35.

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. 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, Payoffs and Consumption. 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 $U = \log(C_i) + \log(c_i), \forall i = \{x, y\}$.[11] Notice that parents do not have an intrinsic preference for children of a particular gender. Denote the maximized utility of the groom's family with income x marrying a bride with income 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 benefits are that if a girl lands a well off boy, she will get to consume a fraction of the wealth her future husband receives as a transfer (or bequest) from his parent. We denote the transfer by t. While parents are altruistic towards their own children, they are not towards their children's marriage partners. Therefore, the boy's parent would like to give nothing to the bride, but given that the husband and wife live together, the groom obtains a fraction $\alpha \geq 1/2$ of the transfer, and the bride cannot be stopped from consuming a fraction $1 - \alpha$ of what her husband receives.[12] The groom's (or the bride's) parent cannot earmark what fraction of the transfer goes to the bride, and thus α is exogenously given.

[11] 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, Browning, and Meghir (1994). Although there are only two generations in the model, we can interpret the weight on the child's consumption utility as reflecting the cummulative (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, but this extension to the model would not change the results that follow.

[12] If future generations are accounted for in the parent's utility function, then the boy's parent could place positive weight on his daughter-in-law's consumption utility (to the extent that this affects the utility of her offspring). However, as long as he places more weight on his son's utility than his daughter-in-law's utility, the results that follow will remain unchanged.

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:

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

The transfer, t, and the dowry, d, are determined endogenously in the model. Based on the solution to the model, derived below, parents and their children end up consuming at different levels. However, altruistic parents would like to share their wealth equally with their children. This mismatch plays a key role in determining sex selection in our model.

3.2 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; i.e. a shortage of girls and, nevertheless, that dowries are positive at every wealth level. Third, we show that sex selection increases higher up the wealth distribution.

Matching. Matching in this marriage market is frictionless with the transfer between the bride and the groom's family d determined competitively. We denote the equilibrium allocation by $\mu(y)$, i.e., the family 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: First, participants in the marriage market choose their best partner given a "Walrasian" schedule of dowries, and the marriage market clears with a resulting equilibrium price d. Then, the parent of the boy chooses the transfer t.[13]

We solve for d and t by backward induction. 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 boy's parent will choose a transfer t that maximizes his family's utility. The maximized utility of the boy's family can be written as:

$$u(x, d) = \max_t \{\log(x - t + d) + \log(\alpha t)\}, \tag{2}$$

which is independent of y, conditional on the dowry d. The first order condition from this maximization problem implies that:

$$t = \frac{x + d}{2}. \tag{3}$$

[13] The sequential nature of the model highlights a key imperfection in the marriage market. The boy's parent cannot credibly commit to transferring a specific amount, which is agreed upon *ex ante* with the girl's parent, directly to his daughter-in-law. The resulting seepage in the dowry exacerbates sex selection in our model. Note that seepage would occur even if dowries and transfers were determined simultaneously.

Given the sequence of decisions, the dowry d is determined competitively in the marriage market, together with the equilibrium matching pattern, $x = \mu(y)$, taking the preceding optimal transfer t as given. 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.

To derive the optimality condition, it will be convenient to express d and t as functions of u. Substituting the expression for t from equation (3) in the boy's family's utility function, we obtain:

$$u = \log\left(\frac{x+d}{2}\right) + \log\left(\alpha\left(\frac{x+d}{2}\right)\right) = \log\left(\alpha\left(\frac{x+d}{2}\right)^2\right). \tag{4}$$

This permits us to write the dowry and the transfer as a function ψ of the utility u obtained by the boy's family, $t = \psi(u)$ and $d = 2\psi(u) - x$, where

$$\frac{x+d}{2} = \sqrt{\frac{e^u}{\alpha}} = \psi(u). \tag{5}$$

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,

$$v(x, y, u) = \log(x + y - 2\psi(u)) + \log((1-\alpha)\psi(u)). \tag{6}$$

This is a matching problem with Imperfectly Transferable Utility (ITU), first analyzed in Legros and Newman (2007). The first order condition to this problem satisfies

$$v_x + v_u u' = 0. \tag{7}$$

Having established the condition for optimality, the remaining condition to be satisfied for a competitive equilibrium is market clearing. To establish market clearing, we must first determine the pattern of matching; $x = \mu(y)$.

Lemma 1 *There is Positive Assortative Matching on wealth, i.e.,* $\mu'(y) > 0$ *if*

$$y \leq x\left(\frac{2}{\sqrt{\alpha}} - 1\right). \tag{8}$$

Proof. In Appendix. ■

Given that $\alpha \in [1/2, 1)$, the preceding condition will be satisfied if $y \leq x$. With positive assortative matching, 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. The *ex post* condition required to establish the existence of an equilibrium with assortative matching, $y \leq x$, is thus satisfied at every wealth level, with or without sex selection.

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 (because their husbands receive a larger bequest). The first order condition, equation (7), ensures that the hedonic price, u, and, hence, the dowry is increasing sufficiently steeply in x so that the matching on wealth is stable.[14] The dowry thus serves as a bequest and as a price to clear the marriage market in our paper.[15]

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).[16]

Sex Selection. As noted, the social norm that all girls must marry plays a key role in generating sex selection in our model. If a boy stays single, the family's wealth is divided equally between the parent and the child (given its objective of maximizing total utility). Non-participation in the marriage market is not an option for the girl, however, due to the social norm and the resulting disutility from staying single. The positive outside option for the boys leaves them with a greater

[14]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 $v(x, y, d) = \log(y - d) + \log((1 - \alpha)t)$. Substituting the expression for t from equation (3), the first order condition for this maximization problem implies that $d = (y - x)/2$. The maximized utility for the girl's family, $v(x, y) = 2 \log\left(\frac{x+y}{2}\right) + \log\left(\frac{1-\alpha}{2}\right)$ is monotonically increasing in x. All girls' parents would want them to match with the wealthiest boys and the market would not clear.

[15]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. 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)).

[16]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).

share of the surplus from marriage and parents are thus better off with boys than with girls, as made precise below. For this endogenously determined preference for 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 $H(k)$.[17] k incorporates the monetary cost, which is relatively small, and the more important ethical cost of sex selection. We assume that k is uncorrelated with wealth and is bounded below at zero. Given this sex selection technology, we can show the following result.

Proposition 1 *In equilibrium there is sex selection and dowries are positive at every wealth level.*

Proof. Suppose that girls with family wealth y match with boys with family wealth x in equilibrium. The total wealth available for consumption for the two families is $x + y$. If the boy did not marry, his parent would share the family wealth equally between them and they both would consume $x/2$. Given the outside option of remaining single, the boy's family must thus receive a consumption utility of at least $2\log\left(\frac{x}{2}\right)$ in marriage. The minimum total wealth that it needs to achieve this utility is x, but this requires that the boy and his parent consume the same amount. If their consumption levels differ, as they must given $\frac{1}{2} \leq \alpha < 1$, then the total requirement will exceed x.[18] In that case, the girl's family is left with less than y and so the maximum utility it can achieve (with equal consumption across generations) is less than $2\log\left(\frac{y}{2}\right)$. It follows that $v(y) < 2\log\left(\frac{y}{2}\right) \leq u(y)$.

A girl's parent with wealth y will manipulate the sex of the child if $k < u(y) - v(y)$. With k bounded below at zero, by assumption, there will be some amount of sex selection at every wealth level. Given that the boy and his parent consume more than their endowment, x, there must also be a positive marriage transfer from girls' parents; i.e. dowries are positive at every wealth level. ■

Because the market clears from the top and there is a shortage of girls at every wealth level, there will be hypergamy; i.e. girls marry wealthier boys, at every wealth level, except at the very top where boys and girls with equal wealth match. There will also be a wealth threshold below which boys are unmatched.

If there was bride-price in equilibrium; i.e. payments from boys to girls at the time of marriage, and the population is growing, then the deficit of girls could be cleared by boys "buying" younger girls (Tertilt (2005), Neelakantan and Tertilt (2008), Bhaskar (2011)).[19] In our model, the last boy

[17]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.

[18]Substituting the expression for t from equation (3) in the expressions for C_x, c_x, it is straightforward to verify that $c_x = \alpha C_x$.

[19]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

to match is indifferent between marrying and staying single and 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.

Wealth and sex selection. Proposition 1 establishes that there will be sex selection at every wealth level. However, it does not tell us how sex selection will vary across the wealth distribution. 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 intuition.

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 1 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 $k^\star$ such that

$$k^\star(y) = u(y) - v(\mu(y), y; u(\mu(y))). \tag{9}$$

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 as follows:

$$F(x) = \int_x^{\overline{x}} (1 + H(k^\star(z)) d\Gamma(z)/2 \quad \text{and} \quad G(y) = \int_y^{\overline{y}} (1 - H(k^\star(z)) d\Gamma(z)/2, \tag{10}$$

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

$$\int_{\mu(y)}^{\overline{x}} dF(x) = \int_y^{\overline{y}} dG(y) \tag{11}$$

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 different from one.

or equivalently:

$$\int_{\mu(y)}^{\overline{x}} (1 + H(k^{\star}(z))d\Gamma(z)/2 = \int_{y}^{\overline{y}} (1 - H(k^{\star}(z))d\Gamma(z)/2. \tag{12}$$

Sex selection determines the distribution of wealth for boys and girls, which, in turn, determines the pattern of matching in equation (12). The pattern of matching determines sex selection in equation (9). 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 (7), $v_x + v_u u' = 0$, which implies $u' = -\frac{v_x}{v_u}$, with respect to x:

$$u(x) = \int_{x^{\star}}^{x} -\frac{v_x(x, \mu^{-1}(x); u(x))}{v_u(x, \mu^{-1}(x); u(x))} dx + u(x^{\star}) \tag{13}$$

where the denominator is negative, and where $x^{\star}$ is the lowest wealth boy who is matched. From the outside option, we know that $u(x^{\star}) = 2\log\frac{x^{\star}}{2}$.

The equilibrium is fully defined by the sex selection condition, the matching condition, and the payoff condition, as specified in equations (9), (12), and (13). 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 (13). 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 3.3). 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 $u(x^{\star}) = 2\log\left(\frac{x^{\star}}{2}\right)$.

Proposition 2 *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* $\alpha \geq \tilde{\alpha}$ *for some* $\tilde{\alpha}$*.*

Proof. In Appendix. ■

If the result in Proposition 2 holds over the entire wealth distribution, then this implies that sex selection worsens monotonically with relative wealth. The intuition for this conjecture, which is validated by the numerical results, is that the shortage of girls grows as we move down the wealth distribution because more and more boys are left unmatched above them. This implies that poorer girls match with relatively wealthy boys; i.e., there is an increase in hypergamy, making the switch (through sex selection) to a boy relatively unattractive. The girls at the bottom of the wealth distribution benefit the most from the sex selection above them and thus the sex ratio at the bottom is least distorted.

3.3 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 (7), $\frac{dv}{dx} = 0$, 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 1 that there will be sex selection at every wealth level. 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.

