Premarital Investments in Physical versus Human Capital with Imperfect Commitment[*]

V. Bhaskar[†]

Wenchao Li[‡]

Junjian Yi[§]

November 21, 2019

Abstract

This paper empirically and theoretically studies how imperfect commitment within marriage affects premarital investments in children undertaken by their parents. Using nationally representative Chinese household survey data, we show that when the sex ratio is biased towards males, parents of boys, relative to those of girls, tend to migrate (a proxy for stronger earnings incentive), and increase housing investment at the expense of lower child educational investment. We interpret these patterns as a result of imperfect commitment: After marriage, labor earnings—determined by human capital—are bargained over with bargaining weights that do not depend upon marriage market conditions, while housing as a bequeathable physical capital is shared equally and thus more attractive. We then develop a model of premarital investments that incorporates imperfect commitment and two forms of capital, with a unique equilibrium. Model predictions when men are oversupplied and have great bargaining power after marriage, match empirical patterns.

Key words: Premarital investments; Human capital investments; Physical capital investments; Imperfect commitment; Human capital development

JEL Codes: J12; J13; J16; J18; J24; D10; O15; J61

---

[*]The data used in this paper are from the China Family Panel Studies (CFPS) survey, funded by the 985 Program and carried out by the Institute of Social Science Survey of Peking University.

[†]Department of Economics, University of Texas at Austin; E-mail: `v.bhaskar@austin.utexas.edu`.

[‡]Corresponding to Wenchao Li. School of Business and Management, Shanghai International Studies University; E-mail: `wenchao@shisu.edu.cn`; Address: 550 Dalian Road (W), 200083 Shanghai, China.

[§]Department of Economics, National University of Singapore; E-mail: `junjian@nus.edu.sg`.

1 Introduction

What determine premarital investments in children when marriage market considerations are important? Following Becker (1973), many economists have examined this question. While most classical work approaches the question in a transferable utility context (Chiappori et al., 2009; Cole et al., 2001; Iyigun and Walsh, 2007; Lafortune, 2013), which implicitly assumes full commitment at the time of marriage,[1] recent work begins to depart from such a context and make a more reasonable assumption of imperfect commitment (Anderson and Bidner, 2015; Galichon et al., 2019). Some theoretical studies look at an extreme case of imperfect commitment—non-transferable utility (Bhaskar and Hopkins, 2016; Peters and Siow, 2002).

The imperfect commitment assumption is particularly compelling in societies like China: Before marriage, prospective brides are in an enviable position due to high sex ratios (defined as the number of men per woman); while after marriage, divorce is prohibitively costly and thereby, the traditional power of husbands reasserts itself. Imperfect commitment comes from the divergence in the relative bargaining powers of men and women at the ex ante stage, before marriage, and ex post, after marriage. In such a situation, what are the consequences for premarital investments?

This paper empirically and theoretically studies how imperfect commitment within marriage affects premarital investments in children undertaken by their parents. The empirical analyses are in the setting of China's marriage markets. As Figure 1 shows, male fraction of births in China has been increasing, which foreshadows a sizable bride shortage. In an influential study, Wei and Zhang (2011) document a competitive saving motive as the sex ratio gets high, whereby parents increase investments in sons in a competitive manner to improve their sons' marriage market prospects. Many following studies show that sex imbalance—and the resulting marriage market competition—induce men to increase premarital investments relative to women. Whether sex imbalance modifies all components of premarital investments towards the same direction, however, has not yet been clear. This paper shows that the answer is no, given imperfect commitment within marriage.

We distinguish between two forms of capital in premarital investments: bequeathed physical capital (such as housing) and human capital. While it has been well documented that both

1. Browning et al. (2014) also give a full commitment example.

forms of capital enhance men's marriage market prospects, we propose that, given imperfect commitment within marriage, the former is more effective than the later, and a man's attractiveness as a marital partner depends not only on total investments, but also on the composition. Specifically, marriage partners are unable to commit, at the time of marriage, to share future household resources in a pre-agreed fashion. If a man invests in human capital, which will increase his future labor earnings, the sharing of this between spouses is determined by ex post bargaining, rather than ex ante bargaining. If a man invests in housing, which is a public good and thus non-excludable, spouses jointly consume it without bargaining. Therefore, when men have high bargaining power over their labor income after marriage, housing signals a credible commitment and thus is more favorable in a competitive marriage market. This creates an incentive for parents with sons to shift their investments towards housing and away from human capital when sex ratios are high.[2]

To investigate the consequences of imperfect commitment for premarital investments and in particular, its composition, we use data from a nationally representative Chinese household survey—the 2010 China Family Panel Studies (CFPS) survey. We examine the effect of high sex ratios at the county level for households with a first-born son, using those with a first-born daughter in the same county as a comparison group.[3] We control for county fixed effects to deal with unobservable heterogeneity. The identification partly relies on a well-recognized demographic regularity that the gender of the first child is plausibly random, while gender selections in China typically occur at second and higher order births (see Figure 1).

Empirical results show that when the sex ratio is high, parents of boys are more likely to increase labor supply and in particular, to migrate. In China, migration substantially raises family income, thereby permitting larger premarital investments in children. Further, the composition of investments is affected by sex imbalance, with the share invested in housing increasing relative to the share in children' education for parents with sons. Specifically, a 0.1 increase in the local sex ratio (the secular increase in China from 2002 to 2010, or the standard deviation of county-level sex ratios in 2010) is associated with a roughly 24.1 percent increase in the

2. We discuss in Section 2 the extent to which housing—the single most important piece of physical capital in Chinese families—can be considered as investments in preparation for marriage. Another example of premarital investments in physical capital is an upfront marital payment. Nunn (2005) builds a model in which bride price serves as a credible commitment from men to women.

3. The administrative divisions of China consist of four practical levels—province, prefecture, county, and community.

probability of having a migrant father, a 4.1 percent increase in house construction area, and a 11.0 percent decline in annual education expenditure per child, for first-son families relative to first-daughter families.

We show that our empirical findings are not mainly driven by differential family structures due to son-preferring fertility stopping rules—Chinese families are more likely to stop childbearing if they have a son. We then provide a comprehensive set of robustness checks to address the concerns about potential endogeneity of sex ratios. In particular, we use recent development in high-dimensional methods to select important sex-ratio confounders and control them in regressions to isolate the independent role of sex imbalance (Ahrens et al., 2018; Belloni et al., 2014; Mullainathan and Spiess, 2017). In addition, while focusing on the ordinary least squares (OLS) estimates, we show that estimations using the instrumental variable (IV) method yield similar results.

We interpret our empirical finding of differential effects of sex imbalance on investments in housing and education as a result of imperfect commitment within marriage, although alternative mechanisms may account for part of the results. This motivates us to develop a model of premarital investments that incorporates imperfect commitment and distinguishes between investments in a public good (bequeathed physical capital) and a private good (human capital). We assume that the two partners cannot commit to a future surplus sharing rule at the time of marriage; the public good is shared without bargaining after marriage, while the private good is shared by ex post bargaining with bargaining weights that do not depend upon marriage market conditions. We show existence and uniqueness of an equilibrium under minimal assumptions. We then show that the model allows us to characterize matching and investment equilibrium that includes as a special case an oversupply together with a high ex post bargaining power of men. In this case, model predictions are consistent with our empirical findings.

This paper contributes to the theoretical literature on premarital investments. These investments affect both future marriage matching and after-marriage resource allocation between spouses and therefore, have important intertemporal consequences (Choo and Siow, 2006; Iyigun and Walsh, 2007; Peters and Siow, 2002; Siow, 1998; Zhang, 2015). Different models of marriage matching and premarital investments have been developed to answer various types of research and policy questions (Chiappori et al., 2018; Low, 2017; Zhang, 2019). Relative to existing models, our model has two important innovations. First, imperfect commitment within the household (after marriage) is incorporated into premarital investment decisions. While most existing work assumes either fully-transferable utility—that is, perfect commitment—or an extreme case of non-transferable utility (an exception is Galichon et al. 2019), we allow for transferability constrained by imperfect commitment. The second innovation is to distinguish between private and public goods in premarital investments in the marriage market equilibrium, while most existing work assumes only one type of investments.

This paper is also related to the literature on limited commitment within the household and the implications for intertemporal resource-allocation decisions. Pioneering insights on limited commitment include separate spheres bargaining models in Lundberg and Pollak (1993, 1996) and Pollak (1985). Mazzocco (2007) presents evidence against perfect commitment and studies its implications for intertemporal consumption. Lise and Yamada (2018) show evidence consistent with limited commitment; see also Chiappori and Mazzocco (2017) who provide a discussion on household models with limited commitment. More relevantly, Anderson and Bidner (2015) develop a theory in which marital payment—with property rights that can be clearly defined and easily transferred at the time of marriage—acts as a more effective commitment device and therefore, is valued more than education in the marriage market. Our paper provides supportive empirical evidence in China's setting, showing that marriage market competition induces families with sons to invest more in housing to signal credible commitment at the expense of lower investments in children's education.

Finally, this paper contributes to the literature on sex imbalance in China and other societies.[4] Our analysis builds upon Wei and Zhang (2011), who show that parents facing high sex ratios competitively save more to improve marriage market prospects of their sons. We show that, for the same purpose, parents with sons shift the composition of investments towards housing and away from education. Such distorted investment patterns may hurt human capital development of young-generation men.[5] Taken into account parental migration induced by an incentive to get more resources for premarital investments, the detrimental effect of sex imbal-

4. This literature includes, among others, Bhaskar (2011); Bhaskar and Hopkins (2016); Cameron et al. (2017); Chiappori et al. (2002); Das Gupta et al. (2013); Ebenstein (2010, 2011); Edlund et al. (2013); Hu and Schlosser (2015); Sharygin et al. (2013); Wei and Zhang (2011).

5. This may be one of the reasons for the fact that while the sex ratio keeps rising in China, the level of high school enrollment rate of young men relative to young women is decreasing (Appendix Figure A3).

ance on boys' human capital is even larger.[6] We provide a detailed discussion in Appendix B, following the recent literature on human development to examine three types of human capital outcomes: cognitive skills, non-cognitive skills, and health (Cunha and Heckman, 2007; Heckman, 2007). Further, long-run costs are expected to follow, since early-stage human capital outcomes have cumulative impacts on late-stage achievements, which in turn affect lifetime productivity (Heckman et al., 2013). Therefore, the associated social loss is likely to be large, especially given the drastic increase in returns to schooling in recent years (Zhang et al., 2005). This implies that the social costs of sex imbalance have been understated.

The next section provides background information. Section 3 describes the data and empirical strategy. Section 4 reports empirical results and shows the robustness. Section 5 provides interpretations of empirical results and motivates Section 6, which sets up the model. The paper concludes with a brief summary and discussions about the implications of this study for human capital development of the next generation and for the study of marriage.

2 Background: Sex imbalance, marriage market, and premarital investments in China

In this section, we first describe sex imbalance and marriage market competition in China. We then discuss the extent to which housing can be considered as premarital investments in physical capital. We finally present evidence for imperfect commitment within marriage in China.

Sex imbalance The sex ratio at birth in China has increased drastically over time, from 1.12 boys per girl in 1990 to 1.2 in 2000. It has stabilized at that high level since then. In the current population under age 15, there are 13 percent more boys than girls. The imbalance primarily is due to sex-selective abortions, which in turn can be attributed to the traditional preference for sons (Ebenstein, 2010) and in part to China's family planning policy (Li et al., 2011). Specifically, parents undertake gender selections to satisfy their dual interests in having a son and complying with the birth quota stipulated by the policy.[7]

Marriage market competition High sex ratios at birth decades ago have led to an oversupply

6. Because of the strict household registration system in China, children of migrants are typically left behind in their hometown; see footnote 10. Based on the population census, there are more than 60 million left-behind children (Zhang et al., 2014).

7. Li and Pantano (2013) structurally estimate demographic consequences of gender-selection technology.

of marriage-age men in the market, which results in intensified competition for prospective brides and increased marriage expenditures facing parents with sons. The CFPS survey data show that household expenditure on marriage ceremonies increases by about 24 percent as the local sex ratio rises by 0.1 (see Appendix Figure A1 for graphical evidence). In particular, grooms' families are spending more on marriage over time, whereas brides' are less affected (see Appendix Figure A2 and Brown et al., 2011).

Premarital housing investment In China, housing as a form of bequeathable physical capital can be considered as investments in preparation for marriage. Wrenn et al. (2019) show empirical evidence for housing investment in China as a provision for marriage entry. In general, a marriage-age man or his family is required to cover the cost, or at least a substantial portion, of providing a home for the newlywed as a prerequisite to get married (Huang, 2010; Pierson, 2010; Shepard, 2016). Family housing wealth enhances a man's marriage market prospects: Typically, a man with more housing wealth is a more desirable partner in the marriage market.[8] In rural areas, a household is much less likely to have an unmarried adult son if they have a higher-quality house; in urban areas, a household is less likely so if they are a homeowner as opposed to a renter (Wei and Zhang, 2011).