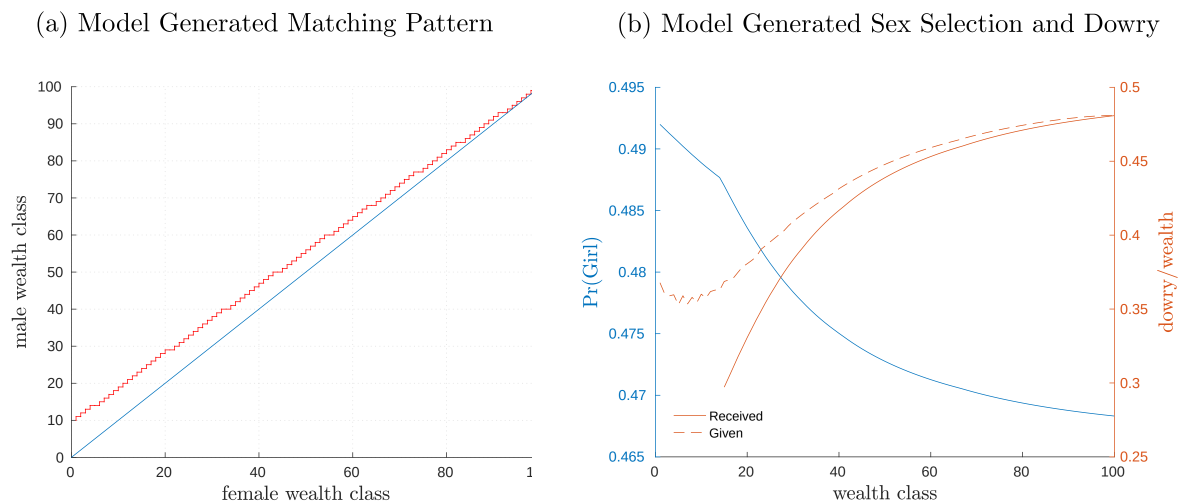
Given any initial guess for $u(\overline{x})$ and the matching pattern, we can solve for $u(x)$ and $v(y)$ in each wealth class. $v(y)$ is a function of y, x, and $u(x)$, as specified in equation (6). Given that girls in the highest and the next to highest wealth class match with the wealthiest boys, 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, $2\log\left(\frac{x^\star}{2}\right)$, 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 census data (within castes). The wealth distribution is divided into 100 classes for the simulations. As in Proposition 2, $k \sim U[0, a]$, which implies that there are two parameters in the model: α and a. We select values for these parameters: $\alpha = 0.6$ and $a = 12$ that are close to the values estimated below.

Figure 1: **Simulated Model (parameter values $\alpha = 0.6; a = 12$)**



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. 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, it must nevertheless be the case that the dowry given is greater than the dowry received at a given wealth level. This is what we observe in Figure 1b. Proposition 1 states that the dowry is positive at all wealth levels and this is also true in our numerical simulation.

We have shown analytically that there will be sex selection at every wealth level (Proposition 1) and that sex selection is increasing in wealth at the bottom and the top of the distribution (Proposition 2). However, we cannot analytically characterize the relationship between wealth and sex selection over the entire distribution. Figure 1b reports this relationship, based on the numerical solution to the model. The proportion of girls declines monotonically as we move up the wealth distribution, with an accompanying increase in the dowry.[20]

Figure 2: **Model Generated Comparative Statics**

(a) Friction: α (b) Cost of Sex Selection: a

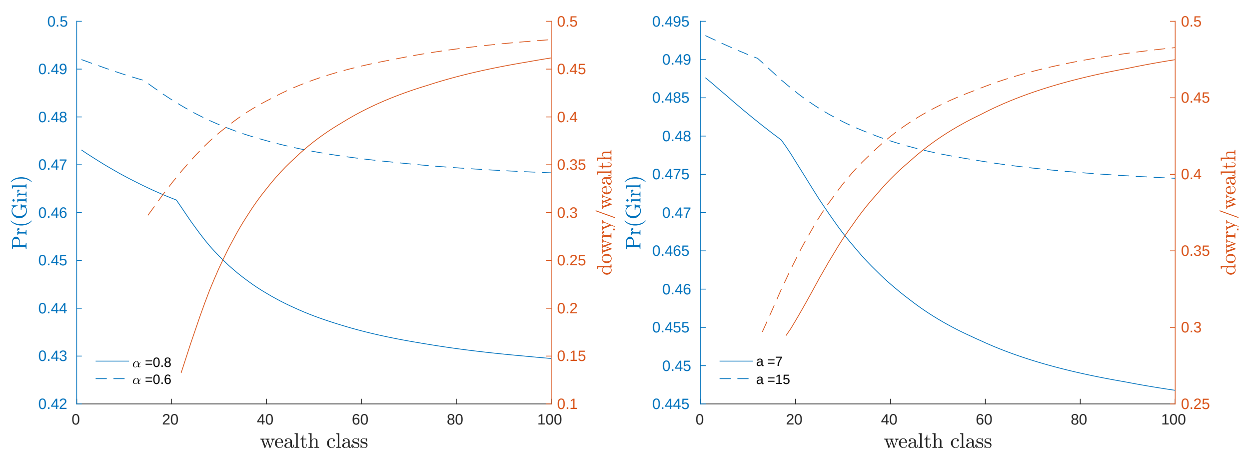


Figure 2a reports the relationship between wealth and both dowries and sex selection for different values of α. The direct effect of an increase in α is to make boys' families better off at the expense of girls' families. However, this effect is partially offset by the decrease in the equilibrium dowry observed in Figure 2a.[21] The direct effect nevertheless dominates, resulting in an increase in sex selection at every wealth level. Figure 2b reports the same plots for different values of a, the sex selection parameter. The direct effect of a reduction in the cost of sex selection is a decline in the fraction of girls at every wealth level. This is partially (but not completely) offset by the reduced competition for boys, which shifts dowries down in Figure 2b.

[20]Notice that the dowry is less than half the family's wealth at the top of the wealth distribution. This satisfies the condition for sex selection to be increasing in wealth at that point in the distribution, as derived in Proposition 2.

[21]If there was sufficient curvature in the utility function, then altruistic parents would compensate for the decline in their daughters' consumption by increasing the dowry (at the cost of their own consumption). This condition is evidently not satisfied with logarithmic preferences. The effect of an increase in α is to reduce competition for boys in the marriage market, shifting down the dowry.

The parameter values in Figure 2 are chosen to cover a relatively wide parameter space. Nevertheless, the robust finding is that the dowry is always positive and increasing in wealth, while the proportion of girls is decreasing in wealth. An increase in α reduces the proportion of girls, while a decrease in a, which implies a reduction in the cost of sex selection, also has the same effect. In our model, sex selection is generated by (i) the social norm that all girls must marry, and (ii) by the mismatch between the child's actual consumption allocation and the parent's preferred allocation. Proposition 1 characterizes one channel through which sex selection occurs, which combines the social norm and the fact that boys end up consuming less than their altruistic parents (because their wives capture some of the parental transfers). The comparative statics analysis in Figure 2a identifies another channel through which sex selection occurs. As α increases, the consumption gap between the girl and her parent grows, making it less attractive for parents to have a girl.

Economic development and sex selection. The α parameter will, in general, reflect the woman's status and bargaining power in her marital home. Although α is treated as exogenously determined and fixed in our model and in the cross-sectional empirical analysis, this parameter could change over the course of the development process. Anderson and Bidner (2015) decompose the marriage payment into two components: a bequest to the girl, which she controls in her marital home and which could include pre-marital investments in her human capital, and a groom-price. They show theoretically that parental resources could be shifted from the bequest to the groom-price with economic development, resulting in a decline in women's status. The equivalent increase in α, in the context of our model, would worsen sex selection, providing one explanation for the increasingly biased child sex ratios in the dynamic South Indian economy over the past decades.

Anderson and Bidner's model also provides an additional mechanism through which sex selection could be reduced. If girls' parents invest in their daughters' human capital (and girls enter the labor force after they marry) then the daughters will be less reliant on the transfers from their husbands' parents. The mismatch between the daughter's actual consumption and the preferred level of consumption from her parent's perspective, which is one direct cause of sex selection, will thus decline. We find no empirical evidence in support of this mechanism, perhaps because female labor force participation (which is a necessary condition for this mechanism to be effective) remains unusually low in India.

Policy experiments. Our model is able to capture key features of the marriage institution and, based on these features, to generate sex selection in equilibrium. The model is also well suited to evaluate the effect of policies designed to address the sex selection problem because most policies will either directly or indirectly work through the marriage market.

Given that the dowry is effectively the price for a boy, one obvious policy lever to reduce the demand for boys would be to tax the dowry. A dowry tax is introduced in our model by assuming that the boy's parent receives an amount θd, where $\theta < 1$. This directly affects the transfer from the boy's parent to his son and equation (3) can be rewritten as,

$$t = \frac{x + \theta d}{2}. \tag{14}$$

Substituting as before, $t = \psi(u)$ and $d = \frac{2\psi(u) - x}{\theta}$, where $\psi(u) = \sqrt{\frac{e^u}{\alpha}}$. This allows us to write the maximized utility for the girl's family as

$$v = \log\left(y - \frac{2\psi(u) - x}{\theta}\right) + \log((1 - \alpha)\psi(u)). \tag{15}$$

If the dowry d remains fixed, then the boy's family's utility u will decrease.[22] The girl's family's utility will also decrease; although her parent's utility is unchanged, the girl's utility declines with the decline in t. Thus, the net effect on sex selection is ambiguous. This ambiguity is compounded by the fact that the dowry d will shift up in response to the dowry tax. The advantage of our model is that it can be solved numerically, using the modified expressions for u and v, to derive the equilibrium price response to the dowry tax and the accompanying consequences for sex selection at each wealth level.

The preceding discussion highlights the value of our model in analyzing the complex equilibrium response to external interventions. Although a tax on the dowry has yet to be implemented, many schemes have already been introduced with the specific objective of reducing the bias in child sex ratios. In the framework of our model, these schemes either provide a wealth transfer to girls' parents, conditional on having a girl, or a direct transfer to the girl. These policies 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,

$$v = \log(y + w + x - 2\psi(u)) + \log((1 - \alpha)\psi(u)). \tag{16}$$

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

$$v = \log(y + x - 2\psi(u)) + \log((1 - \alpha)\psi(u) + w). \tag{17}$$

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 when evaluating existing transfer schemes that are targeted at less wealthy parents. While the targeted families may be induced to have more girls, there will be spillover effects through the equilibrium marriage

[22]The transfer t declines, from the expression above. Given that $t = \psi(u)$ and that ψ is an increasing function of u, it follows that u will decrease as well.

market price that could *increase* sex selection at other points in the wealth distribution. These spillover effects, which are found to be quantitatively important below, would not be captured by alternative models of sex selection that are based on an exogenous preference for boys.

Alternative models. Models in which there is an exogenous preference for sons; e.g. Edlund (1999), Bhaskar (2011), will also generate increasingly biased sex ratios higher up in the wealth distribution for the same reason as in our model. However, there are a number of important differences between our model and these models (and models based on an exogenous preference for sons, more generally).

First, models with exogenous son preference in which marriage payments clear the market *ex post* predict that there will be bride price in equilibrium because there is a shortage of girls.[23] While this prediction is consistent with the bride-prices that are observed in countries like China, where sex selection is determined by the demand for a son, it is inconsistent with the marriage transfers from the girl's side to the boy's side that are universally reported in India. Our model, which incorporates relevant features of the Indian marriage institution, predicts that there will be positive dowries at each point in the wealth distribution, even for the last boy to marry, despite the fact that girls are on the short side of the market.[24]

Second, we posit that sex selection is *determined* by particular features of the Indian marriage institution, and the resulting imperfections in the marriage market. Thus, while sex selection and dowries are jointly determined in our model, we would expect an exogenous change in the dowry to have a causal effect on sex selection. Bhalotra, Chakravarty, and Gulesci (2016) construct a statistical instrument for the dowry, based on shocks to the price of gold (an important component of the dowry) around the time of conception. Their instrumental variable estimates indicate that the dowry does indeed have a causal effect on sex selection, consistent with our model.

Third, a more nuanced view of son preference, which is commonly assumed in the literature, is that parents want at least one son. One consequence of this view is that the sex ratio will vary with birth-order because parents are more likely to manipulate the sex of a child (conditional on not having had a boy) as they move closer to the desired family size. One additional consequence of this view, even if parents want just two children, is that first births will be unbiased. In contrast, our model, in which sex selection is generated by marriage market imperfections, applies to all births. We will see below, using multiple data sets, that the child sex ratio for first born children is indeed biased relative to the natural benchmark, providing support for our model.[25] However, it is less

[23] Bhaskar's model will also generate a groom-price, due to a net shortage of men, if men marry younger women and the population is growing sufficiently fast. However, his model cannot explain why a groom-price is observed in states like Tamil Nadu, the setting for our research, where fertility rates have been below replacement since the mid-1990's and where there is a surplus of men due to sex selection.

[24] This does not imply that the shortage of girls plays no role in our model. It is this shortage that pins down the dowry for the last boy to match such that he is indifferent between marrying and staying single.

[25] This result is not inconsistent with the assumption in many studies on sex selection that first births are unbiased. The statistics that we report below are based on the child (aged 0-6) sex ratio, which could be biased even if the sex

biased relative to the sex ratio for all children, indicating that an exogenous preference for sons is also empirically relevant in our setting.

4 Quantitative Analysis

4.1 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.[26]

Table 1 reports the age distribution, marriage patterns, literacy rates, and labor force participation 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 accepts 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. Notice that 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. The extremely low female labor force participation rates will be relevant later when we discuss potential policies to reduce sex selection.

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.[27] The survey collected information on key aspects of the marriage

ratio at birth is unbiased.

[26]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.

[27]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

Table 1: **Comparison of Demographic and Socioeconomic Characteristics**

Region	Men			Women		
	South India	Tamil Nadu	Study Area	South India	Tamil Nadu	Study Area
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	0.1	0.0	0.0	1.6	0.6	0.8
20-24	2.2	1.0	1.0	7.0	5.4	5.9
25-29	5.4	4.4	4.8	7.4	7.8	8.1
30-34	6.9	6.8	6.5	7.4	7.4	6.6
35-39	6.5	6.6	7.1	6.2	6.2	7.7
40-44	6.2	6.3	3.7	5.9	6.2	3.1
45-49	5.4	5.8	6.8	4.5	4.8	5.8
50-54	4.6	5.0	4.8	3.2	3.7	3.5
55-59	3.4	4.0	4.1	3.5	3.9	3.3
60-64	2.8	3.1	3.8	1.8	2.0	1.9
65-69	2.1	2.4	2.4	1.1	1.1	1.1
70-74	1.3	1.4	1.6	0.6	0.6	0.3
75-79	0.8	0.9	0.8	0.2	0.2	0.1
80-84	0.3	0.3	0.4	0.1	0.1	0.0
85>Yrs	0.2	0.2	0.2	0.0	0.0	0.0
Kolmogorov-Smirnov test of equality (p-value)	1.00	1.00	–	1.00	0.75	–
Literacy rate (%)	79.2	82.1	76.9	61.1	65.5	62.4
Labor force participation rate (%, 15-59 years)	79.8	81.1	81.0	44.9	42.6	40.0

Notes: % married measures the number of married individuals in each age category as a fraction of the total population, seperately for men and women. South India includes Maharashtra, Karnataka, Andra Pradesh, and Tamil Nadu. Literacy defined by the Goverment of India as 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 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.
Sources: For Tamil Nadu and South India, age/marriage data from Ministry of Home Affairs, GOI, and literacy data from 2011 census reported by Office of the Registrar General and Census Commissioner, GOI; Labor force participation: Ministry of Labor and Employment, Government of India, 2009-10. For Study Area, all statistics based on SICHS Census.

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 2 provides information on marriages over the two generations based on data from the SICHS survey.[28] In line with nationally representative survey evidence and genetic evidence for the

caste in the study area, excluding castes with less than 100 households in the census.

[28]The larger number of marriages for daughters versus sons in the last 5 years is because girls marry younger than boys in India. Although most girls marry in their twenties, men will marry into their thirties. Given that the fathers

Table 2: **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 (in thousand Rupees)	–	138.32	187.46
Mean fraction of annual income	–	2.94	3.83
Observations	3,524	421	611

Source: SICHS household survey.

country as 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, 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 2 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.[29] 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. In line with past studies; e.g. Rao (1993b), Jejeebhoy and Sathar (2001), and Rahman and Rao (2004), the dowry is three to four times the household's annual income on average, which is a substantial sum in an economy where access to market credit is severely restricted.[30] Notice that dowries paid by girls are larger than dowries received by boys. One explanation for the gender-gap is reporting bias, with respondents inflating the amount they gave and under-reporting the amount they received. A second explanation is hypergamy; i.e. that girls marry wealthier boys on average, which emerges

are aged 25-60, there are more girls of marriageable age in the households in our sample.

[29]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).

[30]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. Note that the amount of money that must be raised is large, even by the standards of a developed economy. For example, the maximum amount that banks lend in developed economies for the purchase of a home is typically 2.5 to 3 times the annual household income.

naturally in our model when there is sex selection and the marriage market clears from the top. Hypergamy results in girls paying a higher dowry than boys receive at each level of wealth in Figure 1b.

Table 3: **Evidence on 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; sample: marriages of children in the last 5 years.

Table 3 provides evidence supportive 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, Reddy, and Caldwell (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. Table 4 reports child (aged 0-6) sex ratios from three sources: the 2005 Demographic and Health Survey (DHS), the 2005 India Human Development Survey (IHDS), and our own South India Community Health Study (SICHS).[31] First, child sex ratios, measured by the num-

[31] The SICHS is situated in rural Tamil Nadu, but sample sizes for that one state are too small in the DHS and the IHDS to permit a meaningful comparison of sex ratios across data sets. We saw in Table 1 that the study area is representative of rural Tamil Nadu and rural South India with respect to socioeconomic and demographic characteristics. We thus compare child sex ratios in the SICHS with child sex ratios from rural South India, obtained from the DHS and the IHDS, in Table 4. Despite the fact that we now include an entire region of the country, the number of children in the DHS and the IHDS is still an order of magnitude smaller than in the SICHS census, emphasizing the novelty of our data.

Table 4: **Sex Ratios**

Population	Rural South India		Rural Vellore
Data Source	DHS2005	IHDS2005	SICHS census
First-born children	105	106	106
All children	109	108	108
Observations	5,750	3,057	79,027

Note: Sex ratios are computed for children 0-6. The unbiased child sex ratio, based on pre-1980 population census data from South India, is 102.5.

ber of boys per 100 girls, are very similar across the three data sets. Given the representativeness of the study area, we would expect our findings to apply more broadly. Second, the sex ratio is less biased for first births, consistent with the common assumption that parents want at least one son. Third, if we take 102.5 as the natural benchmark (based on sex ratios in South India prior to the 1980's) then first births are, nevertheless, biased. This indicates that the marriage market mechanism, which is the focus of our analysis and which would bias all births, co-exists with an exogenous preference for sons.

4.2 Measuring Wealth

The model generates predictions for variation in hypergamy (the wealth-gap between boys and girls), dowries, and sex selection across the wealth distribution within castes. To test these predictions with family-level data, we must measure each family's position in its caste's wealth distribution. 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 in the empirical analysis by using *per capita* wealth to determine the family's position in the caste wealth distribution.

The SICHS census and the SICHS survey both collected information on the household's income in the preceding year. 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. The household's income flow will be positively correlated with its stock of wealth. However, the income flow will also include a transitory component, which is especially relevant in a rural setting where weather shocks have a large impact on short-run income. The resulting measurement error will bias the relative wealth coefficient in our analysis towards zero.

If we observed the household's income realizations over multiple years, then the income shocks could be purged by using the average income over time. This is what we do for the analysis based on SICHS survey data, by taking the average of the household's income from the census and the survey. Although reported incomes from the census and the survey are highly correlated for a given household, the correlation is not perfect; the SICHS census was completed in 2014, whereas the SICHS survey was conducted in 2016.[32] For analyses based on the SICHS census, however, this approach is not an option because only a small sub-sample of households are included in the survey. For those analyses, we purge the income shocks by using historical wealth to predict current wealth.

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. Information on the agricultural revenue tax, per acre of cultivated land, that was collected from these villages by the British colonial government in 1871 is available from the British Library in London. The colonial government carefully measured soil quality, irrigation, and other growing conditions in each village. The revenue tax was based on the potential output per acre, rather than the actual output per acre. Given the detailed information that was used to predict the output per acre, this statistic would have been highly correlated with actual agricultural productivity and, hence, with historical wealth (but would not have been contaminated by income shocks).

Soil quality is a fixed characteristic, which will continue to determine productivity and agricultural income today. Irrigation in the study area in the nineteenth century was almost entirely provided by surface tanks. A relatively small number of villages, which had access to tank irrigation, could grow rice, which increased their income substantially. Tank irrigation has been largely replaced by well irrigation, which is less geographically constrained. However, historical advantage could have persisted in an economy with imperfect credit markets by allowing households in historically wealthy villages to make profit-enhancing investments over time.[33]

We test the hypothesis that a household's current wealth is determined by historical agricultural productivity in its village by estimating the relationship between household income, obtained from the SICHS census, and historical agricultural productivity, measured by the tax revenue per acre

[32] With two realizations, the average income from the SICHS census and the SICHS survey will not purge the income shocks completely. An alternative strategy would be to regress household income from the SICHS survey on household income from the SICHS census and then use the estimated coefficients (together with the census income) to predict each household's wealth. This method will purge measurement error in the survey income completely. Although measurement error in the census income will remain, its effect will be reduced because the estimated coefficient on census income in the predicting equation is less than one. Results with the alternative methods are comparable, as shown below.

[33] The implicit assumption underlying the historical persistence 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)). The 1871 population census provides the caste composition of each historical village in the study area. This allows us to construct the population share of each caste in each village in 1871 and the corresponding statistic today, based on the SICHS census. The correlation between these statistics is as high as 0.8.

in 1871. We allow for the possibility that this relationship could vary across castes, given that they have historically been engaged in different occupations, by including caste fixed effects and the interaction of these fixed effects with the historical agricultural productivity variable.[34] Historical agricultural productivity strongly predicts current household income, with the F-statistic measuring joint significance of the productivity coefficient and the productivity-caste interactions as high as 20.4. Measurement error in the wealth variable for households in the SICHS census can thus be purged by replacing reported household income with predicted household income.

We account for heterogeneity in family size in the empirical analysis by using *per capita* wealth to measure each family's position in its caste's 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.[35] Failure to adjust for family size, conditional on household wealth, would also generate measurement error that biases the relative wealth coefficient towards zero.[36] We will see below that purging the transitory income component from reported household income increases the magnitude of this coefficient, as does the correction for family size.

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 wealth within the caste is stable across age cohorts. Our predictor of household wealth, based on historical productivity, is independent of the age composition of the household by construction. However, *per capita* wealth, which is used to determine relative wealth, depends on family size, which could vary with the cohort of the child. This is unlikely to be a concern in a setting where fertility has been just around replacement since the mid-1990's. Nevertheless, 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

[34] Standard errors in this regression are clustered at the *panchayat* level.