In China, housing capital bought by the parents (and potentially used by the parents) at the time when the future groom is young, can be regarded as one for his marriage as well, for three reasons. First, housing purchased by parents is a form of investments in their children in the sense of the bequeathable nature of housing capital, which has a dominant role in household wealth composition (Xie and Jin, 2015). Second, a marriage-age man often has not yet accumulated enough wealth to afford a house on his own and needs his parents' assistance. Housing purchase usually is the most crucial component of household expenditure for children's marriage in China (Pierson, 2010). Third, intergenerational family coresidence is common in China, especially in rural areas. While in some cases the groom's family buys a new house for the couple, more than 70 percent of young adults move in the house of the groom's parents within the first few years after marriage, partly because of China's traditional patrilocal norms (Chu and Yu, 2010). Therefore, this paper focuses on family housing investment as premarital

8. In some personal interviews, most respondents shared that they would not like to get married if they still had to rent (Xinhua, 2011). According to recent surveys, nearly 70 percent of women indicated that housing consideration was a priority in choosing a husband (Beijing, 2013; Huang, 2010).

investments in physical capital.

It is also worth noting that while both family housing wealth and education grant marriage premium for men, the premium of the former turns out to be much higher than that of the latter in China. Analyses based on the population census data verify that better housing conditions improve a man's probability of finding a martial partner more effectively than high education (see Appendix Table A1).

Imperfect commitment within marriage Imperfect commitment within marriage is partly reflected by frictions in the marriage market—or more specifically, the difficulty in divorce. In China, the share of the population divorced and the divorce rate are sufficiently low for different genders, age cohorts, and education levels (see Appendix Table A2). As marriage is costly to reverse, marriage market conditions have little effect on the bargaining position of husband and wife within marriage. This indicates that once married, divorce is unlikely an outside option if within-marriage negotiation breaks down, consistent with the setting of separate spheres bargaining model which incorporates imperfect commitment (Lundberg and Pollak, 1993, 1996). Given that female bargaining power within the household is generally low and the enforcement of female alimony rights is generally weak in developing countries, a woman who is in a favorable position at the time of marriage will lose this advantage after being married. Such an asymmetry between ex ante and ex post bargaining power in marriage gives rise to imperfect commitment.

3 Data and regression model

To investigate how high sex ratios affect premarital investments and the composition, we use data from the CFPS survey. This section introduces the data, describes our main outcome variables, and presents the regression model.

3.1 The China Family Panel Studies survey

The CFPS survey is widely considered nationally representative due to its large sample size and advanced sampling design. The survey contains datasets with high-quality longitudinal information at the individual (both adult and child), household, and community levels. It consists of a total of 14,798 households, and includes 33,600 adults and 8,990 children who were successfully interviewed. The CFPS survey covers 645 communities in 25 designated provinces (out of 34 province-level units), representing 95 percent of the entire population in contemporary China (Xie, 2012).[9] In addition, this large-scaled survey implements a scientifically stratified multi-stage sampling design.

One strength of the CFPS survey lies in its comprehensive information. The family-level dataset contains details of family activities and household characteristics such as migration, expenditures, investments, income and wealth, as well as fertility. For each family surveyed, detailed information is available on demographic and labor-market characteristics of all family members, such as age, gender, schooling years, occupation, and working location, as provided in the CFPS individual-level dataset. The community-level dataset offers regional demographic and socioeconomic information. The datasets are linked across different levels by a set of identification numbers, and in particular, a household identification number allows us to group individuals by living unit. Parent-child relationship can also be precisely identified. Outcome variables of interest are thus readily linked with potential covariates, enabling systematic empirical analyses.

Sample construction Our empirical analyses are based on a cross-sectional sample of households drawn from the 2010 nationwide CFPS baseline survey—which provides the most comprehensive set of information on household activities compared with other waves—and exploit cross-county variation in the sex ratio. This does not render our analyses less strong, partly because variation in sex ratios across time is not that large within a county.

Specifically, we extract a sample of households in the 2010 CFPS family dataset in which the first-born child was between the ages of zero and 15 years, both parents were alive and at most 50 years old, and at least one parent participated in the adult survey. We focus on families in which the eldest child was under the age of 15 to minimize the possibility that the children have started work or participated actively in household decision-making. We impose a constraint on the ages of parents to minimize the probability of their retirement or their ineffectiveness in making investment decisions due to, for example, health reasons. By placing age limits on both parents and children, we maximize comparability across families. The main sample contains 4,314 observations.

9. Hong Kong, Macao, Taiwan, Xinjiang, Tibet, Qinghai, Inner Mongolia, Ningxia, and Hainan are not included.

3.2 Main outcome variables

Our empirical analyses mainly use three sets of outcome variables: parental labor supply (as a proxy for stronger earning/investment incentive), housing investments, and child educational investments.

Parental labor supply We construct five measures of parental labor supply. The first three are, respectively, a binary variable that equals one if the father, mother, and at least one of them, works away from the hometown. The information comes from a question in the CFPS family survey that asks whether any member in the family works in a place that is not where the household is registered or where its permanent address is. According to China's household registration system that is used to differentiate permits of where people are allowed to live and work—the hukou system—it is prohibitively difficult for migrants to assimilate with the local population.[10] Instead, migrants usually leave home temporarily to increase their earnings. This kind of circular migration is considered a crucial form of labor supply in China (Zhao, 1999), which is typically associated with a large increase in gross family income (see Appendix Table A3); migration remittances sent back by migrant family members can be used to increase total premarital investments. As Panel A, Table 1 shows, approximately 9.8 percent of fathers and 2.5 percent of mothers in our sample work outside their hometown; 11 percent of households have at least one migrant parent.

The other two measures we examine are yearly working hours of the father and mother, information on which is from the CFPS adult dataset. More working time generally is associated with more labor income, and therefore, can ease the household budget constraint and increase total premarital investments in children. In our sample, the average father works 2,466 hours per year, and the average mother works 2,416 hours per year.

Housing investments We have discussed in Section 2 that in China, parents see their property as investment in their sons' physical capital in preparation for marriage, even at the time when their sons are young. We construct three variables from the CFPS family dataset for housing investments: construction area, an ownership dummy, and mortgage debts. The ownership dummy indicates whether the family owns the property right of any house, and equals one

10. The hukou system results in institutionalized discrimination against migrants: They have limited access to various benefit schemes that are available to local residents, and their children are often denied access to public schools (Zhao, 1999).

if the property deed and other relevant contracts of one or more houses belong solely to this family; self-constructed houses in rural regions are also counted. Construction area refers to the area for residential use, and is specified for home owners. A larger construction area or a higher mortgage signals houses of higher quality, and typically is indicative of greater housing investments. Panel B, Table 1 shows that 83.1 percent of the families in our sample own a house. Mean construction area is 126 square meters among homeowners. An average household has a mortgage of ¥5,392.

Child educational investments We construct two variables of child educational investments, focusing only on the first-born child in the family. The first, education expenditure, is yearly total expenses on the child's education, including tuition fees, book and stationery costs, after-class tutoring expenses, accommodation fees, and commuting fees, yet excluding living expenses. The second variable, an education funding dummy, equals one if the family has put aside a specialized fund for the child's education. Both variables are defined for children who are at least two years old, and the information comes from the CFPS child dataset. Panel C, Table 1 presents that the average yearly education expenditure is ¥1,507 for the first child, and 29.7 percent of families have been preparing an education fund for the child.

3.3 Regression model

We estimate the following regression model:

$$k_{ic} = \beta_0 + \beta_1 FirstSon_{ic} + \beta_3 FirstSon_{ic} * SexRatio_c + X_{ic}\Gamma + \lambda_c + \epsilon_{ic}, \tag{1}$$

where k_{ic} represents outcome variables for household i in county c (parental labor supply, housing investments, and child educational investments), $FirstSon_{ic}$ is a binary indicator that equals one if the first-born child in the family is a boy, and $SexRatio_c$ refers to the county-level sex ratio. A vector of additional control variables, X_{ic}, includes various parental and household characteristics: both parents' age, schooling years, hukou status, political status, plus age of the first child, region of residence, and ethnicity (column 1 of Table 2 reports summary statistics for main control variables).[11] Regressions also control for county fixed effects, λ_c, to account

11. In robustness checks, we include different sets of controls. For example, we add the number of children, average household income, among others.

for unobservable cross-county heterogeneity. The error term is denoted by ϵ_{ic}.

We assume that parents infer the local sex ratio from the premarital-age cohort. In our main regressions, we use sex ratios for those between the ages of ten and 24 years, which are obtained from the 2010 population census.[12] Prior work has shown that empirical findings appear insensitive to using sex ratios for different age brackets (Wei and Zhang, 2011). Later we check whether this is the case in our study (Section 4.4). Sex ratios are. Sex ratios are at the county level, as each county can be treated as a local marriage market. China's hukou system presents a formidable obstacle to marriage migration (Davin, 2005; Wei and Zhang, 2011). The census shows that more than 90 percent of rural residents and 62 percent of urban residents live in their county of birth and, 89 percent of couples are from the same county. Of migrant couples in cities, 82 percent are from the same place, suggesting that migrants often get married before leaving their hometown.

We focus on the interaction-term coefficient β_3, which measures the effect of high sex ratios for first-son families relative to first-daughter families on outcome variables of our interest. For example, a positive estimate of β_3 when the outcome variable is parental migration, as we may expect, suggests that when sex ratios are high, parents of boys have a desire to earn more and invest more in children than parents of girls. The sign and magnitude of the estimate of β_3 when the outcome variables measure housing investments and child educational investments tell how sex imbalance affects premarital investments in different forms of capital.

3.3.1 Identification assumptions

Obtaining unbiased ordinary least squares (OLS) estimates of β_3 requires that in equation (1), the error term is not substantially correlated with the interaction between the first-son dummy and the local sex ratio.

The identification partly relies on a well-recognized demographic regularity: The first child in a family being a boy or a girl is plausibly random. Data from China population censuses (1982, 1990, 2000, and 2010) reveal that high sex ratios in China are driven by imbalances between second- and higher-order births, while the sex ratio for first births is rather stable and falls in the biologically normal range; see Figure 1. Parents are least likely to practice

12. Instead of knowing the exact local sex ratios, parents are more likely to estimate such statistics based on the experiences of their relatives' or colleagues' marriage-age children in finding partners, or the prevailing marriage expenditure that signals the level of competitiveness in the marriage market.

gender selection on the first birth, despite their son preferences. Before 2015, a second child was officially permitted if the first one was a girl for households in most rural areas, where son preferences appear stronger. This "1.5 children" policy was applicable to residents who accounted for more than 60 percent of the Chinese population—among whom son preferences appear more common—and markedly alleviated their motivation to abort the first daughter.[13]

Statistical evidence from our sample also validates the randomness of first-child gender. An average family in our sample has 1.5 children, consistent with the above-mentioned policy. Nearly half of the families have a first son and the other half have a first daughter. Specifically, the mean of the first-son dummy is 0.507, which implies a sex ratio of 1.03, well within the normal range; the standard deviation is 0.5, which suggests that first-child gender is like a random Bernoulli trial in which having a boy or a girl has an equal probability;[14] see Table 2. In addition, we regress the first-son dummy on the full set of control variables used in our analyses, and find no significant effect of these variables.

The strongest evidence in favor of the randomness of first-child gender is that first-son and first-daughter families have similar predetermined parental and household characteristics in our data, as the balance test in Table 2 shows. For example, 12.1 percent of first-son families and 12.8 percent of first-daughter families belong to minority ethnic groups. The difference is -0.007, which is statistically indifferent from zero at the ten percent level or below.

In the next section, we will present a broad range of robustness analyses to test whether our empirical results are driven by identification issues, which mainly come from the potential endogeneity of local sex ratios.

4 Empirical results

This section presents empirical results on how sex imbalance affects parental labor supply, housing investments, and child educational investments for families with a first-born son relative to those with a first-born daughter. The results are shown to be robust to potential issues related to son-preferring fertility stopping rules and potential endogeneity of local sex ratios.

13. The policy was replaced by a nationwide two-child policy in 2015, which further alleviates the motivation to abort the first daughter. Ebenstein (2010, 2011) shows that most Chinese families prefer one boy and one girl to two boys.

14. A Bernoulli random variable with a mean of 0.5 has a standard deviation of 0.5.

4.1 Sex imbalance and parental labor supply

We begin our empirical analyses by estimating equation (1) based on our sample, using measures of parental labor supply constructed in Section 3.2 as outcome variables. Results are reported in the first panel of Table 3. In these and the following estimations, we mainly use the OLS method;[15] estimations are weighted by the CFPS survey sampling weights; standard errors are clustered at the county level and given in parentheses.

In column (1) in which the outcome variable indicates father's migration, the estimate is 0.235 (standard error 0.094) for the coefficient on the interaction between the first-son dummy and sex ratio, β_3, as reported at the top of the panel.[16] The positive sign and statistical significance (at the five percent level) of the estimate suggest that a high sex ratio is much more likely to induce the migration of fathers of a first-born son relative to fathers of a first-born daughter. Specifically, the estimate implies that with an increase in the sex ratio for adolescents from 1.08 in 2002 to 1.18 in 2010, the probability of having a migrant father, on average, significantly increases by 2.4 percentage points for a first-born boy relative to a first-born girl, all else held equal. This difference is economically significant, representing a 24.1 percent difference relative to the baseline father-migration probability of about 0.1 in our sample, as reported in the middle of the panel. (Appendix Figure A4 presents graphical patterns that are consistent with results here.)

We obtain qualitatively similar findings for mother's migration and the migration of at least one parent, according to columns (2) and (3) of panel A, Table 3. The effect of sex imbalance for first-son families relative to first-daughter families is substantial in both percentage point and percentage terms. For example, a 0.1 increase in the sex ratio would increase the probability of a first boy's mother, relative to a first girl's mother, working outside the hometown by about one percentage point or 38.6 percent. In addition, working time increases more with a rise in the sex ratio for parents of first sons than parents of first daughters, as shown in columns (4) and (5). These results support that relative to first-daughter families, high sex ratios boost parental labor supply among first-son families.

As we discussed earlier, migration significantly increases family income, and so does in-

15. When the outcome is binary, OLS estimates are consistent with marginal effects from Probit models in most empirical practices (Angrist and Pischke, 2008). We have verified that our study is no exception.

16. The estimates of coefficients on other control variables, which are unreported for brevity, have the expected sign and magnitude.

creasing working time. Results in this section indicate that parents attempt to increase total premarital investments in their children when marriage market conditions are disadvantageous, as indicated by having a first-born son together with facing high sex ratios.

4.2 Sex imbalance and premarital investments

We next examine how sex imbalance affects premarital investments in physical capital and human capital. We estimate equation (1) based on our sample, using measures of housing investments and child educational investments constructed in Section 3.2 as outcome variables. Results are reported in panel B, Table 3.

Housing investments We focus on housing investments when considering premarital investments in bequeathed physical capital, for reasons given in Section 2. Housing investment results are reported in the first three columns of panel B, Table 3. In column (1) where the outcome variable is (log) house construction area, the estimate for the interaction-term coefficient is 0.413 (standard error 0.205), positive and statistical significant at the five percent level. This indicates that parents with first-born sons prepare a much larger house relative to those with first-born daughters as the sex ratio becomes more biased towards males. Based on the estimate, as the sex ratio rises from 1.08 in 2002 to 1.18 in 2010, house construction area for home possessors with a first son would increase by about 4.1 percent relative to home possessors with a first daughter.

In column (2) that concerns housing ownership and column (3) that concerns housing mortgage loan, the estimates of the interaction-term coefficient are positive and significant in terms of both statistical sense and economic magnitude (see also Appendix Figure A5 for graphical evidence). All these results imply that, in the presence of a high sex ratio, parents with first-born boys become considerably more aggressive in investments in housing relative to parents with first-born girls.

Child educational investments Results on child educational investments, reported in the remaining two columns of panel B, Table 3, are in contrast with the patterns for housing investments. In column (4) where the outcome variable is annual education expenditure for the first-born child in the family (in thousand), the estimated interaction-term coefficient is -1.663 (standard error 0.800), negative and statistically different from zero at the five percent level.

Accordingly, with a 0.1 increase in the sex ratio, annual education expenditure is ¥166 less for a first-born boy relative to a first-born girl. The economic magnitude is sizeable compared with a mean expenditure of ¥1,507 per child, representing a 11.0 percent difference (see Appendix Figure A6 for graphical evidence). Result in column (5) shows that a high sex ratio reduces the probability that parents with a first-born boy have put aside a specialized fund for his education relative to parents with a first-born girl.

Taken together, empirical results in panel B, Table 3 show that the effects of sex imbalance on the composition of premarital investments are in opposite directions: A combination of having a son and experiencing a scarcity of prospective brides induces more physical capital investments but less human capital investments. Intuitively, parents with sons are motivated to invest their available financial resources in the capital form that can more effectively enhance their sons' marriage market prospects.

4.3 Robustness: Potential issues related to son-preferring fertility stopping rules

While we have shown in the previous section that the gender of the first-born child can be viewed as random, such an argument for identification is potentially not quite complete: Due to son preferences, subsequent fertility decisions may be different depending on the gender of the first child. Families with a first-born daughter typically are more likely to have a second or third child in order to get a boy, while families with a first-born son are more likely to stop childbearing and therefore, have a smaller family (Ebenstein, 2011). This section shows that these concerns do not confound our main results.

Family size One might worry that our findings on parental labor supply decisions, housing investments, and educational investments per child may reflect the the effect of family size. Specifically, it is possible that fewer resources are available for a first-born girl—who may have a sibling—compared with a first-born boy—who may have no sibling, as implied by the theory of child quantity-quality trade-off (Becker and Lewis, 1973). This is what our results on housing investments show; results on educational investments for the first child, however, show the opposite. That is, our empirical results are not fully in line with this interpretation.

To further address the concern with the family-size effect, we control for the total number of children in a family in robustness checks. Table 4, panel A reports the results. The outcome variables in columns (1), (2), and (3) are the paternal-migration dummy, house construction area, and education expenditure for the first child, respectively (using other measures of parental labor supply and premarital investments yields similar patterns). Results after controlling for family size are not significantly different from the baseline (the first row of the table). In particular, we perform a generalized Hausman test to formally show that the estimated coefficient on the interaction term is statistically equivalent to the baseline.

We then control for the number of children plus its interaction with the first-son dummy, to take into account the possibility that family size may have differential effect depending on the gender of the first child. We still get similar results.

An alternative strategy to deal with the potentially confounding family-size effect is to restrict the sample to families with only one child (about 65 percent of the main sample). Results are reported in Table 4, panel B and show a similar pattern to the baseline in terms of the sign and magnitude of the estimated interaction-term coefficient. The pattern remains similar when we further restrict the sample to families that are less likely to have a second child. Specifically, we restrict the sample to one-child families in which the only child is above the age of four (about 50 percent of the main sample). We note that for these two groups of families, the effect on housing investments is less pronounced than the baseline, while the effect on child educational investments is more pronounced.

Marriage market conditions Another issue raised by using the gender of the first child as a regressor is that it may not be an appropriate proxy for marriage market conditions the family faces. As we have discussed, it is common in China that a first daughter is followed by a second son. For these families, despite having a first daughter, they have to worry about the marriage prospects of their second son—similar to first-son families. To isolate the effect of marriage market conditions, we use variables that better represent actual child-gender composition in a family. Specifically, we replace the first-son dummy in equation (1) with a dummy for having any son and the proportion of sons. This yields qualitatively similar results as before.[17] We then control for the number of children (and its interaction with the child-gender measure) in

17. When we use a dummy for having any son, the estimates compare the effect for families with at least one son with families with no son. When we use the proportion of sons, the estimates compare the effect for all-son families with all-daughter families.

these empirical exercises, and again obtain similar results; see Table 4, panel C.

While these patterns confirm that how families make labor supply and premarital investments decisions in response to high sex ratios depends on child gender, the two child-gender measures we use may be endogenous. To partly address this issue, we use the first-son dummy as an instrument for the two variables and repeat all empirical analyses above (first-stage results are given in Appendix Table A4). We observe that this does not change the pattern of results.

In summary, robustness analyses in this section show that concerns related to son-preferring fertility stopping rules are not likely to be the main driver of our findings.

4.4 Robustness: Potential issues related to local sex ratios

We have followed the practice of focusing on OLS estimates for the effect of sex imbalance (Edlund et al., 2013; Wei and Zhang, 2011). But strictly speaking, county-level sex ratios may not be exogenous. Below we provide evidence that helps alleviate this concern and isolate the marriage market effect of sex ratios. We also provide evidence that our results are not sensitive to using sex ratios for different age cohorts.

Unobserved county-level characteristics Counties with higher sex ratios perhaps have unobserved characteristics—for example, culture—that may affect household decisions like premarital investments. This concern is partly addressed in our research design, as we control for county fixed effects and focus on the coefficient on the interaction between local sex ratio and first-child gender. That is, we compare first-son families with first-daughter families, which reduces the confounding effects of unobserved county-level characteristics as long as they affect premarital investments of the two types of families within a county in a similar manner.

To check the extent to which unobserved cross-county heterogeneity is an issue in our estimations, we exclude county dummies, which are previously controlled for (to saturate the model, we include county-level sex ratios). Results are reported in Table 5, panel A, showing a similar pattern to the baseline (the first row of the table). This suggests that county-level unobservables might not be an important issue in our data.

Potential sex-ratio confounders If certain factors are correlated with the sex ratio and affect premarital investments of first-son and first-daughter families in a different manner, our estimates may be biased. Below we discuss some possible factors and check whether they play an important role in generating our results.

We first consider son preferences. People in counties with higher sex ratios have on average stronger son preferences than those in counties with balanced sex ratios. In the latter type of counties, parents may stop childbearing after the first child regardless of the child's gender. In counties with high sex ratios, parents who have a son as the first child also stop, while those who have a daughter may have a second child to get a son. This may lead to differential family structures between high- and balanced-sex-ratio regions. We have discussed in detail in Section 4.3 that issues raised by this son-preferring fertility stopping rule are not a primary concern in our results. We also consider the fact that some wealthier areas of China may retain stronger demand for sons. This motivates us to control for average household financial wealth—defined as the sum of liquid and illiquid assets—and household income at the community level (the sub-level of county) in robustness regressions.

The second factor we examine is gender difference in earnings. We therefore control for a community-level gender wage differential.[18] The third factor is social old-age support, a lack of which increases the demand for sons (sons serve as a better source of insurance against old age than daughters in China). We accordingly control for a variable indicating social insurance at the household level. The fourth factor is the implementation of China's family planning policy, which varies from place to place. As fertility is the direct target of this policy, and thus can be regarded as a proxy for the implementation, this can be addressed by controlling for the number of children. The fifth factor is technological development, and particularly gender-selection technology. This again can be proxied by local average household wealth or income. The last factor is grandparental coresidence, which may be correlated with both sex ratios and various household decisions. For each of these factors, we allow it to affect first-son and first-daughter families differently, by controlling for its interaction with the first-son dummy.

Table 5, panel B reports the robustness results using the three representative measures as the dependent variables. It shows that controlling for these potentially confounding factors—either individually or collectively—leads to a very small difference in our results. In each robustness regression, the estimated coefficient on the interaction term β_3 is not significantly different from the baseline estimate.

18. Alternatively, we can control for gender wage ratio. This does not make much difference.

Sex-ratio confounders selected by high-dimensional method The above robustness analyses discuss potential sex-ratio confounders based on traditional economic reasoning. The current development in methods with high-dimensional data enables us to consider a much more comprehensive set of sex-ratio confounders, and select the most important ones with the help of machine learning (Ahrens et al., 2018; Belloni et al., 2014; Mullainathan and Spiess, 2017).

In this robustness analysis using high-dimensional method, the initial set of variables that are considered as being potentially correlated with the sex ratio include local residents' age, schooling years, hukou status, political status, marital status, region of residence, the number of siblings, ethnicity, income, social insurance, scores for a word test and a math test, depression score, and coresidence with parents. We consider the average, the average for men, the average for women, and the gender difference at the county level. The final set consists of 363 variables made up of the levels and quadratics in each of the initial variables, and interactions of all the preceding variables with each other, as in the example discussed in Belloni et al. (2014). We regress county-level sex ratios on these variables and use high-dimensional methods to select potentially important ones—those are strongly predictive of local sex ratios. Then we control for the selected variables and their interactions with the first-son dummy in estimating equation (1). In this way, we take into account factors whose effects are most likely to be confounded with the effect of sex imbalance. Results show that after partialling out the confounding effects, sex imbalance still has a significant role in accounting for the main empirical patterns; see the last part of Table 5, panel B.

Together, results reported in Table 5, panel B show that our findings are robust to the inclusion of various potentially confounding factors and are mainly driven by marriage market considerations. This implies that to a very small extent omitted variable bias is an issue, and that the potential endogeneity of sex ratios may not be a primary concern (Altonji et al., 2005).

IV estimations To further alleviate the concern about the possible endogeneity of sex ratios, we use the IV estimation method to estimate equation (1) as a robustness check. Cross-county variation in sex ratios may be accounted for by variation in financial penalties for violating the family planning policy and quota of births stipulated by the policy (depending on whether the household belongs to ethnic minorities, since ethnic minorities are generally exempted from the policy). Therefore, the interactions of these variables (as well as their interactions with a dummy for ethnic minority) with the first-son dummy are used as instruments for the interaction of the sex ratio and the first-son dummy in equation (1).

As reported in Table 5, panel C, IV regressions results reveal qualitatively similar patterns to OLS estimates in the benchmark. Results regressing sex ratios on policy-violation penalty, birth quota, and their interactions with a dummy for ethnic minority—in lieu of first-stage results—suggest that heavier financial penalties levied for unauthorized births and fewer births allowed by the policy are associated with higher sex ratios (see Appendix Table A5). This is consistent with the notion that more stringent enforcement of the family planning policy leads to more aggressive gender-selective abortions (Li et al., 2011).

We, however, exercise caution in interpreting the IV results and still focus on the OLS results, to avoid problems with common candidates for instruments of sex ratios such as the ones used here. China's family planning policy is passed down the administrative chain of command until it is interpreted and adapted to suit local needs (Short and Zhai, 1998). Therefore, financial penalties and birth quotas are likely to be endogenously stipulated by local governments based on local conditions, and may be correlated with various household decisions independent of the local sex ratio (Ebenstein, 2011; Wei and Zhang, 2011).

Sex ratios for alternative age cohorts We also check whether our findings are robust to using sex ratios for a particular age cohort. Instead of the premarital-age cohort between the ages of ten and 24 years, we recalculate sex ratios for local population in the age brackets of 10–14, 15–19, and 20–24. We find that using sex ratios for each age bracket gives a qualitatively similar pattern of estimates for the interaction-term coefficient to the benchmark: The sign is preserved and the magnitude varies only moderately (panel D, Table 5). This is perhaps due to the persistence of the local level of sex-ratio distortion over time. If we take any potential measurement error in sex ratios—which tends to produce attenuated coefficient estimates—into account, our results may be interpreted as lower bounds of the true effect of sex imbalance.