[35] 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, in any case, robust to constructing *per capita* wealth on the basis of household size.

[36] When we do not account for family size, the estimating equation implicitly assigns the average family size to all households in the sample. A term which is a function of the difference between the family's own size and the average family size thus enters the residual of the estimating equation, biasing the wealth coefficient downward.

almost identical with both measures for each outcome.

4.3 Evidence on Hypergamy and Dowries

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. While the household wealth of the primary respondent and his children is available from the survey, this information is not directly available for their spouses. To construct a measure of the spouses' family wealth, we take advantage of the fact that the identity of their natal village was collected in the survey. Recall from Table 2 that women move from their natal village to a different village in about 80% of marriages. As noted, historical agricultural productivity is a strong predictor of current wealth. The equation linking current income to historical productivity that we estimate with the SICHS census data can thus be used to predict family wealth for the spouses.[37] To be consistent, we use the same equation to predict family wealth for the primary respondent and his children. One limitation of this approach is that the same wealth level will be predicted for families belonging to a given caste residing in a given village. Information on family size is not available for the spouses and so we cannot construct a measure of *per capita* wealth. This will introduce measurement error in the wealth variable, as discussed above.

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 family 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.

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.[38]

[37]The information on historical productivity 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 wealth even for villages outside the study area.

[38]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

Table 5: **Hypergamy and Relative Wealth**

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

Source: SICHS survey. 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. *** $p<0.01$

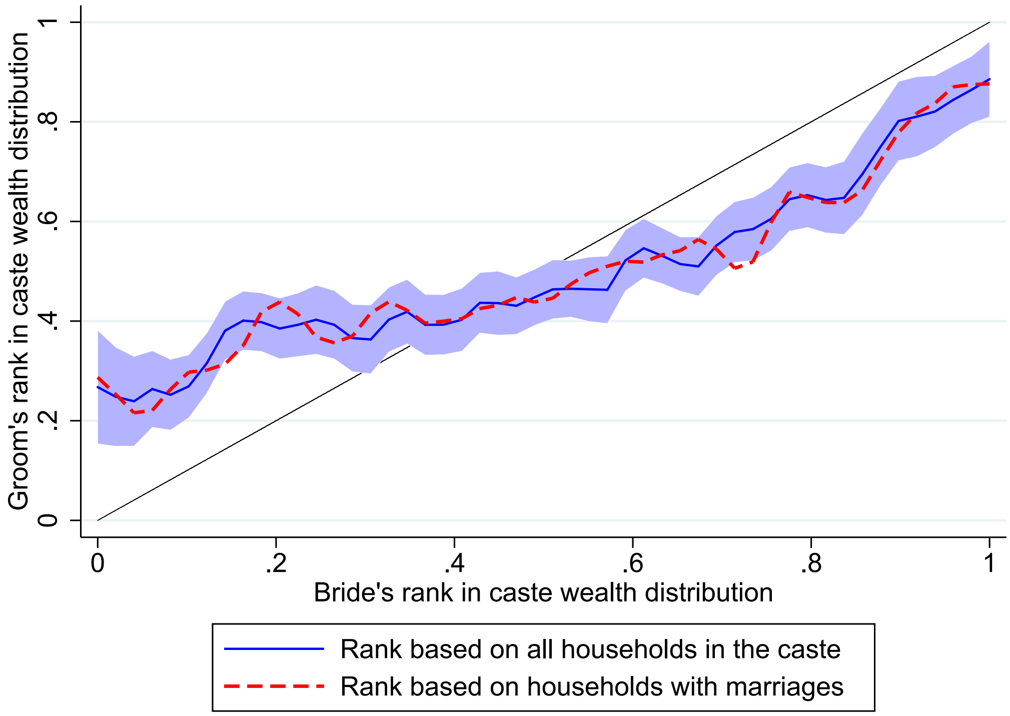
The coefficient on the education variable is small in magnitude and statistically insignificant. A comparison of Columns 1 and 2 indicates that the relative wealth coefficient is hardly affected by the introduction of female education, supporting the assumption in the model that matching on family wealth 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.[39]

Figure 3 subjects the results in Table 5 to closer scrutiny by reporting nonparametric estimates of the relationship between the male's family's relative wealth and the female's family's relative wealth. The family's position in its caste's wealth distribution in Table 5 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 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 3 does not quite match Figure 1a. In particular, although the intercept is positive and there is convergence to the 45 degree line, this convergence

down to the caste-gender-age level.

[39]A 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. We would expect educated parents, who invest more heavily in their children's education in any case, to be more likely to take advantage of this mechanism. This implies that the children of more educated parents should be more likely to be 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.

Figure 3: **Hypergamy and Relative Wealth**



is too sharp with the result that the wealthiest boys do not end up matching with the wealthiest girls. We conjecture that this discrepancy is due to measurement error, for which we find supporting evidence presented below; we are using family wealth rather than *per capita* wealth to construct the relative wealth statistic and this will bias the slope coefficient downward.[40]

The model predicts that dowries are positive at each point in the wealth distribution and increasing with relative wealth. Table 6 reports the relationship between dowries and relative wealth. The sample 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 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.[41]

[40]This bias will generate a positive upward bias in the intercept (to match predicted and actual average relative wealth on the male side). This is why the estimated intercept is substantially higher than the excess of boys; given the level of sex selection in our study area, this should be around 0.1.

[41]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 half his family wealth.

Table 6: **Dowry and Relative Wealth**

Dependent variable	Dowry		
Wealth measure	SICHS survey	Average of SICHS survey and census	Average per capita
	(1)	(2)	(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

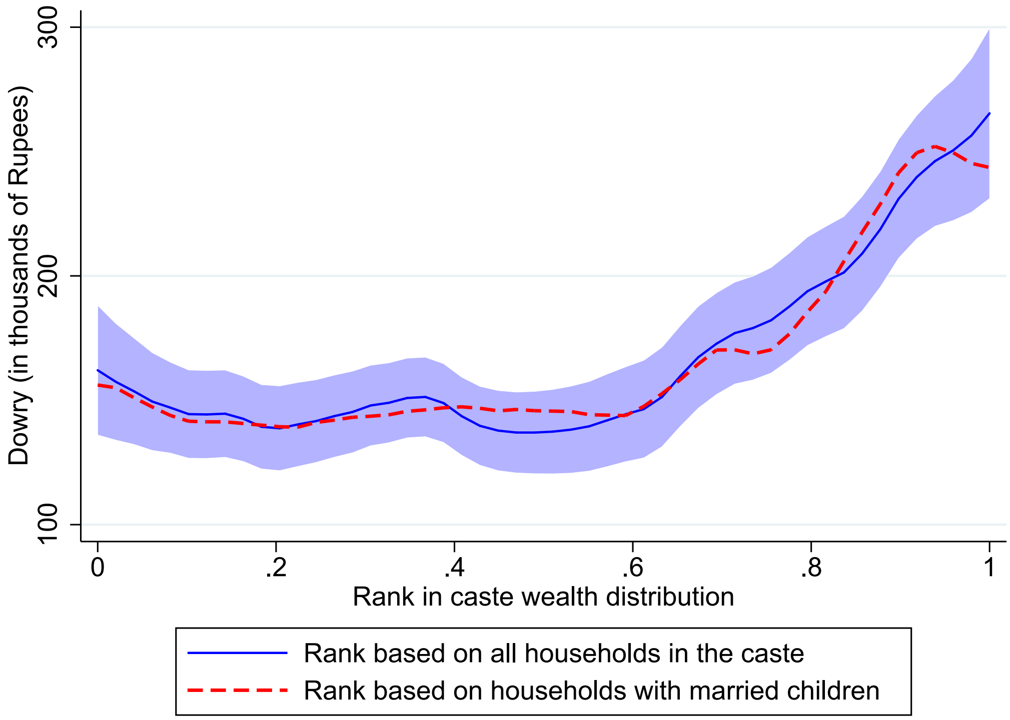
Source: SICHS survey. Sample based on all children's marriages in the past 5 years. Dowry measured in thousands of Rupees. Relative wealth measured by rank in the caste wealth distribution, from 0 (poorest) to 1 (wealthiest). *** $p<0.01$, ** $p<0.05$

Table 6, Column 1 reports estimates with the most basic specification, where the family's relative wealth is based on the income reported in the SICHS household survey. As discussed, this statistic is subject to two sources of measurement error, associated with (i) the transitory component of reported income, and (ii) the failure to adjust for family size. Table 6, Column 2 accounts for the first source of measurement error by using the average of the reported income from the SICHS census and the SICHS survey to measure household wealth. Table 6, Column 3, in addition, adjusts for family size by using *per capita* wealth to construct the relative wealth statistic. As expected, the coefficient on the relative wealth variable increases in size as we sequentially remove different sources of measurement error. Although a similar exercise cannot be implemented with the hypergamy analysis, these results provide indirect support for the argument that measurement error in the relative wealth variable, on account of the inability to adjust for family size, biased the coefficient on the female's family wealth downward in Table 5 and Figure 3.

Figure 4 reports nonparametric estimates corresponding to Table 6, Column 3 (our most preferred specification).[42] The family's position in its caste's wealth distribution in Table 6 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. Notice that dowries are positive at each point in the wealth distribution and are increasing particularly steeply higher up the distribution. This

[42] The gender dummy and the caste fixed effects are partialled out in a first step prior to estimating the nonparametric regression. The same two-step procedure is used in all the nonparametric regressions that follow.

Figure 4: **Dowry and Relative Wealth**



last observation is particularly striking because we will see below that the shortage of girls is most acute in precisely that region of the wealth distribution.[43] Figure 5 reports the relationship between dowries and relative wealth, separately for boys and girls. The amount that is given by the girls exceeds the amount that is received by the boys at each wealth level, as in Figure 1b, with little overlap in the confidence intervals.[44]

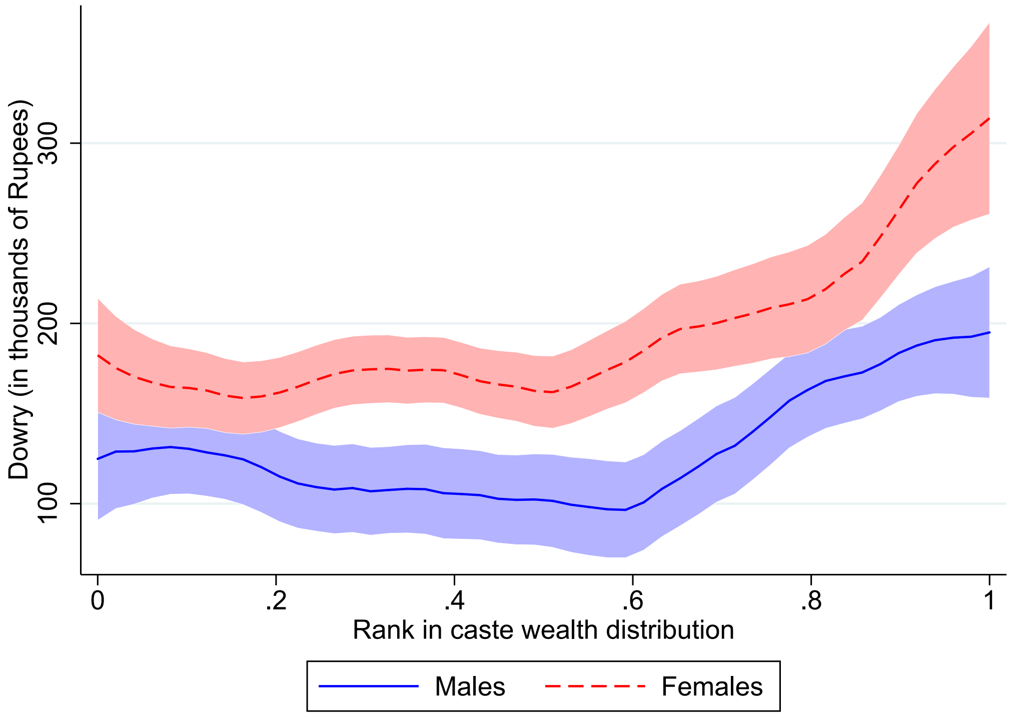
4.4 Evidence on Sex Selection

The model predicts 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 4 that there are nearly 80,000 children aged 0-6 in the study area.