5 Interpretations of results

We have shown that when sex ratios are high, parents of boys, relative to those of girls, tend to increase labor supply, and shift investments toward housing and away from children's education. In the following, we provide plausible interpretations of these results and also discuss competing hypotheses.

In line with prior literature, we propose that parents facing steep marriage odds due to sex imbalance increase labor supply—and in particular, work away from home—in a competitive manner in order to increase total resources available for premarital investments in their children. Further, we propose that the effect of sex imbalance on the composition of premarital investments is because of imperfect commitment in marriage. As men and women are unable to commit, at the time of marriage, to share future household resources in a pre-agreed fashion, the future labor income will be subject to ex post bargaining, where bargaining power depends on who earns the income. Therefore, a man who brings with him more housing at the time of marriage—which will not be subject to ex post bargaining—is a more desirable marriage partner than one with higher labor earnings but a smaller house. This explains why parents who want to ensure the marriage of their son direct investments towards more housing than education.

One competing hypothesis centers around the possibility that the difference in outcome between first-son and first-daughter families is affected by factors other than sex imbalance—such as household structures. In high-sex-ratio regions where son preferences are stronger, a first-born girl is more likely to have sibling(s) relative to a first-born boy, while in low-sex-ratio regions, the first child may be the only child regardless of the gender. Possibly, parents with more children have to devote less time to wage earning in the labor market, and in particular, are less likely to work away from home; they are also poorer and have less residual wealth to invest in real estate, as more children dilute household resources. Although in this respect, the hypothesis is consistent with part of our empirical findings, we have provided various robustness analyses in Sections 4.3 and 4.4 to verify that our results are not mainly driven by issues raised by son-preferring fertility stopping rules. In particular, our results on human capital investments show that when sex ratios are high, parents tend to invest less on a first-born boy relative to a first-born girl, which contradicts the sibling size effect that in first-son families there are more per child resources available.

Another challenge to our interpretation centers around interpreting housing as premarital investments. Household investments in housing may reflect the desire to get higher returns. It is possible that some unobserved county-specific shocks lead more boys to be born, and is also linked to higher returns to real estate investment. But any county-specific factors would impact housing investment of local families within the area in similar ways, and the effect would not depend on child gender. Therefore, comparing investments between first-son and first-daughter families differences out the effect of such shocks. It is also possible that wealthier people—that is, parents with a first son in high-sex-ratio regions, possibly because of higher probability of migration or fewer children—just happen to keep their savings in the form of housing. But we have controlled for parental education levels in our main regressions, and further added household financial wealth and household income in robustness regressions, to account for any wealth effect.

Therefore, it is primarily due to marriage market considerations that parents with sons attach greater importance to housing investment when the sex ratio gets higher. We have provided supportive evidence in Section 2 that parents see their property as investment in their sons' physical capital in preparation for marriage, at the time when their sons are young. A closer look at the intended purpose of migration remittances in our sample also indicates the marriage market effects on parental decisions even if children are still young: When faced with a higher sex ratio, son families relative to daughter families are more willing to spend the migration remittances on the son's marriage such as building or buying a house, based on answers to a survey question in the CFPS (see Appendix Table A6).

Another piece of evidence in favor of our empirical findings mainly driven by marriage market considerations comes from heterogenous effects of sex imbalance across families. We split the main sample into two groups based on first-born children's proximity to marriage age. For the subsample that contains households with a first-born child above the age of 11, the effect of sex imbalance on housing investment for son families relative to daughter families appears much more prominent—the magnitude more than double the benchmark (see Appendix Table A7). This finding is in line with the interpretation that housing investment patterns mainly reflect parental considerations for children's marriage market prospects.

Status seeking may also be a potential explanation. Perhaps, the level of status competition among son families is higher than among daughter families in counties with high sex ratios. A strong desire to conform to norms in the same social strata induces families with a son to engage in earning activities and housing investment more aggressively. This interpretation is in line with Brown et al.'s (2011) finding that grooms' families spend more on weddings as local competition for status intensifies. However, it is incompatible with our finding that parents facing higher sex ratios invest relatively less in sons' education.

To sum up, while some other stories seem to rationalize part of our findings on parental labor supply and premarital investments, a natural and highly plausible interpretation is that bequeathable physical capital of men is more attractive to potential brides in the marriage market than human capital, as spouses are unable to commit to an agreement regarding the future division of the latter. These interpretations of our empirical results motivate us to build a theoretical model with imperfect commitment within marriage that incorporates two different types of premarital investments, which we turn to next.

6 A model of premarital investments with imperfect commitment

In this section we build a model of premarital investments with imperfect commitment within marriage. We highlight the difference between bequeathable physical capital (housing) and human capital. In particular, the model predictions in a setting with oversupplied men and large ex post bargaining power of men, match the empirical patterns we have shown.

6.1 The basic model

Setup We assume a continuum of men and a continuum of women. Let us consider first a situation with equal measures on both sides of the market. At the ex ante stage, the parents of a boy have to choose a vector of investments for their son, (x_B, y_B). x_B is investment in a private good, such as the son's human capital. y_B is investment in good which is public within marriage, such as the purchase of a house. The financial costs of investment are given by a function $c : \mathbb{R}^2_+ \to \mathbb{R}_+$, that is strictly increasing and strictly convex, with $c_x(.)$ and $c_y(.)$ denoting the partial derivative of costs with respect to investment in the two goods. We normalize the return on investments on each good to one, so that one unit of investment in the good yields, one average, one unit of the good – this is without loss of generality since the cost functions capture any non-linearity. Similarly, the parents of a girl choose a vector of investments (x_G, y_G). For simplicity, we assume that the cost functions are the same for the two sexes, an assumption that is easy to relax.

Finally, we assume that returns to the private good are stochastic: each boy is subject to a zero-mean shock ε, which has distribution function F. Consequently, the realized return on investment in a boy equals $x_B + \varepsilon$. Similarly, each girl is subject to a zero-mean shock η, which is distributed according to G, so that the realized return from equals $x_G + \eta$. As in Bhaskar and Hopkins (2016), the shocks ensure existence and uniqueness of a (quasi-symmetric) equilibrium in investment levels, under appropriate distributional assumptions.

We assume that two parties who agree to marry cannot commit to a sharing rule of the returns to the parental investments in any private good, including the stochastic component. Instead, the shares in the private good are determined by ex post bargaining. We model this by assuming that a man has a share λ_B in the returns, with his partner securing the remaining share, $1 - \lambda_B$. Similarly, a woman has share λ_G of the returns to in her investment in the private good good. We assume that public goods consumption in the couple is given by the sum $y := y_B + y_G$, with a man's payoff being $v_B(y)$ and a woman's payoff being $v_G(y)$, where these evaluation functions are strictly increasing and strictly concave.

Utilitarian efficient investments Consider first a social planner, who chooses the levels of investments, but who cannot dictate the sharing rule. Consider first the case where the social planner maximizes the ex-ante expected utility of the parent, before she observes the sex of her child. Thus, since the child is equally likely to be a boy or a girl, the planner will give their respective utilities equal weight. In this case, it is straight-forward to see that the investment profile (x_B^{**}, y_B^{**}) must satisfy the first order conditions:

$$c_x(x_B^{**}, y_B^{**}) = 1, \tag{2}$$

$$c_y(x_B^{**}, y_B^{**}) = v_B'(y_B^{**} + y_G^{**}) + v_G'(y_B^{**} + y_G^{**}). \tag{3}$$

Investments in the marriage market Let us now consider the marriage market. Matching takes place after investments and payoff shocks are realized. Suppose that a boy with investment profile (x_G, y_G) and shock value ε matches with girl with profile of investments (x_G, y_G) and shock realization η. Then overall payoff of the boy from this match equals

$$\lambda_B(x_B + \varepsilon) + (1 - \lambda_G)(x_G + \eta) + v_B(y_B + y_G). \tag{4}$$

The overall payoff of the girl from this match equals

$$\lambda_G(x_G + \eta) + (1 - \lambda_B)(x_B + \varepsilon) + v_G(y_B + y_G). \tag{5}$$

Our focus is on a quasi-symmetric equilibrium where all men invest (x_B, y_B) and all women invest (x_G, y_G). In such an equilibrium, men with higher levels of ε are uniformly more attractive to any woman, since $\lambda_B < 1$. Similarly, women with higher values of η are uniformly more attractive to men. Thus, any stable matching must be assortative in the shocks, and since it must also be measure preserving, the matching function $\phi(\varepsilon)$ satisfies $F(\varepsilon) = G(\phi(\varepsilon))$.

Let us examine the incentives for investment. Let us first consider marginal incentives for investment in a private good, such as human capital. In this case, if a man invests $x_B + \Delta$, the marital return on this investment arises from the fact that he is now more attractive to every woman. In particular, since a woman gets the same fraction $(1 - \lambda_B)$ of this Δ increment, as she does from a larger shock, then for any ε, he is as attractive as a man with a higher shock value ε', which satisfies $\varepsilon' - \varepsilon = \Delta$. Since the shock value of his marriage partner will equal $\phi(\varepsilon')$, and he gets a fraction $(1 - \lambda_G)$ of this value, his marginal marriage market benefit from investment equals

$$(1 - \lambda_G) \int \phi'(\varepsilon) d\varepsilon,$$

where

$$\int \phi'(\varepsilon) d\varepsilon = \int \frac{f(\varepsilon)}{g(\phi(\varepsilon))} f(\varepsilon) d\varepsilon =: \theta_B.$$

Consequently, the first order condition for optimal investment in the private good by a boy, i.e. the best response $\hat{x}_B$ is given by

$$c_x(\hat{x}_B, y_B) = \lambda_B + (1 - \lambda_G)\theta_B. \tag{6}$$

Observe that the return to investment in the private good is independent of the common investment level chosen by girls. Similarly, for women, the first order condition for investment in the private good is

$$c_x(\hat{x}_G, y_G) = \lambda_G + (1 - \lambda_B)\theta_G, \tag{7}$$

where

$$\theta_G := \int \frac{g(\eta)}{f(\phi^{-1}(\eta))} g(\eta) d\eta.$$

Consider next the public good. The marginal benefit from increasing $\hat{y}_B$ to $\hat{y}_B + \Delta$ is that for any ε, you are as attractive as type $\hat{\varepsilon}$ that satisfies

$$v_G(\hat{y}_B + \Delta, y_G) + (1 - \lambda_B)\varepsilon = v_G(\hat{y}_B, y_G) + (1 - \lambda_B)\hat{\varepsilon},$$

Thus, you will be matched with type $\phi(\hat{\varepsilon})$ rather than $\phi(\varepsilon)$, and the benefit of this is $(1 - \lambda_G)[\phi(\hat{\varepsilon}) - \phi(\varepsilon)]$. Thus the marginal return on the marriage market at any realization of ε is given by

$$\frac{1 - \lambda_G}{1 - \lambda_B} \phi'(\varepsilon) v'_G(\hat{y}_B + y_G).$$

Averaging over all realizations of ε, we see that the marriage market return from investment in the public good equals

$$v'_G(\hat{y}_B + y_B) \frac{(1 - \lambda_G)}{(1 - \lambda_B)} \int \frac{f(\varepsilon)}{g(\phi(\varepsilon))} f(\varepsilon) d\varepsilon.$$

In addition, an individual also benefits from his own consumption of the public good, at rate $v'_B(\hat{y}_B + y_G)$. Thus the first order condition for optimal investment by men in the public good is given by

$$c_y(\hat{x}_B, y_B) = v'_B(\hat{y}_B + y_G) + \frac{1 - \lambda_G}{1 - \lambda_B} \theta_B v'_G(\hat{y}_B + y_G). \tag{8}$$

In the case of the public good, we see that the best response for men, $\hat{y}_B$ does directly depend upon y^G, the investment level chosen by girls, since the payoff from the public good is strictly concave. In consequence, since the marginal costs of investment in the private good depend upon public good investment (unless the cost function is separable), $\hat{x}_B$ and $\hat{y}_B$ are both functions of y_G. However, they do not depend upon x_G, the level of girls' investments in the private good.

Since the argument is identical for women, the first order condition for optimal investment in the public good by women is

$$c_y(\hat{x}_G, \hat{y}_G) = v'_G(y_B + \hat{y}_G) + \frac{1-\lambda_B}{1-\lambda_G}\theta_G v'_B(y_B + \hat{y}_G). \tag{9}$$

Equations (6), (7), (8) and (9) characterize the best response investments in the private good and the public good by both the sexes.

Equilibrium We will now show that quasi-symmetric equilibrium exists and is unique. Furthermore, such an equilibrium is necessarily and stable, under the usual best response dynamics, which in turn implies that the comparative statics predictions will be intuitive. Consider the best response on the boys' side to a pair (x_G, y_G). Since the first order conditions for the boys' investments are unaffected by x_G, we may write this as a pair of functions $\hat{x}_B(y_G)$ and $\hat{y}_B(y_G)$. Thus, $\hat{y}_B(y_G)$ is the best response investments in the public good, when both types of investments are chosen optimally by the boy. Similarly, we may define best responses on the girls' side, $\hat{y}_G(y_B)$. Let ζ denote the composition of the functions $\hat{y}_G$ and $\hat{y}_B$ so that $\zeta(u) = \hat{y}_G(\hat{y}_B(u))$. The fixed points of ζ correspond to quasi-symmetric equilibria. More precisely, y_B is a fixed point of ζ if and only if $y_B = \hat{y}_G(y_B))$, so that the pair $y_B, \hat{y}_G(y_B)$ are mutual best responses, with the private good investments being given by $\hat{x}_B(\hat{y}(y_B))$ and $x_G = \hat{x}_G(y_B)$.