[43] An alternative method to purge measurement error regresses household income from the survey on household income from the census, and then uses the estimated coefficients (together with the census income) to predict household wealth. With this method, the results are even stronger; the dowry is increasing in relative *per capita* wealth at each point in the wealth distribution, as reported in Appendix Figure A1.

[44] The dowry is measured as a fraction of household wealth in Figure 1b, and with that measure there is convergence in the amount given and received as we move up the wealth distribution. We cannot measure the dowry as a fraction of wealth in the empirical analysis because wealth is a regressor, and any measurement error that remains in that variable will thus bias the estimated coefficient on relative wealth downward. Notice, however, that dowries for boys and girls track roughly in parallel in Figure 5. If we normalized by wealth, this implies that there would be convergence moving up the wealth distribution, consistent with Figure 1b.

Figure 5: **Dowry and Relative Wealth (by gender)**



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 wealth distribution. 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. Relative *per capita* wealth in Table 7, Column 1 is based on the household's reported income in the SICHS census, which we know is subject to measurement error. The transitory component of the reported income is purged in Column 2 by replacing reported income by predicted income, determined by agricultural productivity in the household's village in 1871 and its caste. The coefficient on relative wealth is negative in both columns, but it grows substantially larger (in absolute magnitude) and is statistically significant at conventional levels when relative wealth is based on predicted income.

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, fertility will be complete.[45] This is not the case for the sex selection analysis. We account for this in Table 7, Column 3 by replacing observed family size with predicted family size, based on observed parental and household characteristics, for families where fertility may not be complete.[46] The estimated coefficient on relative wealth is

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

[46]The completed family size in Column 3 is predicted in two steps. First, the Ordered Probit model is used to estimate the relationship between the number of children and household wealth (predicted income), mother's age,

Table 7: **Sex Selection and Relative Wealth**

Dependent variable	Girl dummy		
Wealth measure	Reported wealth	Predicted wealth	
Family size measure	Observed		Predicted
	(1)	(2)	(3)
Rank in caste per capita wealth distribution	-0.00554	-0.0455***	-0.0453***
	(0.00847)	(0.00682)	(0.00674)
Mean of dependent variable	0.480	0.480	0.480
Observations	78,979	78,979	78,979
Caste FE	Yes	Yes	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). Predicted wealth is based on historical agricultural productivity in the village and the household's caste. Predicted family size is based on household wealth and parental characteristics. Standard errors (in parentheses) are calculated on the basis of 500 bootstrap replications. All standard errors are clustered at the panchayat level. *** $p<0.01$

hardly affected by this adjustment.

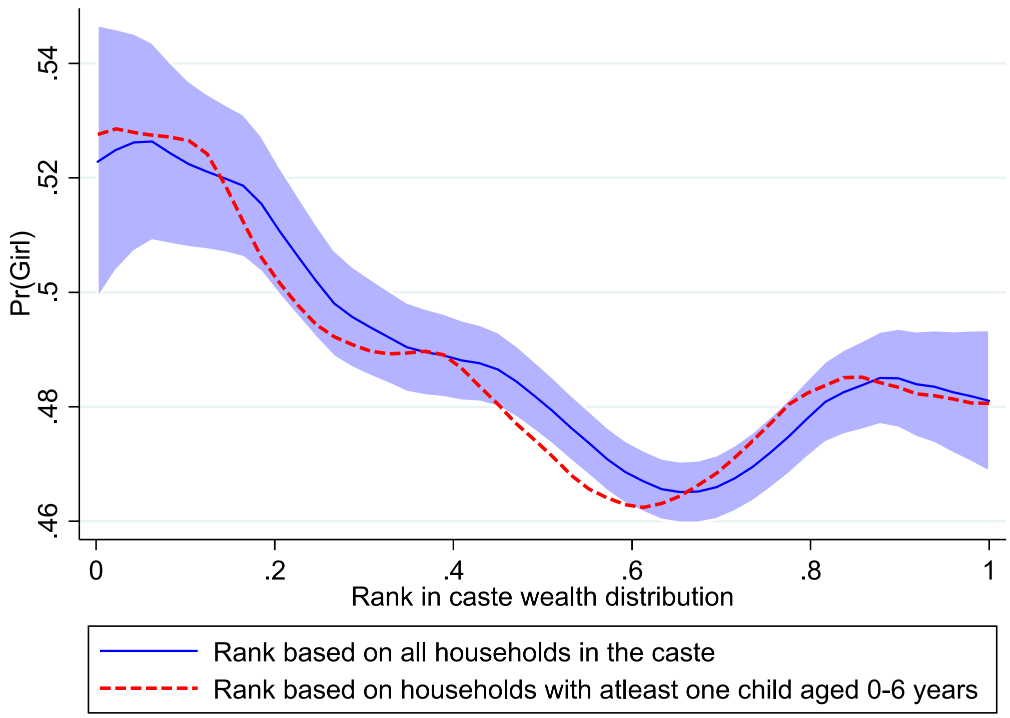
Figure 6 reports nonparametric estimates corresponding to Table 7, Column 2. The family's position in its caste's wealth distribution in the benchmark specification shown in blue, with the accompanying 95% confidence interval, is based on all households in the caste. The alternative construction of relative 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.

Sex selection in our model is determined by imperfections in the marriage market, organized within each caste. However, other determinants of sex selection could co-exist with this mechanism. 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, Kishor, and Roy (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)).[47] The same factors could generate cross-sectional variation

mother's age squared, mother's education, father's education. This equation is estimated with families whose children are aged 5-15. Because birth spacing rarely exceeds five years in the census data and children rarely leave their natal home before age 15, the size of these families can be assumed to be complete and accurately measured. In the second step, we use the estimated coefficients to predict the ultimate family size for families with children less than 5 years old (who may not have completed their fertility).

[47] As noted, sex ratios will worsen at higher birth-orders when there is a demand for a son. Exogenous fertility

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



in sex selection.

We take advantage of two features of the theory and the data to identify the marriage market mechanism: First, sex selection is determined by the family's *relative* position in its caste's *per capita* wealth distribution in our model. Second, data from multiple castes are available.

The family's *per capita* wealth is determined by household wealth and family size. Its relative wealth is based on not just its own *per capita* wealth, but also the *per capita* wealth of other members of the caste. The latter term is subsumed in the caste fixed effects, which we include in the estimating equation. Thus, the only threat to identification is that household wealth or family size, either independently or jointly, are correlated with other direct determinants of sex selection. Household wealth could determine access to ultrasound technology or investments in children's human capital (which differentially affect economic returns to having boys versus girls). Family size determines sex ratios when there is a demand for at least one son, as discussed above. Our solution to this identification problem exploits the fact that data from multiple castes, with distinct wealth distributions and within which independent marriage markets are organized, are available. This allows us to include a flexible control function, with household wealth and family size as arguments, in the estimating equation. By implementing this procedure, we are effectively comparing families with the same wealth and the same size, but which occupy different positions in their caste's *per capita* wealth distribution.

decline makes the sex ratios worsen earlier, resulting in an overall increase in the bias.

Figure 7: **Sex Selection and Relative Wealth (control function)**

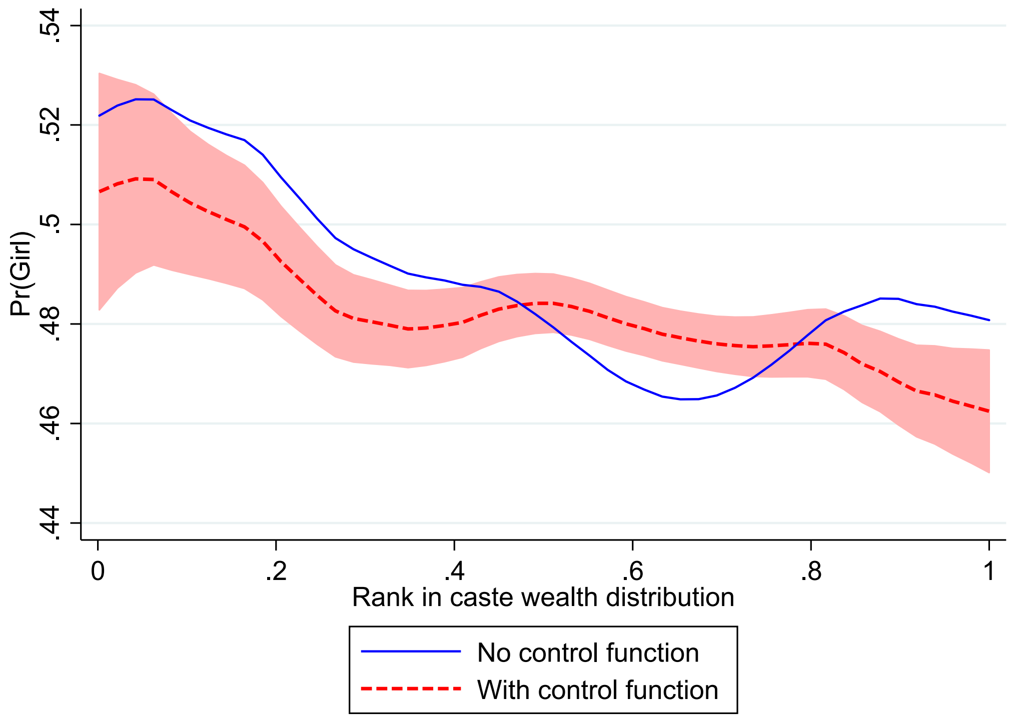
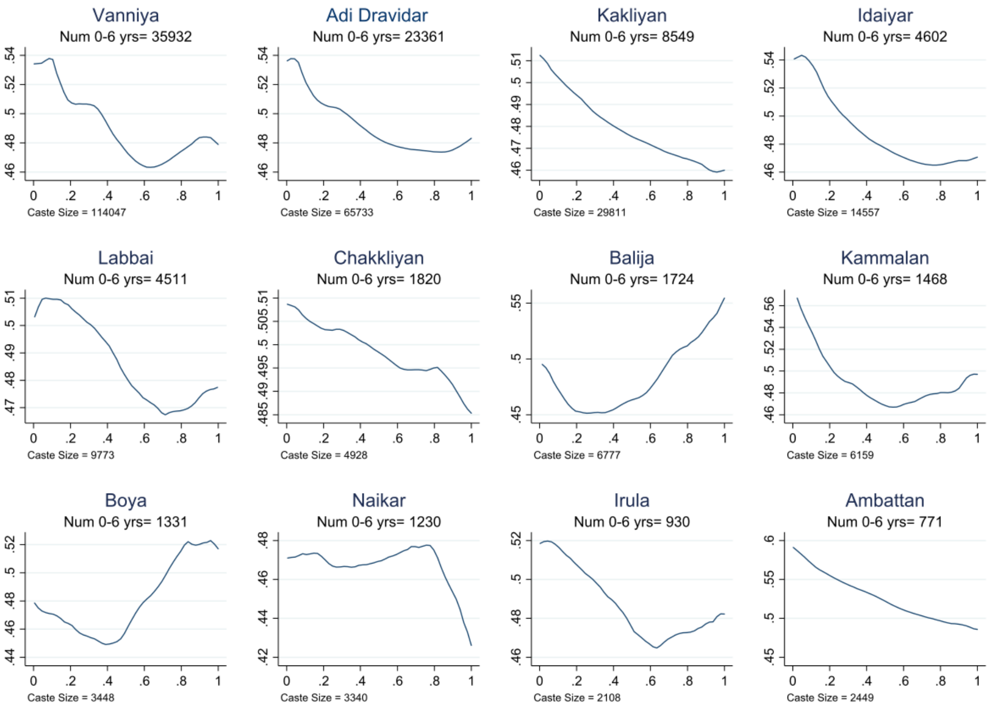


Figure 7 reports the relationship between sex selection and relative wealth, with and without the control function. The control function includes linear, quadratic, and cubic wealth terms, family size dummies (with three children; i.e. a family size of five, and above as the reference category) and a full set of interactions. The control function and the caste fixed effects are partialled out prior to estimating the relationship as usual. Inclusion of the control function does affect the estimated relationship, judging from the fact that estimates with the benchmark specification (with caste fixed effects alone) lie outside the 95% confidence interval with the augmented specification (which also includes the control function) at almost every point in the wealth distribution. However, the probability that a child is a girl continues to be decreasing in relative wealth; indeed, this decline is now observed at each point in the wealth distribution, as implied by the model.