Observe that ζ is continuous and differentiable on the positive reals. Since $c_y(x, 0) = 0$, $\zeta(0) > 0$. Also, since $v'_B(y) \to 0$ and $v'_G(y) \to 0$ as $y \to \infty$, while $c_y(x, y) \to \infty$, $\zeta(y) < y$ for y sufficiently large. Thus there exists a fixed point of ζ, which we denote by y^*_B. Let us denote this quasi-symmetric equilibrium by $((x^*_B, y^*_B), (x^*_G, y^*_G))$.

To show uniqueness and stability, consider the slope of the best responses that compose ζ. The derivative of the boys' best response is

$$\frac{d\hat{y}_B}{dy_G} = \frac{\Omega^B c^B_{xx}(.)}{\Delta^B - \Omega^B c^B_{xx}(.)},$$

where $\Omega^B := v''_B(.) + \frac{1-\lambda_G}{1-\lambda_B}\theta_B vG''(.) < 0$, and Δ^B is the determinant of the Hessian of the cost function, $c(.)$, evaluated at (x_B^*, y_B^*). Since the cost function is strictly convex, $\Delta^B > 0$ and $c_{xx}^B > 0$, so that $\frac{d\hat{y}_B}{dy_G} \in (-1, 0)$.

Similarly,

$$\frac{d\hat{y}_G}{dy_B} = \frac{\Omega^G c_{xx}^G(.)}{\Delta^G - \Omega^G c_{xx}^G(.)},$$

where $\Omega^G := v_G''(.) + \frac{1-\lambda_B}{1-\lambda_G}\theta_G vB''(.) < 0$, and Δ^G is the determinant of the Hessian of the cost function, evaluated at (x_G^*, y_G^*). Thus $\frac{d\hat{y}_G}{dy_B} \in (-1, 0)$. Consequently, at any fixed point, ζ has a slope that is positive and strictly less than 1. Thus, there can be at most one fixed point. We summarize our results in the following proposition.

Proposition 1 *Under assumption 1, there exists a unique quasi-symmetric equilibrium, which is stable.*

6.2 Implications

We now proceed the qualitative properties of equilibrium investments and their relation to the utilitarian efficient investments.

The following lemma will be very useful.

lemma 2
- *$\theta_B \theta_G \geq 1$ for any F and G.*
- *$\theta_B \theta_B = 1$ if and only if distributions F and G are of the same type, i.e. $F(x) = G(ax+b)$.*
- *If $F = G$, then $\theta_B = \theta_G = 1$.*
- *If $G \geq_d F$, then $\theta_B > 1 > \theta_G$.*

The first order conditions have the following implications:

Proposition 3 *Suppose that $F = G$ and $\lambda_B = \lambda_G$, so that both sexes have the same distribution of shocks and the two bargaining powers, over labor income, are equal. Then investments in public goods and private goods are utilitarian efficient.*

This result may appear unexpected. Even if bargaining powers over the returns from the investments are asymmetric and unequal, with men having greater bargaining power over both their own labor income and their partner's labor income, this does not result in investment inefficiency. Indeed, boys' investments in both public and private goods are efficient as long as $\frac{1-\lambda_G}{1-\lambda_B}\theta_B = 1$; girls' investments are efficient as long as $\frac{1-\lambda_B}{1-\lambda_G}\theta_G = 1$. Of course, if the shock distributions are the same, $\theta_B = \theta_G = 1$, and if bargaining power over own labor incomes are equal, then $\frac{1-\lambda_G}{1-\lambda_B} = 1$.

Proposition 4 *Suppose that $F \geq_d G$, so that the shocks are more dispersed for men than for women. If the bargaining power of men over shocks is not greater than that of women, then women overinvest in the private good, and also overinvest in the public good, while men underinvest in both types of goods, relative to the utilitarian investment levels. Women overinvest more in private goods over which their bargaining power is lower, while men's underinvestment is less pronounced in private goods where their bargaining power is higher.*

Bhaskar and Hopkins (2016) examined investments with a single private good, and established that the sex whose shocks are less dispersed will over-invest, while that with dispersed shocks will underinvest. By allowing for multiple avenues for investment, the above proposition generalizes this result. Intuitively, when shocks are more dispersed for boys, there is more competition for boys, and less competition for girls. Consequently, girls investments are driven more by considerations of marriage market competition, while investments of boys are not, and are directed more towards their own returns. Girls will invest more in private goods where they have lower bargaining power, since this increases their attractiveness on the marriage market.

The next proposition shows that differences in bargaining power over shocks play a similar role to dispersion:

Proposition 5 *If men have more bargaining power over own income than women so that $\lambda_B > \lambda_G$, and if $F = G$, then men overinvest in the private good, and also overinvest in the public good, while women underinvest in both types of goods, relative to the utilitarian investment levels. Men overinvest more in private goods over which their bargaining power is lower, while women's underinvestment is less pronounced in private goods where their bargaining power is higher.*

The above proposition shows that having greater bargaining power over shocks has the same implication as having less dispersed shocks. Intuitively, if men have greater bargaining power over shocks, then since women get a smaller fraction of the man's shocks, this is effectively less dispersed. It is plausible that men's shocks are more dispersed than women, and also that men have greater bargaining power over the shock component. In this case, the above two propositions show that the two forces can offset each other, and make investments more efficient.

6.3 Comparative statics

We are now in a position to examine the effects of a change in bargaining power of one of the sexes upon equilibrium investments on both sides. Since we have shown that the equilibrium is stable, comparative statics will be intuitive. Nonetheless, some complications arise due to the fact that the cost function $c(x, y)$ is not separable in its two arguments. Consequently, if there is a greater incentive in both goods, it is possible that investment in one good might conceivably fall if $c_{xy}(.)$ is very large.

Suppose that the bargaining power of boys over their own income is larger, i.e. consider an increase in λ_B. Examining the first order condition for boys for private goods, (6), we see that boy have a greater incentive to invest in the private good, since the marginal return increases. This is intuitive – since boys retain a greater share of their earnings, there incentive to invest in the private good is greater. More interesting is the effect on public good investments. Examining the first order condition (8), we see that boys have a greater incentive to invest in the public good, since this is now a more effective way to compete on the marriage market. In other words, investing in the public good is a better commitment device, since a man can now no longer promise as large a share of labor income to his potential spouse as before. Indeed, since a larger value of ε is less attractive now to a woman – since she gets a smaller share – an increment to public good investment becomes more competitive on the marriage market.

Consider now the effect on girls' investments. Since it now less important to attract a better quality of boy, girls have reduced incentives to invest in both the private good and in the public good. The following proposition summarizes our results.

Proposition 6 *An increase in the bargaining power of boys, λ_B, increases boys' investments in both private good and public good as long as $c_{xy}(x_B^*, y_B^*)$ is not too large, and reduces girls' investments in both types of good as long as as long as $c_{xy}(x_G^*, y_G^*)$ is not too large.*

Proof. See Appendix A.

■

6.4 Modelling sex imbalance

The distortionary effect on investments arise when bargaining powers are asymmetric, and favor one sex. This is most likely to be men, given their superior legal and customary position in many societies, and in this case, there will be overinvestment in boys and underinvestment in girls, with boys investing in those goods where they can commit to share the rewards more equally. On the other hand, distortions can also arise due to difference in ex ante competitive position, since the sex that faces more competition has a greater incentive to invest, in order to improve its competitive position. We now see that how the two distortions can reinforce each other, in traditional societies where men have greater bargaining power within marriage, and are also in greater number on the marriage market. This is particularly relevant in countries such as India and China.

Our modelling strategy incorporates the following innovation, which allows the sex ratio to affect investment incentives in a continuous fashion.[19] Suppose that the ratio of women to men is $r < 1$. We assume that the overall marriage market is composed of many local marriage markets, where the sex ratio varies. In some of these marriage markets, there is an excess of men, while in the others, the marriage market is balanced. A reduction in r, the aggregate ratio of women to men, increases the likelihood that an individual woman resides in a market where there is an excess of men. A simple way of modeling this is as follows. Fix $\hat{r} < 1$, and let this be sufficiently small so that the aggregate sex ratio, r, lies in the interval $(\hat{r}, 1]$. The sex ratio in a local market takes one of two values, $\hat{r}$ and 1, where the probability of the first value is $\rho(r)$. Since the aggregate sex ratio equals r, $\rho(r)$ must satisfy the equation:

$$\rho(r)\hat{r} + (1 - \rho(r)) = r,$$

which implies that

$$\rho(r) = \frac{1 - r}{1 - \hat{r}}.$$

19. In previous work such as Bhaskar and Hopkins (2016), the sex ratio has discontinuous effects on investment incentives at $r = 1$.

It is straightforward to verify that $\rho(1) = 0$, which is consistent with our analysis of the case of a balanced sex ratio. Further, $\rho'(r) = -\frac{1}{1-\hat{r}} < 0$.

Utilitarian efficient investments Before analyzing equilibrium investments, let us consider the conditions for utilitarian efficiency. For private goods, the return on investment equals 1, and this return is either shared if the individual marries, or accrues entirely to the individual if he remains single. Since the utilitarian planner puts equal weight on both partners, the first order condition for utilitarian efficiency remains $c_x(.) = 1$, for both men and women.

However, utilitarian investments in the public good do depend upon the sex ratio, since the likelihood of marriage determines whether the public good is shared, or consumed singly. Observe that a man is married with probability r. In the event that he is married, the benefit of the public good accrues also to his partner, while if he is not married, it does not. Consequently, utilitarian efficiency requires:

$$c_y(x_B^{**}, y_B^{**}) = r[v_B'(y_B^{**} + y_G^{**}) + v_G'(y_B^{**} + y_G^{**})] + (1-r)v_B'(y_B^{**}). \tag{10}$$

For a woman, her probability of marriage equals one, and hence the efficiency condition is:

$$c_y(x_G^{**}, y_G^{**}) = v_B'(y_B^{*,*} + y_G^{**}) + v_G'(y_B^{**} + y_G^{**}). \tag{11}$$

The effects of the sex ratio r upon efficient investments is, in general, ambiguous. If men have a greater matching probability, due to an increase in r, then the planner would like them to invest more, since the investments benefit their partner. However, they are also more likely to benefit directly from their partner's investment in the public good, and are less likely to remain single, and this is a force towards reducing men's investments in the public good.

Equilibrium investments Let us now turn to equilibrium investments. The matching function in the local marriage market now takes two different forms, depending upon whether there is an excess of men or not. In a local market where the sex ratio is balanced, the matching function is ϕ, as we have already analyzed. So consider a local market with an excess of men, so that the local sex ratio is $\hat{r}$. Let $\hat{\varepsilon}$ denote the lowest quality boy that is matched, an let the matching function in this case be denoted by ϕ_+. Since the matching it must be measure preserving, it must now satisfy

$$1 - F(\varepsilon) = \hat{r}[1 - G(\phi_{+}(\varepsilon)]. \tag{12}$$

The derivatives of $\phi'_{+}(.)$ is given by

$$\phi'_{+}(\varepsilon) = \frac{f(\varepsilon)}{\hat{r}g(\phi_{+}(\varepsilon))}.$$

When boys are in excess supply, so that $\hat{r} < 1$, this magnifies the impact of an increase in the boy's quality shock upon his match quality. Intuitively, since there is smaller measure of girls than boys, the qualities of the girls are more dispersed relative to the boys. Thus the marriage market return to own quality is greater for boys.

Let $\xi_{+}(\eta) = \phi_{+}^{-1})'(\eta)$, i.e. ξ_{+} is the inverse of the matching function in a market with an excess of boys, and specifies which quality of boy is matched to type η of girl By the same logic,

$$\xi'_{+}(\eta) = \frac{\hat{r}g(\eta)}{f(\xi_{+}(\eta))}.$$

Let us now define θ_{B+}, as follows:

$$\theta_{B+} := \int_{\hat{\varepsilon}} \phi'_{+}(\varepsilon) f(\varepsilon) d\varepsilon = \frac{1}{\hat{r}} \int_{\hat{\varepsilon}} \frac{f(\epsilon)}{g(\phi_{+}(\epsilon))} f(\epsilon) d\epsilon.$$

Similarly, we define θ_{G+}, as follows:

$$\theta_{G+} := \int \xi'_{+}(\eta) d\eta = \hat{r} \int \frac{[g(\eta)]^2}{f(\xi_{+}(\eta))} d\eta$$

Note that the expressions θ_B and θ_G, that apply to a balanced local marriage market, are as defined previously.

Let $\hat{f} := f(\hat{\varepsilon})$, and let $\bar{U}$ denote the utility gain of the boy from being matched to the lowest quality girl, as compared to being unmatched. The first order condition for a boy's optimal investment in the private goods given by

$$\begin{aligned} c_x(x_B^*, y_B^*) = \rho(r) &\left[(1 - \hat{r}) + \hat{r}[\lambda_B + (1 - \lambda_G)\theta_{B+}] + \hat{f}\bar{U}\right] \\ &+ (1 - \rho(r)) \left[\lambda_B + (1 - \lambda_G)\theta_B\right]. \end{aligned} \tag{13}$$

The first line on the right-hand side considers the payoff in a local marriage market where there is an excess of boys. The investment return in such a market consists of three terms. With probability $1 - \hat{r}$ the boy is single, and enjoys the entire return on his investment. With probability $\hat{r}$, he is married, and must share the return with his spouse. However, in this case, an increment to investment also increases his rank in the marriage market, and therefore, there is a marriage market return on his investment. Observe that this marriage market investment return is magnified, since it divided by $\hat{r}$. Intuitively, since the (relative) measure of girls in the local market is only $\hat{r}$, the effective dispersion amongst girls is larger than amongst boys, increasing the marriage market returns to investment for boys. Finally, the third term, reflects the fact that by increasing investment, the boy increases his chances of being married, by overtaking the lowest ranked boy, of quality $\hat{\varepsilon}$. In other words, the desire not to left unmatched magnifies investment incentives.

The second line reflects the payoff in a balanced local marriage market. In this case, the boy gets a fraction λ_B of his own return, plus the marriage market return, which is lower since the effective dispersion on girls' qualities is lower in a local market where there is an excess of girls.

The first order condition can be simplified as follows:

$$c_x(x_B^*, y_B^*) = \lambda_B + (1 - \lambda_G)\theta_B + \rho(r)\left[(1 - \hat{r})(1 - \lambda_B) + (1 - \lambda_G)(\hat{r}\theta_{B+} - \theta_B) + \hat{f}\bar{U}\right]. \quad (14)$$

The first order condition for the public good for men is given by

$$c_y(x_B^*, y_B^*) = \begin{array}{r} \rho(r)\left[(1 - \hat{r})v_B'(y_B^*) + \hat{r}v_B'(y_B^* + y_G^*) + \hat{r}\frac{1-\lambda_G}{1-\lambda_B}v_G'(y_B^* + y_G^*)\theta_{B+} + \frac{v_G'(y_B^*+y_G^*)}{1-\lambda_B}\hat{f}\bar{U}\right] \\ +(1 - \rho(r))\left[v_B'(y_B^* + y_G^*) + \frac{1-\lambda_G}{1-\lambda_B}v_G'(y_B^* + y_G^*)\theta_B\right]. \end{array} \quad (15)$$

This can be simplified as

$$c_y(x_B^*, y_B^*) = \begin{array}{c} v_B'(y_B^* + y_G^*) + \frac{1-\lambda_G}{1-\lambda_B}v_G'(y_B^* + y_G^*)\theta_B \\ +\rho(r)[(1 - \hat{r})[v_B'(y_B^*) - v_B'(y_B^* + y_G^*)] \\ +\frac{1-\lambda_G}{1-\lambda_B}v_G'(y_B^* + y_G^*)\theta_{B+}(\hat{r}\theta_{B+} - \theta_B) + \frac{v_G'(y_B^*+y_G^*)}{1-\lambda_B}\hat{f}\bar{U}]. \end{array} \quad (16)$$

For women, the first order condition for investment in the private good is simpler:

$$c_x(x_G^*, y_G^*) = \lambda_G + (1-\lambda_B)\left[\rho(r)\theta_{G+} + (1-\rho(r))\theta_G\right]. \tag{17}$$

Since a woman is always matched, she gets a fraction λ_G of her own return, and with probability $\rho(r)$ she gets the marital return in the market where women are short supply, and with the remaining probability, the marital return in a balanced marriage market.

The first order condition for women's investment in the public good is

$$c_y(x_G^*, y_G^*) = v_G'(y_B^* + y_G^*) + \frac{1-\lambda_B}{1-\lambda_G} v_B'(y_B^* + y_G^*)[\rho(r)\theta_{G+} + (1-\rho(r))\theta_G)]. \tag{18}$$

Women invest less in the public good, for two reasons. First, they underweight the effect on their utility of their partner by a fraction r, even though efficiency dictates that they give this weight one. Second, there is a crowding out effect: since men invest more in public goods, women have less incentives to invest.

It is more illuminating the examine the coefficient on $\rho(r)$ in the first-order conditions for optimal investments, which represents the difference from the case with a balanced sex ratio analyzed in Section 5.1 (where $\rho(r) = 0$):

$$\textit{FOC(Boys, Private)} : (1-\hat{r})(1-\lambda_B) + (1-\lambda_G)(\hat{r}\theta_{B+} - \theta_B) + \hat{f}\bar{U} \tag{19}$$

Suppose that F and G are both uniform, and the ratio of the two densities is k. Then, it is straightforward to verify that $\theta_{B+} = k, \theta_B = k$, and $\theta_{G+} = r/k, \theta_G = 1/k$. In the uniform case, this reduces to:

$$\textit{FOC(Boys, Private, Unif)} : (1-\hat{r})[(1-\lambda_B) - k(1-\lambda_G)] + \hat{f}\bar{U} \tag{20}$$

The first order condition for the public good for men is given by

$$\textit{FOC(Boys, Public)} : (1-\hat{r})[v_B'(y_B^*) - v_B'(y_B^* + y_G^*)] + \frac{v_G'(y_B^* + y_G^*)}{1-\lambda_B}\left[(1-\lambda_G)(\hat{r}\theta_{B+} - \theta_B) + \hat{f}\bar{U}\right] \tag{21}$$

and

$$FOC(Boys,\ Public, Unif): (1-\hat{r})[v_B'(y_B^*) - v_B'(y_B^*+y_G^*)] + \frac{v_G'(y_B^*+y_G^*)}{1-\lambda_B}[\hat{f}\bar{U} - (1-\lambda_G)(1-\hat{r})k] \tag{22}$$

For women,

$$FOC(Girls,\ Private): (1-\lambda_B)(\theta_{G+} - \theta_G). \tag{23}$$

$$FOC(Girls,\ Private,\ Unif): \frac{(1-\lambda_B)(\hat{r}-1)}{k}. \tag{24}$$

$$FOC(Girls,\ Public): \frac{v_B'(y_B^*+y_G^*)}{1-\lambda_G}[\theta_{G+} - \theta_G)]. \tag{25}$$

$$FOC(Girls,\ Public,\ Unif): -\frac{v_B'(y_B^*+y_G^*)}{1-\lambda_G}\frac{1-\hat{r}}{k}. \tag{26}$$

We summarize these results in the following proposition.

Proposition 7 *Suppose that $r < 1$, so that there is an oversupply of men relative to women. If the utility gain of men from being matched to the lowest quality women, as compared to being unmatched, is large enough, then men overinvest in the public good, and also overinvest in the private good, while women underinvest in both types of goods, compared to the case where $r = 1$ (sex ratio is balanced). In other words, men overinvest in both types of goods, relative to women.*

6.4.1 When men have large ex post bargaining power

Suppose that the sex ratio is imbalanced, so that women have an advantage on the marriage market. However, men have a high bargaining power over the returns from investments in private goods, i.e. λ_{Bi} is large. Consider the limit as $\lambda_{Bi} \to 1$. In this case, the incentive of men to overinvest in the private good disappears, and the limit investments in (14) satisfy

$$\lim_{\lambda_{Bi}\to 1} c_i'(x_B^*, y_B^*) = 1.$$

On the other hand, the incentive to overinvest in public goods remains. Since the investments in the public good do not depend upon men's bargaining power over the the public good, the expression for (21) is unaffected. Indeed, one might argue that public good investment incentives may be magnified, if boys' bargaining power over shocks also increases. If we $\lambda_{B0} \to 1$, then the right hand side increases to ∞.

In summary, we have the following proposition.

Proposition 8 *Suppose that $r < 1$, so that there is an oversupply of men relative to women, and that the bargaining power of men, λ_B, is large. If the utility gain of men from being matched to the lowest quality women, as compared to being unmatched, is large enough, then men overinvest in the public good, relative to women. For private good over which men's bargaining power is higher, men underinvest relative to women, as long as λ_B or $c_{xy}(x_B^*, y_B^*)$ is large enough.*

7 Conclusion

This paper empirically and theoretically studies how imperfect commitment in marriage affects premarital investments in children made by their parents. Using data from a nationally representative Chinese household survey, we find that high sex ratios lead to increased parental migration (a proxy for stronger earnings incentive), increased housing investments, and reduced educational investments for families with a first-born son relative to families with a first-born daughter. To get reliable estimates, we control for unobserved county-level heterogeneity and compare the effect of sex imbalance for first-son families with the effect for first-daughter families within the county. We also provide a variety of robustness checks to address the concerns about potential endogeneity of sex ratios—including using high-dimensional method and IV method—and other concerns.

We propose that parents increase labor supply in order to increase premarital investments in their sons to improve their future marriage prospects. We further propose that the underlying reason for the pattern of premarital investments is bequeathable physical capital such as housing—which will be shared equally between spouses in its consumption after marriage and therefore, indicates a credible commitment—being more attractive in the marriage market than human capital—the sharing of which will be subject to bargaining after marriage. We provide supportive evidence for our interpretations and also discuss some competing hypotheses.

Motivated by these interpretations of our empirical findings, we develop a model where imperfect commitment combines with sex imbalance to affect the magnitude and composition of premarital investments in children. The model differentiates premarital investments in bequeathable physical capital (housing) and human capital, and assumes constant after-marriage bargaining powers that do not depend on marriage market conditions. With sex imbalance, the model predicts that men—the oversupplied side—would increase investment to compete for marital partners. When men have great control over their own labor earnings after marriage, the model predicts that they would increase physical capital investment at the expense of lower human capital investment. The predictions are consistent with our empirical findings.

This paper highlights the distinction between premarital investments in physical capital and human capital. It has important implications for human capital development of the next generation. Underinvestment in education as well as increased parental migration driven by sex imbalance may undermine the development of boys relative to girls. (Appendix B provides a detailed discussion.) The paper also has important implications for the study of marriage. While classic work on marriage markets focuses on one-dimensional assortative matching on income, wage, or education (Becker, 1973, 1974), recent studies begin to realize the importance of marriage matching along multiple dimensions (Chiappori et al., 2012, 2017; Galichon and Salanié, 2010). Our results show the different roles in marriage matches of multiple types of capital. This area deserves more attention in future research.

References

Ahrens, A., C. B. Hansen, and M. E. Schaffer (2018). pdslasso and ivlasso: Programs for post-selection and post-regularization ols or iv estimation and inference. *URL http://ideas. repec. org/c/boc/bocode/s458459. html*.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy 113*(1), 151–184.

Anderson, S. and C. Bidner (2015). Property rights over marital transfers*. *Quarterly Journal of Economics 130*(3), 1421–1484.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Becker, G. S. (1973). A theory of marriage: Part i. *Journal of Political Economy*, 813–846.

Becker, G. S. (1974). A theory of marriage. In *Economics of the Family: Marriage, Children, and Human Capital*, pp. 299–351. University of Chicago Press.

Becker, G. S. and H. G. Lewis (1973). On the interaction between the quantity and quality of children. *Journal of Political Economy 81*(2), S279–S288.

Beijing (2013). Married to the mortgage. *Economist*.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives 28*(2), 29–50.

Bhaskar, V. (2011). Sex selection and gender balance. *American Economic Journal: Microeconomics 3*(1), 214–244.

Bhaskar, V. and E. Hopkins (2016). Marriage as a rat race: Noisy premarital investments with assortative matching. *Journal of Political Economy 124*(4), 992–1045.

Brown, P. H., E. Bulte, and X. Zhang (2011). Positional spending and status seeking in rural china. *Journal of Development Economics 96*(1), 139–149.

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

Cameron, L., X. Meng, and D. Zhang (2017). China's sex ratio and crime: Behavioural change or financial necessity? *Economic Journal*.

Chiappori, P.-A., M. C. Dias, and C. Meghir (2018). The marriage market, labor supply, and education choice. *Journal of Political Economy 126*(S1), S26–S72.

Chiappori, P.-A., B. Fortin, and G. Lacroix (2002). Marriage market, divorce legislation, and household labor supply. *Journal of Political Economy 110*(1), 37–72.

Chiappori, P.-A., M. Iyigun, and Y. Weiss (2009). Investment in schooling and the marriage market. *American Economic Review 99*(5), 1689–1713.

Chiappori, P.-A. and M. Mazzocco (2017). Static and intertemporal household decisions. *Journal of Economic Literature 55*(3), 985–1045.

Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2012). Fatter attraction: anthropometric and socioeconomic matching on the marriage market. *Journal of Political Economy 120*(4), 659–695.

Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2017). Bidimensional matching with heterogeneous preferences: education and smoking in the marriage market. *Journal of the European Economic Association*, jvx012.

Choo, E. and A. Siow (2006). Who marries whom and why. *Journal of Political Economy 114*(1), 175–201.

Chu, C. C. and R.-R. Yu (2010). *Understanding Chinese families: A comparative study of Taiwan and Southeast China*. Oxford University Press.

Cole, H. L., G. J. Mailath, and A. Postlewaite (2001, December). Efficient non-contractible investments in large economies. *Journal of Economic Theory 101*(2), 333–373.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Das Gupta, M., A. Y. Ebenstein, and E. J. Sharygin (2013). China's marriage market and upcoming challenges for elderly men. *World Bank Policy Research Working Paper Series, Vol.*

Davin, D. (2005). Marriage migration in china: The enlargement of marriage markets in the era of market reforms. *Indian Journal of Gender Studies 12*(2-3), 173–188.

Ebenstein, A. (2010). The "Missing Girls" of China and the Unintended Consequences of the One Child Policy. *Journal of Human Resources 45*(1), 87–115.

Ebenstein, A. (2011). Estimating a Dynamic Model of Sex Selection in China. *Demography 48*(2), 783–811.

Edlund, L., H. Li, J. Yi, and J. Zhang (2013). Sex ratios and crime: Evidence from china. *Review of Economics and Statistics 95*(5), 1520–1534.