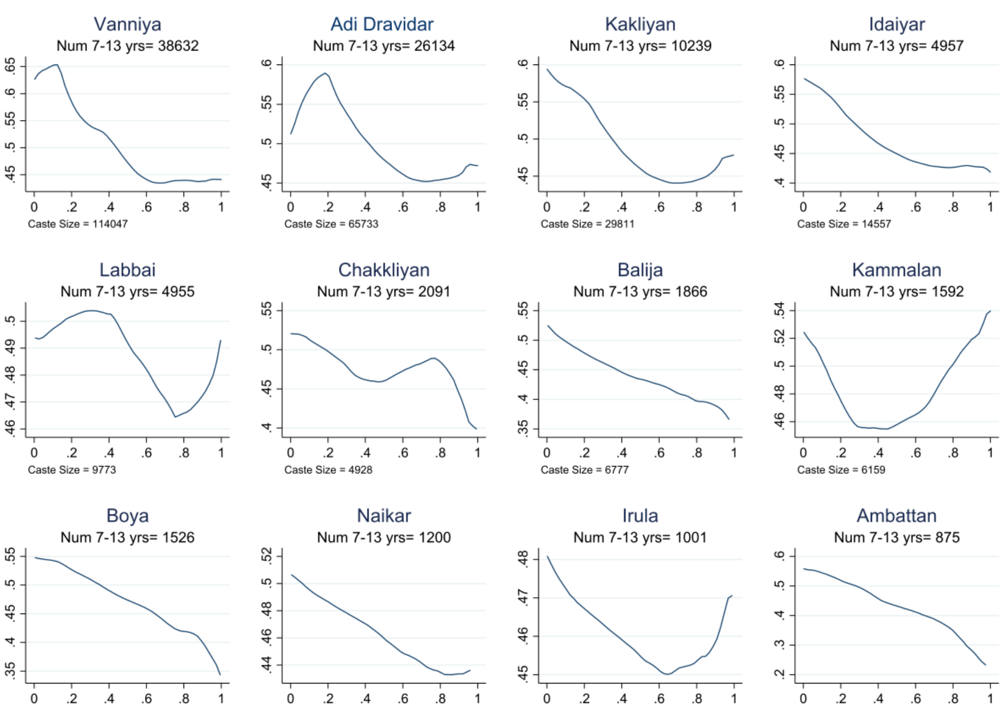
The advantage of including all castes in the analysis is that a flexible control function can be included in the estimating equation, which allows us to estimate the causal relationship between relative wealth and sex selection. An alternative test of the model is to nonparametrically estimate the relationship between relative wealth and sex selection, 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 8a reports this test for the 0-6 age group for the 12 largest castes, 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

Figure 8: **Sex Selection and Relative Wealth (12 largest castes)**

(a) Ages 0-6



(b) Ages 7-13



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 relationship between relative wealth and the probability that the child is a girl for the 7-13 year olds in Figure 8b. These older cohorts would also have been affected by sex selection and, thus, we expect to observe the same relationship for them. This is indeed what we find and, reassuringly, the relationship is negative at each point in the wealth distribution for the three castes for which anomalous patterns were observed in Figure 8a.

Two castes, the Vanniyas and the Adi Dravidars, 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. Despite their social differences, and in line with the view that the predictions of our model should apply to all castes, the probability that a child is a girl is decreasing with wealth within each of these castes. As a final robustness test, we estimated 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 A3, are very similar to what we obtain with the full sample of 0-6 year olds in Table 7, Column 2. There is a robust negative relationship between a family's position in its caste's *per capita* wealth distribution and the probability that a child will be a girl.

5 Structural Estimation and Quantification

5.1 Magnitude of Within-Caste Variation

A consistent finding from the preceding analysis is that the fraction of girls is decreasing as we move up the wealth distribution within castes. To quantify the magnitude of this variation, we partition the households with children aged 0-6 in each caste into eight 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, 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 0.23 when the sample is restricted to the 30 largest castes and 0.39 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 0.33 for the sample with 30 castes and 0.45 for the sample with 12 castes (the coefficient estimates for all specifications are reported in Appendix Table A4). This implies that within-caste variation accounts for 70% of the explained variation with the 30-caste sample and as much as 87% 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 eight wealth classes. Converting the fraction of girls to the number of boys per 100 girls, to be consistent with Government of India statistics once again, the sex ratio ranges from 97 to 117. The same range is obtained with the more flexible nonparametric estimates, with or without the control function, reported in Figure 7. To put these statistics in perspective, the sex ratio ranges from 115 to 120 in the three worst states in the country.

5.2 Structural Estimation and Counter-Factual Simulations

Our estimates indicate that the variation in sex ratios within castes is as large as the variation across states in the country. This variation is driven by marriage market frictions that apply to all castes. In addition, 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. Although our analysis provides new evidence on the extent of the sex selection problem, 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. Once the structural parameters have been estimated, counter-factual simulations with our model can be used to assess the impact of different schemes.

The estimation of the structural parameters, α and a, is straightforward. The algorithm we used to solve the model for given values of α and a was described above. To estimate the parameters, we search over all combinations of α and a to find the combination for which the predicted fraction of girls across the eight 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. Bootstrapped means and confidence intervals for the parameters are reported in Table 8, Column 2 by drawing repeated samples with replacement, where the probability that a particular caste is drawn is proportional to its size. Given the numerical dominance of a small number of large castes, the bootstrapped estimates will be largely determined by those castes. To assess the

Table 8: **Structural Estimates**

Sampling	Fraction of girls			
	Weighted		Unweighted	
	Actual	Predicted	Actual	Predicted
	(1)	(2)	(3)	(4)
Wealth class				
1	0.509	0.487	0.522	0.489
		[0.483, 0.491]		[0.484, 0.493]
2	0.499	0.481	0.494	0.483
		[0.475, 0.487]		[0.476, 0.490]
3	0.472	0.477	0.471	0.480
		[0.470, 0.484]		[0.472, 0.488]
4	0.465	0.476	0.475	0.479
		[0.469, 0.483]		[0.471, 0.488]
5	0.460	0.476	0.464	0.479
		[0.468, 0.483]		[0.470, 0.488]
6	0.468	0.475	0.474	0.478
		[0.467, 0.482]		[0.469, 0.487]
7	0.483	0.473	0.488	0.477
		[0.465, 0.481]		[0.468, 0.486]
8	0.477	0.473	0.471	0.476
		[0.465, 0.481]		[0.467, 0.486]
Parameters				
a	–	12.975	–	16.367
		[9.053, 16.897]		[8.009, 24.726]
α	–	0.613	–	0.620
		[0.589, 0.637]		[0.552, 0.688]

Source: SICHS census. Weighted estimates sample castes in proportion to their size. Unweighted estimates sample castes with equal probability. The 12 largest castes are used for the structural estimation. 95% confidence intervals in brackets.

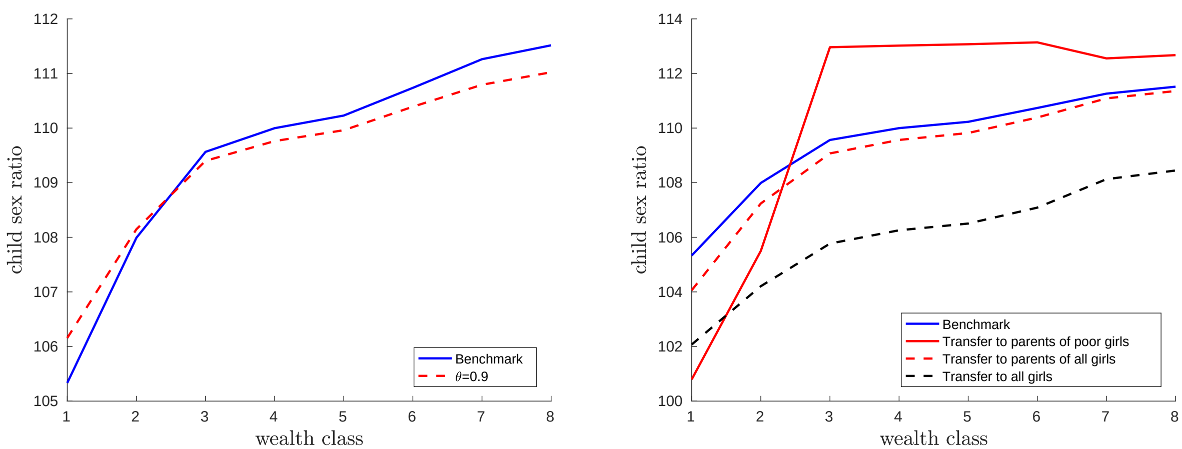
robustness of our results to alternative sampling procedures, we also report bootstrapped estimates in Table 8, Column 4 where all castes are sampled with equal probability.

The estimated α parameter is just over 0.6 with both sampling procedures, indicating that boys consume a greater fraction of the transfers from their parents than their wives. The a parameter is more sensitive to the sampling procedure, but the point estimate with each procedure nevertheless lies within the 95% confidence interval generated by the other. This is also true for the predicted fraction of girls in each wealth class. Despite the fact that the confidence intervals for the predicted fraction of girls are very narrow, the actual fraction (reported in Column 1 and Column 3) lies within the confidence interval for 3 of the 8 wealth classes, and just outside for the remainder.

Although dowries have been illegal in India since 1961, families can easily circumvent the law by claiming that the dowry is a gift. Given that the dowry is effectively the price for a boy, one potential policy instrument to reduce the (excess) demand for boys would be a gift tax on the dowry. As described above, such a tax would affect the welfare of both boys' and girls' families in ways that are difficult to determine analytically. To quantify the impact of a gift tax, we consider a policy that levies a 10% tax on the dowry; i.e. the θ parameter in the model is equal to 0.9. While the dowry tax does reduce the fraction of boys in the upper wealth classes, it has the opposite effect in the lower wealth classes in Figure 9a. One reason why this result is obtained is that as sex selection at the top of the wealth distribution decreases, the advantage of marrying up at the bottom is mitigated, and as a result sex selection there increases. This result emphasizes the fact that any policy intervention will affect sex ratios through the marriage market equilibrium channel, with potentially unintended consequences.

Figure 9: **Counterfactual Experiments**

(a) Tax (b) Transfers



The second set of policies that we consider are based on the Conditional Cash Transfer schemes that are currently in place. 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 or 20. In the framework of our model, a transfer to the girl's parent is equivalent to an exogenous increase in the wealth of the girl's family. Although governmental transfers when she is young go directly to her parents, the insurance payment, when it matures, goes directly into a bank account that is set up for the girl. Even though the money is in a bank account in the girl's name, however, it is not clear whether the insurance payment should be seen as a wealth transfer to the girl's family or a direct transfer to her. We allow for the latter possibility by examining a policy that provides a direct transfer to the married daughter (in addition to her share of the transfer that her husband receives from his parent). Although some of the welfare schemes are available to all families with girls, many are restricted to families below the poverty line. While a Conditional Cash Transfer program will encourage eligible families to have a girl, it will, in addition, affect all families in a caste by shifting the equilibrium marriage price (dowry). Our model, which allows for these pecuniary externality effects, is perfectly suited to examine the impact of programs with restricted eligibility.

The blue solid line in Figure 9b is the benchmark sex ratio (the number of boys per 100 girls) predicted by the model in each wealth class. The first counter-factual policy experiment that we consider 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 sex ratio declines 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 six wealth classes.[48] Combining all wealth classes, the net effect of this scheme is to *worsen* the overall sex ratio.

The next policy experiment that we consider 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 four (because the beneficiaries are now in 8 rather than 2 of the equal-sized wealth classes). Although the transfer now reduces the sex ratio bias in each wealth class, the magnitude of the effect is small. Moreover, part of the subsidy is transferred to boys' families via higher dowries.

The final policy experiment that we consider 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. Crucially, the 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 9b, there is now a substantial increase in the fraction of girls in each wealth class.[49] This is because the (optimal) 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 parents when they are children, may thus be especially effective in reducing the bias in child sex ratios in India.

[48] 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.

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

6 Conclusion

Sex selection continues to be a serious problem in India, despite many decades of economic progress. Much research and policy attention has been devoted to sub-populations in the country with severe sex selection and to explaining changes over time. The first contribution of our research is to document substantial variation in sex ratios in the cross-section along a new dimension; within castes or *jatis* which are the building blocks of Indian society. Our second contribution is to provide an explanation for this variation. We propose a model of assortative matching in the marriage market organized within each caste in which wealth-dependent sex selection arises endogenously. The root cause of sex selection at each level of wealth is specific imperfections in the marriage market, which arise due to the structure of the marriage institution in India. These imperfections are shown to affect relatively wealthy households in the caste more severely.[50]

We test the predictions of the model for wealth-matching, groom-prices (dowries), and sex selection across the wealth distribution within castes using unique data we have collected, covering a rural population of 1.1 million individuals in South India. Because data on multiple castes with distinct wealth distributions are available, a flexible control function approach can be implemented to estimate the causal relationship between relative wealth and sex selection. We find that the variation in sex ratios within castes in a single (unexceptional) district is comparable to the variation across all states in the country. The marriage market is organized the same way in all castes and, thus, we would expect this result to apply more widely. Sex selection may be more serious and more pervasive than currently believed.

The design of policies to reduce the sex selection problem assumes special significance in light of our findings. In order to evaluate the impact of alternative policy interventions, we estimate the structural parameters of the model using the data we have collected. Given these parameter estimates, the model can be used to quantify the impact of counter-factual policy interventions, which will work through the marriage market equilibrium to change patterns of sex selection. One class of policy interventions finds ways to reduce sex selection, taking as given the marriage market imperfections that generate the problem. A gift tax on the groom-price or dowry is a seemingly obvious solution to the problem by reducing the demand for grooms (boys). However, any intervention will work through the marriage market affecting the entire equilibrium price (dowry) schedule. Our counter-factual simulations indicate that a gift tax on the dowry would have mixed effects, reducing sex selection higher in the wealth distribution, but increasing it lower down.

A second set of policy interventions in this broad class, which have been implemented by the central government and many states, effectively shift the wealth distribution on the girls' side of the marriage market. Once again, these interventions affect the equilibrium prices in the marriage

[50] Rich girls marry boys of similar wealth, but the shortage of girls grows as we move down the wealth distribution because more and more boys are left unmatched above them. This implies that poorer girls match with relatively wealthy boys, reducing the incentive for sex selection.

market, and the effects are sometimes surprising. Cash transfers to less wealthy parents, conditional on having a girl, do generate an increase in the fraction of girls. However, there is a negative pecuniary externality on wealthier households in their caste, through the accompanying change in the dowry. This results in an overall worsening of the sex ratio. A transfer to all parents, conditional on having a girl, reduces sex selection across the wealth distribution. But the effects are small because the marriage market imperfections dampen the incentives of parents to change their behavior. Based on the counter-factual simulations, by far the most effective policy is to give wealth transfers directly to girls when they are adults; forward-looking altruistic parents will take these transfers into account and the resulting equilibrium price schedule leads to a substantial reduction in sex selection.

Ultimately, the most effective and the most sustainable class of policies would address the root causes of the sex selection problem: (i) the social norm that all girls must marry, and (ii) the mismatch between the actual consumption of the children and the preferred consumption level from the perspective of their parents. The mismatch is determined by the woman's bargaining power in her marital home (the α parameter in the model). One way to simultaneously shift the social norm and increase the woman's bargaining power would be to increase female labor force participation. With economic independence, marriage is no longer the only option, and the woman's bargaining power (conditional on being married) will also increase. Female labor force participation remains extremely low, even in South India, despite large increases in female education. Our analysis indicates that policies that increase female labor force participation would not only generate economic growth,[51] but also improve the sex ratio, through a new channel that has not been previously identified in the literature.

[51] Numerous studies, going back to Rosenzweig and Schultz (1982), have documented the positive effect of female labor force participation on sex ratios. Their interpretation of these findings is that female labor force participation is positively correlated across generations and that the accompanying increased economic returns to having girls reduces sex selection.

Appendix A Tables and Figures

Table A1: **Child Sex Ratios in India**

	Census Year			
	1991		2011	
Worst states				
	Haryana	114	Haryana	120
	Punjab	114	Punjab	118
	Delhi	109	Delhi	115
All India		106		109
South India		105		108

Source: Indian Population Census. Children aged 0-6.

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

Source: SICHS census. 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. *** $p<0.01$

Table A3: **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

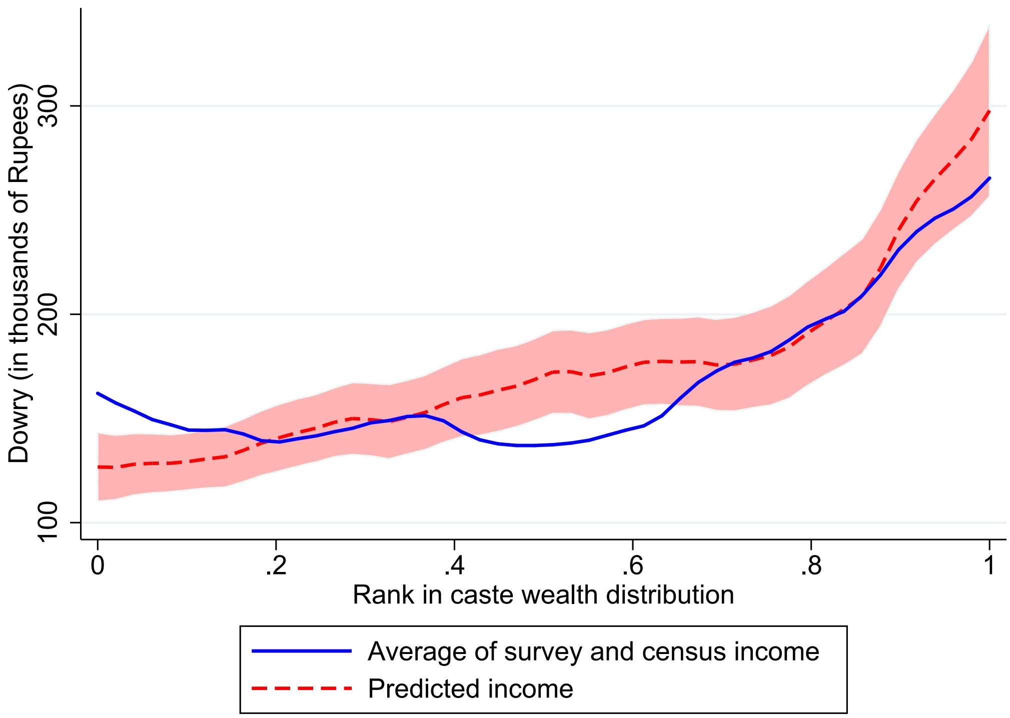
Source: SICHS census. Standard errors (in parentheses) are calculated on the basis of 500 bootstrap replications. All standard errors are clustered at the panchayat level. *** p<0.01

Table A4: **Within and Between Variation in Sex Ratios**

Dependent variable	Fraction of girls			
Sample	12 castes		30 castes	
	(1)	(2)	(3)	(4)
Wealth class				
1	0.0247***	0.0250***	0.0248***	0.0249***
	(0.00889)	(0.00759)	(0.00851)	(0.00745)
2	0.0139	0.0140	0.00973	0.00964
	(0.00978)	(0.00977)	(0.00979)	(0.0101)
3	-0.0119	-0.0117	-0.0138	-0.0136
	(0.0129)	(0.0117)	(0.0119)	(0.0109)
4	-0.0197**	-0.0195**	-0.0190**	-0.0194**
	(0.00848)	(0.00767)	(0.00824)	(0.00750)
5	-0.0232**	-0.0230***	-0.0233**	-0.0231***
	(0.00979)	(0.00824)	(0.00927)	(0.00802)
6	-0.0108	-0.0108	-0.00958	-0.00957
	(0.00905)	(0.00936)	(0.00859)	(0.00886)
7	-0.00204	-0.00208	-0.00129	-0.00187
	(0.0104)	(0.0111)	(0.00972)	(0.0104)
Constant (wealth class 8)	0.484***	0.484***	0.485***	0.485***
	(0.00789)	(0.00680)	(0.00751)	(0.00666)
Observations	96	96	237	237
R-squared	0.392	0.453	0.231	0.332
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: **Dowry and Relative Wealth (alternative methods to purge measurement error)**



Appendix B Omitted Proofs

B.1 Proof of Lemma 1

Proof. The solution to the girl's family maximization problem is to choose the optimal x that maximizes the family's utility. The first order condition is:

$$v_x + v_u u' = 0 \quad \Rightarrow \quad u' = -\frac{v_x}{v_u}. \tag{18}$$

Then the surplus is supermodular and the allocation will be PAM (see Legros and Newman (2007) and Chade, Eeckhout, and Smith (2017)) provided:

$$\frac{\partial^2 v(x,y,u)}{\partial x \partial y} = v_{xy} + v_{uy}u' > 0 \quad \Rightarrow \quad v_{xy} > \frac{v_x}{v_u}v_{uy}. \tag{19}$$