Galichon, A., S. D. Kominers, and S. Weber (2019). Costly concessions: An empirical framework for matching with imperfectly transferable utility. *Journal of Economic Perspectives (Forthcoming).*

Galichon, A. and B. Salanié (2010). Matching with trade-offs: Revealed preferences over competing characteristics.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–2086.

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences 104*(33), 13250–13255.

Hu, L. and A. Schlosser (2015). Prenatal sex selection and girls' well-being: Evidence from india. *Economic Journal 125*(587), 1227–1261.

Huang, S. (2010). Housing and marriage. *Chinadaily*.

Iyigun, M. and P. R. Walsh (2007). Building the family nest: Premarital investments, marriage markets, and spousal allocations. *Review of Economic Studies 74*(2), 507–535.

Lafortune, J. (2013). Making yourself attractive: Premarital investments and the returns to education in the marriage market. *American Economic Journal: Applied Economics 5*(2), 151–178.

Li, H., J. Yi, and J. Zhang (2011). Estimating the effect of the one-child policy on the sex ratio imbalance in china: Identification based on the difference-in-differences. *Demography 48*(4), 1535–1557.

Li, Q. and J. Pantano (2013). The demographic consequences of sex-selection technology.

Lise, J. and K. Yamada (2018). Household sharing and commitment: Evidence from panel data on individual expenditures and time use. *The Review of Economic Studies 86*(5), 2184–2219.

Low, C. (2017). A reproductive capital model of marriage market matching. *Manuscript, Wharton School of Business*.

Lundberg, S. and R. A. Pollak (1993). Separate spheres bargaining and the marriage market. *Journal of Political Economy*, 988–1010.

Lundberg, S. and R. A. Pollak (1996). Bargaining and distribution in marriage. *Journal of Economic Perspectives 10*(4), 139–158.

Lundberg, S. and E. Rose (2002). The effects of sons and daughters on men's labor supply and wages. *Review of Economics and Statistics 84*(2), 251–268.

Lyle, D. S. (2006). Using military deployments and job assignments to estimate the effect of parental absences and household relocations on children's academic achievement. *Journal of Labor Economics 24*(2), 319–350.

Mazzocco, M. (2007). Household intertemporal behaviour: A collective characterization and a test of commitment. *Review of Economic Studies 74*(3), 857–895.

McKenzie, D. and H. Rapoport (2011). Can migration reduce educational attainment? evidence from mexico. *Journal of Population Economics 24*(4), 1331–1358.

Mullainathan, S. and J. Spiess (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives 31*(2), 87–106.

Nunn, N. (2005). A model explaining simultaneous payments of dowry and brideprice. *Unpublished Paper. http://www. econ. ubc. ca/nnunn/dowries. pdf*.

Peters, M. and A. Siow (2002). Competing premarital investments. *Journal of Political Economy 110*(3), 592–608.

Pierson, D. (2010). China's housing boom spells trouble for boyfriends. *Los Angeles Times*.

Pollak, R. A. (1985). A transaction cost approach to families and households. *Journal of Economic Literature 23*(2), 581–608.

Sharygin, E., A. Ebenstein, and M. Das Gupta (2013). Implications of china's future bride shortage for the geographical distribution and social protection needs of never-married men. *Population Studies 67*(1), 39–59.

Shepard, W. (2016). Demand for housing in china's heated property market is real. *Forbes*.

Short, S. E. and F. Zhai (1998). Looking locally at china's one-child policy. *Studies in Family Planning*, 373–387.

Siow, A. (1998). Differential fecundity, markets, and gender roles. *Journal of Political Economy 106*(2), 334–354.

Wei, S.-J. and X. Zhang (2011). The competitive saving motive: Evidence from rising sex ratios and savings rates in china. *Journal of Political Economy 119*(3), 511–564.

Wrenn, D. H., J. Yi, and B. Zhang (2019). House prices and marriage entry in china. *Regional Science and Urban Economics 74*, 118–130.

Xie, Y. (2012). The user's guide of the china family panel studies (2010). *Beijing: Institute of Social Science Survey, Peking University*.

Xie, Y. and Y. Jin (2015). Household wealth in china. *Chinese Sociological Review 47*(3), 203–229.

Xinhua (2011). Housing pressure leads to delayed marriage. *Chinadaily*.

Zhang, H. (2015). Pre-matching gambles. *SSRN Working Paper*.

Zhang, H. (2019). An investment-and-marriage model with differential fecundity.

Zhang, H., J. R. Behrman, C. S. Fan, X. Wei, and J. Zhang (2014). Does parental absence reduce cognitive achievements? evidence from rural china. *Journal of Development Economics 111*, 181–195.

Zhang, J., Y. Zhao, A. Park, and X. Song (2005). Economic returns to schooling in urban china, 1988 to 2001. *Journal of Comparative Economics 33*(4), 730–752.

Zhao, Y. (1999). Labor Migration and Earnings Differences: The Case of Rural China. *Economic Development and Cultural Change 47*(4), 767–782.

Table 1 Summary statistics of main outcome variables

	Mean	Std. Dev.	Min	Max	Observations
A: Parental labor supply					
Paternal migration	0.098	0.297	0	1	4,314
Maternal migration	0.025	0.158	0	1	4,314
At least one parent migration	0.111	0.314	0	1	4,314
Paternal working hours, thousand	2.466	0.947	0.400	5.400	1,534
Maternal working hours, thousand	2.416	0.902	0.240	5.400	978
B: Housing investment					
Housing construction area, thousand sq.m	0.126	0.086	0.008	1	4,169
Housing ownership	0.831	0.375	0	1	4,314
Housing mortgage, thousand	5.392	32.04	0	750	4,314
C: Child educational investment					
Education expenditure, thousand	1.507	2.629	0	40	3,978
Having an education funding	0.297	0.457	0	1	3,978

Notes: Data are from the 2010 CFPS survey. The main sample includes all households in the 2010 CFPS family dataset in which the first-born child was between the ages of zero and 15, both parents were alive and at most 50 years old, and at least one parent participated in the 2010 CFPS adult survey. In panel C, child educational investment is measured for first-born children who are at least two years old. Descriptive statistics are weighted by the CFPS survey sampling weights.

Table 2 A balance test: First-son versus fist-daughter families

	Mean (Std. Dev.)				
	All	First-son families	First-daughter families	Difference	SE
	(1)	(2)	(3)	(4)	(5)
First son	0.507	–	–	–	–
	(0.500)				
Sex ratio (M/F)	1.077	1.076	1.077	-0.001	0.003
	(0.101)	(0.100)	(0.101)		
Ethnicity (minority=1)	0.124	0.121	0.128	-0.007	0.010
	(0.330)	(0.326)	(0.334)		
Region of residence (urban=1)	0.438	0.452	0.424	0.028	0.015
	(0.496)	(0.498)	(0.494)		
First-child age	8.746	8.623	8.874	-0.251	0.138
	(4.543)	(4.531)	(4.552)		
Father's age	36.14	36.03	36.27	-0.240	0.187
	(6.149)	(6.137)	(6.162)		
Father's schooling years	7.818	7.890	7.745	0.145	0.131
	(4.308)	(4.266)	(4.350)		
Father's political status (party=1)	0.091	0.090	0.092	-0.002	0.009
	(0.287)	(0.286)	(0.289)		
Mother's age	34.30	34.21	34.40	-0.190	0.190
	(6.251)	(6.264)	(6.239)		
Mother's schooling years	6.549	6.591	6.506	0.085	0.143
	(4.693)	(4.652)	(4.735)		
Mother's political status (party=1)	0.026	0.030	0.023	0.007	0.005
	(0.160)	(0.171)	(0.149)		
Observations	4,314	2,186	2,128		

Notes: Data are from the 2010 CFPS survey. In the first three columns, standard deviations are given in parentheses. In the last column are standard errors for the difference between characteristics of first-son and first-daughter families, none of which are statistically significant at the five percent level.

Table 3 Baseline results: Parental labor supply and premarital investments

A: Parental labor supply	Migration			Working hours, log	
Dependent variable	Father	Mother	At least one parent	Father	Mother
	(1)	(2)	(3)	(4)	(5)
First son * Sex ratio (β_3)	0.235**	0.098*	0.264***	0.569***	0.473
	(0.094)	(0.059)	(0.093)	(0.169)	(0.408)
Observations	4,314	4,314	4,314	1,534	978
R-squared	0.109	0.064	0.113	0.164	0.256
Dependent variable mean	0.098	0.025	0.111	7.726	7.701
Percentage difference sex ratio+0.1	24.1	38.6	23.8	5.7	4.7
Model	OLS	OLS	OLS	OLS	OLS
Other controls?	YES	YES	YES	YES	YES
County fixed effects?	YES	YES	YES	YES	YES

B: Premarital investments	Housing investment			Child educational investment	
Dependent variable	Construction area, log sq.m	Ownership	Mortgage, thousand	Education expenditure, thousand	Having an education funding
	(1)	(2)	(3)	(4)	(5)
First son * Sex ratio (β_3)	0.413**	0.233**	15.403**	-1.663**	-0.337**
	(0.205)	(0.117)	(7.141)	(0.800)	(0.161)
Observations	4,169	4,314	4,314	3,978	3,978
R-squared	0.278	0.177	0.145	0.323	0.135
Dependent variable mean	4.650	0.831	5.392	1.507	0.297
Percentage difference sex ratio+0.1	4.1	2.8	28.6	-11.0	-11.3
Model	OLS	OLS	OLS	OLS	OLS
Other controls?	YES	YES	YES	YES	YES
County fixed effects?	YES	YES	YES	YES	YES

Notes: Data are from the 2010 CFPS survey. In columns (1)–(3) of panel A and columns (2)–(5) of panel B, the difference in the effect of sex imbalance between first-son and first-daughter families is reported in both percentage points (β_3) and percentages (β_3/dependent variable mean); in the remaining columns, the difference is reported in percentages. In columns (4) and (5) of panel B, child educational investment is measured for the first-born child in the family who is at least two years old. Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 4 Addressing issues related to son-preferring fertility stopping rules

Dependent variable		Paternal migration	House construction area, log sq.m	Education expenditure, thousand
		(1)	(2)	(3)
		Interaction-term coefficient (β_3)		
Benchmark		0.235**	0.413**	-1.663**
A: Family-size effect				
Adding number of children		0.240**	0.409**	-1.689**
		[0.218]	[0.478]	[0.285]
Adding number of children		0.245**	0.410*	-1.689**
	& Interaction with first son	[0.215]	[0.745]	[0.467]
B: Families with one child				
One-child families	No age limit	0.234**	0.336	-1.776**
	Child $\geq$ 4	0.223**	0.217	-2.411**
C: Alternative measures of marriage market conditions				
Having any son	OLS	0.223***	0.310	-1.168*
	OLS, adding number of children	0.221***	0.313	-1.151
	OLS, adding number of children & interaction	0.220***	0.313	-1.154
	IV	0.355**	0.528**	-2.505**
	IV, adding number of children	0.360**	0.522**	-2.644**
	IV, adding number of children & interaction	0.356**	0.505**	-2.608**
Share of sons	OLS	0.300***	0.398*	-1.095
	OLS, adding number of children	0.302***	0.394*	-1.112
	OLS, adding number of children & interaction	0.301***	0.394*	-1.114
	IV	0.305**	0.495**	-2.173**
	IV, adding number of children	0.312**	0.493**	-2.231**
	IV, adding number of children & interaction	0.308**	0.474**	-2.243**

Notes: Data are from the 2010 CFPS survey. In column (3), education expenditure is measured for first-born children who are at least two years old. The difference in the effect of sex imbalance between first-son and first-daughter families is reported in percentage points (β_3). In panel A, p-values of Hausman's general specification test for the equality of β_3 are given in square brackets. In panel C, the instrument for having any son and the share of sons in IV regressions is the first-son dummy; see Appendix Table A4 for first-stage results. Estimations are weighted by the CFPS survey sampling weights. Standard errors are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 5 Addressing issues related to county-level sex ratios

Dependent variable		Paternal migration (1)	House construction area, log sq.m (2)	Education expenditure, thousand (3)
		Interaction-term coefficient (β_3)		
Benchmark		0.235**	0.413**	-1.663**
A: Unobservable cross-county heterogeneity				
No county fixed effects		0.233**	0.245	-1.857***
		[0.914]	[0.017]	[0.428]
B: Potential sex-ratio confounders				
Adding average household financial wealth		0.236**	0.397**	-1.665**
		[0.688]	[0.479]	[0.939]
Adding average household financial wealth		0.236**	0.396*	-1.675**
	& Interaction with first son	[0.738]	[0.413]	[0.885]
Adding average household income		0.237**	0.402*	-1.662**
		[0.592]	[0.363]	[0.911]
Adding average household income		0.239***	0.405**	-1.632**
	& Interaction with first son	[0.663]	[0.593]	[0.748]
Adding gender earning differential, m-f		0.251***	0.356*	-1.756**
		[0.142]	[0.029]	[0.441]
Adding gender earning differential, m-f		0.252***	0.356*	-1.766**
	& Interaction with first son	[0.176]	[0.025]	[0.453]
Adding social insurance		0.236**	0.432**	-1.694**
		[0.911]	[0.418]	[0.560]
Adding social insurance		0.242***	0.429**	-1.679**
	& Interaction with first son	[0.494]	[0.550]	[0.858]
Adding grandparental coresidence		0.232**	0.394*	-1.661**
		[0.567]	[0.526]	[0.824]
Adding grandparental coresidence		0.237**	0.393*	-1.664**
	& Interaction with first son	[0.857]	[0.532]	[0.966]
Adding all variables above		0.249***	0.347*	-1.794**
		[0.298]	[0.271]	[0.321]
Adding all variables above		0.260***	0.339*	-1.802**
	& Interactions with first son	[0.245]	[0.156]	[0.331]
Adding variables selected by high-dimensional method		0.251***	0.519**	-1.734**
	& Interactions with first son	[0.786]	[0.359]	[0.844]
C: IV results		0.374*	1.283*	-3.291*
		(0.224)	(0.776)	(1.993)
	First-stage F-statistic	3.429	5.294	3.282
	P-value	0.000	0.000	0.000
D: Sex ratios for alternative age cohorts				
Cohort of sex ratio	10–14	0.209**	0.289*	-0.852
	15–19	0.249***	0.284*	-1.166
	20–24	0.106	0.299*	-1.376**