Calculating each of the derivatives, we get:

$$v_x = \frac{1}{x+y-2\psi(u)} \tag{20}$$

$$v_u = \frac{-2\psi'(u)}{x+y-2\psi(u)} + \frac{\psi'(u)}{\psi(u)} = \frac{x+y-4\psi(u)}{2(x+y-2\psi(u))} \tag{21}$$

$$v_{xy} = \frac{-1}{(x+y-2\psi(u))^2} \tag{22}$$

$$v_{uy} = \frac{\psi(u)}{(x+y-2\psi(u))^2}, \tag{23}$$

where we have used the fact that $\psi'(u) = \frac{e^{u/2}}{2\sqrt{\alpha}}$ and $\frac{\psi'(u)}{\psi(u)} = \frac{1}{2}$. Then there is PAM provided:

$$\frac{-1}{(x+y-2\psi(u))^2} > \frac{\frac{1}{x+y-2\psi(u)}}{\frac{x+y-4\psi(u)}{2(x+y-2\psi(u))}} \times \frac{\psi(u)}{(x+y-2\psi(u))^2} \tag{24}$$

or equivalently

$$\frac{x+y-2\psi(u)}{x+y-4\psi(u)} < 0 \tag{25}$$

The numerator is equivalently to $y-d$ (and the denominator to $(y-d)-(x+d)$). Clearly the girl's family cannot pay a dowry more than their wealth so the numerator here is positive. Hence, for PAM we must have $x+y < 4\psi(u)$, or the net wealth of the girls' family after giving dowry must be less than that of the boys after the transaction ($y-d < x+d$). This is true when:

$$x+y < 4\sqrt{\frac{e^u}{\alpha}} \tag{26}$$

or

$$u > 2\log\left(\frac{\sqrt{\alpha}}{4}(x+y)\right). \tag{27}$$

A sufficient condition for this to be satisfied is that the equilibrium payoff exceeds the outside option: $u(x) \geq 2\log\left(\frac{x}{2}\right)$ or

$$\frac{\sqrt{\alpha}}{4}(x+y) \leq \frac{x}{2} \tag{28}$$

$$\sqrt{\alpha}(x+y) \leq 2x \tag{29}$$

For $x = y = \overline{x}$, this is true when $\sqrt{\alpha} \leq 1$. Since $\alpha \in [0,1]$, we have PAM at the top for any α. For all other x, y the inequality is satisfied when $y < x\left(\frac{2}{\sqrt{\alpha}} - 1\right)$. This establishes the proof. ■

B.2 Proof of Proposition 2

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

$$k^{\star}(y) = u(y) - v(\mu(y), y, u(\mu(y))). \tag{30}$$

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

$$k^{\star'}(y) = u'(y) - \left(v_x \mu' + v_y + v_u u' \mu'\right). \tag{31}$$

From the first order condition (7), along the equilibrium matching $\mu(y)$, it must be that $v_x + v_u u' = 0$, so the derivative can be written as:

$$\begin{aligned} k^{\star'}(y) &= u'(y) - \left((v_x + v_u u')\mu' + v_y\right) && (32)\\ &= u'(y) - v_y(\mu, y, u(\mu)). && (33)\end{aligned}$$

This is increasing provided:

$$\frac{-2}{y + \mu^{-1}(y) - 4\psi(u(y))} - \frac{1}{\mu(y) + y - 2\psi(u(\mu(y)))} > 0, \tag{34}$$

since $u' = -\frac{v_x}{v_u}$ from the First-Order Condition (18) and substituting for v_x and v_u from equations (20) and (21), and v_y obtains from partially differentiating expression (6).

1. At the top of the wealth distribution. At $x = \overline{x}$, under positive sorting we have $\overline{y} = \mu(\overline{x}) = \overline{x}$. Then condition (34) can be written as:

$$\frac{-2}{2\overline{y} - 4\psi(u(\overline{y}))} - \frac{1}{2\overline{y} - 2\psi(u(\overline{y}))} > 0 \tag{35}$$

or

$$\begin{aligned} \psi(u) &< \frac{3}{4}\overline{x} && (36)\\ \sqrt{\frac{e^u}{\alpha}} &< \frac{3}{4}\overline{x} && (37)\\ u(\overline{x}) &< 2\log\frac{3\overline{x}\sqrt{\alpha}}{4}. && (38)\end{aligned}$$

Now we know that $u(\overline{x}) = \log\left(\frac{(\overline{x}+d)\sqrt{\alpha}}{2}\right)^2$. Therefore a sufficient condition for $k^{\star}$ increasing at $\overline{x}$ is:

$$\frac{(\overline{x}+d)\sqrt{\alpha}}{2} < \frac{3\overline{x}\sqrt{\alpha}}{4} \tag{39}$$

or $d < \frac{\overline{x}}{2}$.

2. At the bottom of the wealth distribution. We verify the condition at the lower bound $\underline{y}$, where $\mu(\underline{y}) = x^\star$. Now $\mu^{-1}(\underline{y})$ is not defined as it outside of the range of matches. But we know that for the unmatched men, the utility is $u(x) = 2\log\frac{x}{2}$ for all $x \in [\underline{x}, x^\star]$. Therefore $u'(x) = \frac{4}{x}$ in that range. At an income level $\underline{y}$, we can then write condition (33) as:

$$\frac{4}{\underline{y}} - \frac{1}{x^\star + \underline{y} - 2\psi(u(x^\star))} > 0. \tag{40}$$

Observe that the marginal man $x^\star$ is indifferent between the outside opinion and being matched, so that $u(x^\star) = 2\log\frac{x^\star}{2}$. After rearranging, we therefore obtain:

$$\alpha > \left(\frac{4x^\star}{4x^\star + 3\underline{y}}\right)^2. \tag{41}$$

This is satisfied for large enough α. $x^\star$ is a function of the distribution Γ and the parameters α and a, and $x^\star \in [\underline{y}, \overline{y}]$. So there is a critical $\tilde{\alpha} \in \left(0.327, \left(\frac{4\overline{y}}{4\overline{y}+3\underline{y}}\right)^2\right)$ such that for all $\alpha > \tilde{\alpha}$ sex selection is increasing at the bottom of the distribution. To see this, at $x^\star = \underline{y}$ this condition implies $\tilde{\alpha} = \left(\frac{4}{7}\right)^2 = 0.327$, and if $x^\star = \overline{y}$ then $\tilde{\alpha} = \left(\frac{4\overline{y}}{4\overline{y}+3\underline{y}}\right)^2$. ■

References

Anderson, S. (2003): "Why Dowry Payments Declined with Modernization in Europe but are Rising in India," *Journal of Political Economy*, 111(2), 269–310.

——— (2007): "The Economics of Dowry and Brideprice," *The Journal of Economic Perspectives*, 21(4), 151–174.

Anderson, S., and C. Bidner (2015): "Property Rights over Marital Transfers," *The Quarterly Journal of Economics*, 130(3), 1421–1484.

Arnold, F., M. K. Choe, and T. K. Roy (1998): "Son Preference, the Family-Building Process and Child Mortality in India," *Population Studies*, 52(3), 301–315.

Arnold, F., S. Kishor, and T. Roy (2002): "Sex-Selective Abortions in India," *Population and Development Review*, 28(4), 759–785.

Attanasio, O. P., and G. Weber (1993): "Consumption Growth, the Interest Rate and Aggregation," *The Review of Economic Studies*, 60(3), 631–649.

Basu, A. M. (1999): "Fertility Decline and Increasing Gender Imbalance in India, Including a Possible South Indian Turnaround," *Development and Change*, 30(2), 237–263.

Becker, G. S. (1973): "A theory of Marriage: Part I," *Journal of Political economy*, 81(4), 813–846.

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

Bhalotra, S. R., A. Chakravarty, and S. Gulesci (2016): "The Price of Gold: Dowry and Death in India," Essex Working paper.

Bhaskar, V. (2011): "Sex Selection and Gender Balance," *American Economic Journal: Microeconomics*, 3(2), 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(2), 129–148.

Blundell, R., M. Browning, and C. Meghir (1994): "Consumer Demand and the Life-Cycle Allocation of Household Expenditures," *The Review of Economic Studies*, 61(1), 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(01), 104–121.

Botticini, M., and A. Siow (2003): "Why Dowries?," *American Economic Review*, 93(4), 1385–1398.

Browning, M., P.-A. Chiappori, and Y. Weiss (2014): *Economics of the Family*. New York: Cambridge University Press.

Caldwell, J. C., P. H. Reddy, and P. Caldwell (1983): "The Causes of Marriage Change in South India," *Population Studies*, 37(3), 343–361.

Caucutt, E. M., N. Guner, and J. Knowles (2002): "Why do Women Wait? Matching, Wage Inequality, and the Incentives for Fertility Delay," *Review of Economic Dynamics*, 5(4), 815–855.

Chade, H., J. Eeckhout, and L. Smith (2017): "Sorting through Search and Matching Models in Economics," *Journal of Economic Literature*, 55(2), 493–544.

Chiappori, P.-A., M. Iyigun, and Y. Weiss (2009): "Investment in Schooling and the Marriage Market," *The American Economic Review*, 99(5), 1689–1713.

Chiappori, P.-A., and S. Oreffice (2008): "Birth Control and Female Empowerment: An Equilibrium Analysis," *Journal of Political Economy*, 116(1), 113–140.

Chiappori, P.-A., B. Salanié, and Y. Weiss (2017): "Partner Choice, Investment in Children, and the Marital College Premium," *American Economic Review*, 107(8), 2109–2167.

Choo, E., and A. Siow (2006): "Who Marries Whom and Why," *Journal of political Economy*, 114(1), 175–201.

Das Gupta, M. (1987): "Selective Discrimination against Female Children in Rural Punjab, India," *Population and Development Review*, pp. 77–100.

Doepke, M., and M. Tertilt (2009): "Women's Liberation: What's in it for Men?," *The Quarterly Journal of Economics*, 124(4), 1541–1591.

Doepke, M., and F. Zilibotti (2017): "Parenting with Style: Altruism and Paternalism in Intergenerational Preference Transmission," *Econometrica*, 85(5), 1331–1371.

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

Edlund, L. (1999): "Son Preference, Sex Ratios, and Marriage Patterns," *Journal of Political Economy*, 107(6), 1275–1304.

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(4), 1079–1092.

GREENWOOD, J., N. GUNER, AND G. VANDENBROUCKE (2017): "Family Economics Writ Large," *Journal of Economic Literature*, 55(4), 1346–1434.

GUILMOTO, C. Z. (2009): "The Sex Ratio Transition in Asia," *Population and Development Review*, 35(3), 519–549.

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(4), 687–712.

KAPADIA, K. (1993): "Marrying Money: Changing Preference and Practice in Tamil Marriage," *Contributions to Indian Sociology*, 27(1), 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(4), 1073–1102.

LINDENLAUB, I. (2017): "Sorting Multidimensional Types: Theory and Application," *Review of Economic Studies*, forthcoming.

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(3), 422–438.

MUNSHI, K., AND M. ROSENZWEIG (2016): "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *The American Economic Review*, 106(1), 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*, pp. 745–782.

NEELAKANTAN, U., AND M. TERTILT (2008): "A Note on Marriage Market Clearing," *Economics Letters*, 101(2), 103–105.

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(2), 239–268.

RAO, V. (1993a): "Dowry 'Inflation'in Rural India: A Statistical Investigation," *Population Studies*, 47(2), 283–293.

——— (1993b): "The Rising Price of Husbands: a Hedonic Analysis of Dowry Increases in Rural India," *Journal of Political Economy*, 101(4), 666–677.

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(4), 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*.

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(6), 1341–1371.