Notes: Data are from the 2010 CFPS survey. In column (3), education expenditure is measured for first-born children who are at least two years old. The difference in the effect of sex imbalance between first-son and first-daughter families is reported in percentage points (β_3). In panels A and B, p-values of Hausman's general specification test for the equality of β_3 are given in square brackets. For IV regressions in panel C, see Appendix Table A5 for results regressing sex ratios on excluded instruments in lieu of first-stage results. Estimations are weighted by the CFPS survey sampling weights. Standard errors are clustered at the county level, and given in parentheses in panel C.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

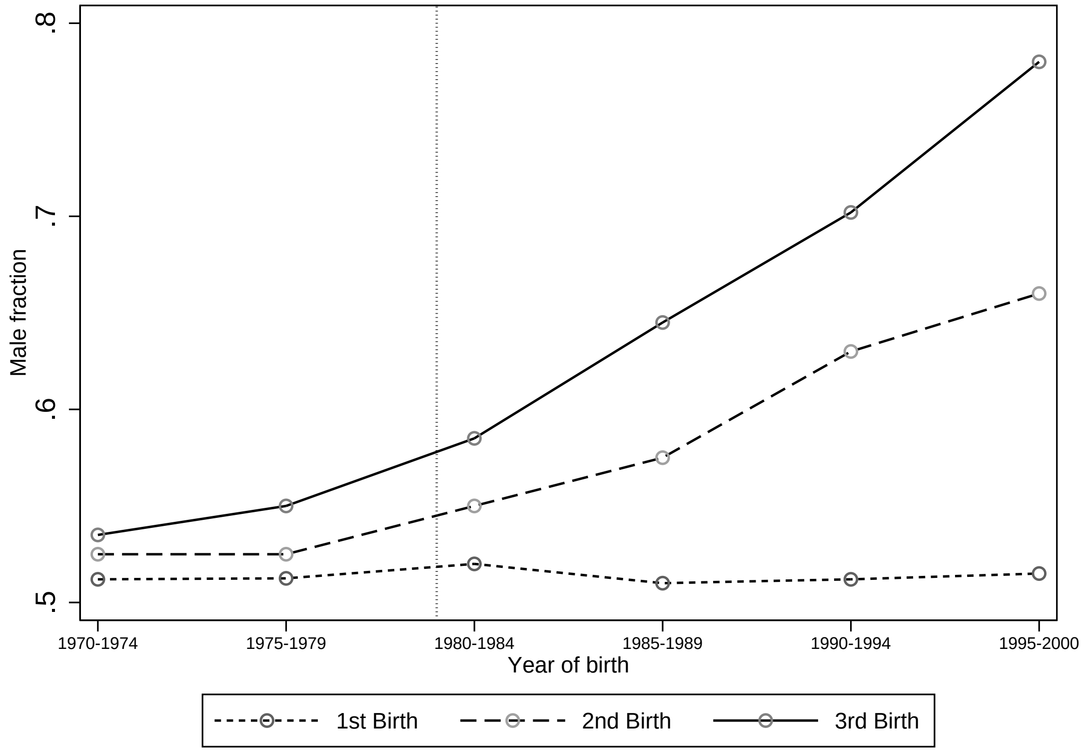

Figure 1 Male fraction of births by birth order in China

Notes: Data are from Ebenstein (2010). The figure shows a steep rise in the sex ratio over the past decades, and the imbalance comes from gender selection among second- and higher-order births, rather than among first-order births.

Online Appendices

A Proof of Proposition 6

Recall the first order condition for private investment by boys is:

$$c_x(x_B^*, y_B^*) = \lambda_B + (1-\lambda_G)\theta_B. \tag{27}$$

Totally differentiating;

$$c_{xx}^B(.)\frac{dx_B}{d\lambda_B} + c_{xy}^B(.)\frac{dy_B}{d\lambda_B} = 1. \tag{28}$$

$$c_{xx}^G(.)\frac{dx_G}{d\lambda_B} + c_{xy}^G\frac{dy_G}{d\lambda_B} = -\theta_G. \tag{29}$$

$$c_{xy}^B\frac{dx_B}{d\lambda_B} + c_{yy}^B\frac{dy_B^*}{d\lambda_B} = \left[v_B''(.) + \frac{1-\lambda_G}{1-\lambda_B}\theta_B v_G''(.)\right]\left(\frac{dy_B}{d\lambda_B} + \frac{dy_G}{d\lambda_B}\right) + \frac{1-\lambda_G}{(1-\lambda_B)^2}\theta_B v_G'(.). \tag{30}$$

$$c_{xy}^G\frac{dx_G}{d\lambda_B} + c_{yy}^G\frac{dy_G^*}{d\lambda_B} = \left[v_G''(.) + \frac{1-\lambda_B}{1-\lambda_G}\theta_G v_B''(.)\right]\left(\frac{dy_B}{d\lambda_B} + \frac{dy_G}{d\lambda_B}\right) - \frac{1}{(1-\lambda_B)}\theta_B v_G'(.). \tag{31}$$

B Sex imbalance, children's human capital outcomes, and potential mechanisms

Table A8 presents estimation results on the effect of sex imbalance on the formation of human capital using equation (1). Dependent variables in columns (1) and (2) are children's cognitive outcomes, defined for first-born children who are at least ten years old. These outcomes are measured by latest class rankings in mathematics and Chinese examinations, and set to one minus the rank over the total number of students in the class so that a larger value implies a better result. (For instance, ranking first in a class of 50 students gives a value of 0.98, while ranking last gives a value of zero.) The negative estimates for the interaction-term coefficient in both columns are large and statistically significant, which implies that a high sex ratio adversely impacts academic achievement or learning outcomes of boys relative to girls.

In columns (3) and (4), we turn to children's non-cognitive outcomes—interpersonal communication skills including openness and cooperation, again defined for first-born children who are at least ten years old. After interviewing each child, the CFPS interviewers were asked to evaluate communication skills of the child. Children's behavior is ranked from one to seven, where a larger number means a better performance. We recode interviewers' ratings as binary variables, a value of one including evaluations of five, six, and seven. Estimates show that an increase in the sex ratio from 1.08 in 2002 to 1.18 in 2010 significantly reduces the fraction of boys exhibiting openness by 5.0 percentage points relative to girls. The analogous statistic for cooperation is 5.7 percentage points.

Columns (5) and (6) report results on health outcomes, measured by z-scores of the child's body weight and height transformed using international child growth standards (UK reference growth charts) as reference. The means of the z-scores are negative, as can be seen from the table. The empirical pattern revealed is similar to the pattern for cognitive and non-cognitive skills. The effect of a 0.1 increase in the sex ratio is weight of a boy being 0.09 standard deviation further below the average of comparable international children, and height being 0.02 standard deviation further below, relative to a girl.

In summary, results in Table A8 suggest that sex imbalance hurts human capital development of boys relative to that of girls. We propose two underlying reasons, (i) distortion in premarital investments, or specifically, underinvestment in education, and (ii) parental migration that results in a shortage of parenting inputs and mental costs related to family separation (Lyle, 2006; McKenzie and Rapoport, 2011; Zhang et al., 2014). The absentee-father problem is magnified for boys, as fathers are more important in modelling social roles for sons than for daughters (Lundberg and Rose, 2002). Our results using the CFPS survey data indicate that migration, especially that of fathers, is accompanied by less satisfactory human capital outcomes, less time devoted to studying and physical exercise, as well as worse psychological well-being for children left behind (see Appendix Table A9). These results indicate that migration is a potential mechanism through which sex imbalance hurts boys' human capital development, especially when taking into account the possibility that parental inputs may be complements to children' own efforts.

C Appendix tables and figures

Table A1 Housing, education, and marital status of men

Dependent variable	Marital status of men (married=1)				
	(1)	(2)	(3)	(4)	(5)
High-quality housing	0.019***				0.013***
(costs $\geq$ 50k=1)	(0.004)				(0.004)
High-quality housing		0.045***			0.044***
(private bathroom=1)		(0.004)			(0.004)
High education			0.002		
(high school and above=1)			(0.004)		
High education				0.010**	0.005
(college and above=1)				(0.005)	(0.005)
Age	0.461***	0.460***	0.461***	0.460***	0.460***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Age square	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Hukou (urban=1)	0.018***	0.015***	0.024***	0.020***	0.008**
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Observations	94,457	94,457	94,457	94,457	94,457
R-squared	0.216	0.217	0.216	0.216	0.217
Dependent variable mean	0.440	0.440	0.440	0.440	0.440
Model	OLS	OLS	OLS	OLS	OLS

Notes: Data are from the 2000 China Population Census. The sample includes men who were: (i) between the ages of 20 and 40; (ii) from families who bought or built a new house between 1997 and 2000; and (iii) unmarried before the house was bought or built. Since the houses were newly got in families with an unwedded son of marriage age, they were most likely for the son's marriage purpose. The outcome variable is a dummy that equals one if the son got married during that period (1997–2000), and zero if he remained single. Housing condition is measured by: (i) a dummy variable that equals one if the new house cost no less than ¥50,000 (about 6.3 times per capita GDP in 2000); and (ii) a dummy variable that equals one if the new house has a private bathroom (as opposed to a shared bathroom). Education level is measured by: (i) a dummy variable indicating whether the man had at least a high school diploma; and (ii) a dummy variable indicating whether the man had at least a college degree. Standard errors are given in parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table A2 Marital status by age, gender, and education level

Age cohort	Secondary school		High school		College and above	
	Male	Female	Male	Female	Male	Female
A: Share of population divorced						
22–31	0.011	0.009	0.007	0.009	0.003	0.004
32–41	0.024	0.018	0.027	0.038	0.018	0.034
42–51	0.024	0.019	0.029	0.047	0.022	0.052
52–61	0.018	0.019	0.019	0.033	0.017	0.042
B: Share of population ever married						
22–31	0.636	0.780	0.505	0.628	0.363	0.453
32–41	0.944	0.984	0.943	0.968	0.945	0.955
42–51	0.979	0.996	0.985	0.992	0.989	0.987
52–61	0.985	0.997	0.992	0.995	0.995	0.990
C: Divorce rate						
22–31	0.018	0.011	0.013	0.014	0.008	0.010
32–41	0.026	0.018	0.029	0.039	0.019	0.036
42–51	0.024	0.019	0.030	0.047	0.022	0.053
52–61	0.018	0.019	0.020	0.033	0.017	0.042

Notes: Data are from the 2010 China Population Census.

Table A3 Parental migration and gross family income

Dependent variable	Gross family income, thousand			
	(1)	(2)	(3)	(4)
Paternal migration	6.935***			
	(2.447)			
Maternal migration		8.891***		
		(3.093)		
At least one parent migration			7.065***	
			(2.248)	
Both parents migration				11.672***
				(3.702)
Observations	4,314	4,314	4,314	4,314
R-squared	0.191	0.190	0.191	0.189
Dependent variable mean	32.1	32.1	32.1	32.1
Percentage increase (migration=1)	21.6	27.7	22.0	36.4
Model	OLS	OLS	OLS	OLS
Other controls?	YES	YES	YES	YES
County fixed effects?	YES	YES	YES	YES

Notes: Data are from the 2010 CFPS survey. The parental-migration effect on gross family income is reported in both percentage points and percentages. Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table A4 First-stage results—Child-gender measures are instrumented

Second-stage dependent variable	Paternal migration	House construction area, log sq.m	Education expenditure, thousand
	(1)	(2)	(3)
A: Endogenous variable is having any son			
(1) IV			
First son	1.206***	1.213***	1.224***
	(0.233)	(0.229)	(0.252)
R-squared	0.630	0.638	0.611
(2) IV, adding number of children			
First son	1.200***	1.183***	1.214***
	(0.204)	(0.193)	(0.223)
Number of children	0.279***	0.282***	0.276***
	(0.024)	(0.022)	(0.023)
R-squared	0.721	0.727	0.703
(3) IV, adding number of children & interaction			
First son	0.442***	0.444***	0.450***
	(0.112)	(0.111)	(0.121)
Number of children	-0.337***	-0.326***	-0.335***
	(0.028)	(0.026)	(0.027)
Number of children * Having any son	0.525***	0.520***	0.521***
	(0.018)	(0.018)	(0.017)
R-squared	0.930	0.931	0.926
B: Endogenous variable is share of sons			
(1) IV			
First son	1.113***	1.099***	1.123***
	(0.165)	(0.156)	(0.177)
R-squared	0.821	0.825	0.809
(2) IV, adding number of children			
First son	1.112***	1.093***	1.121***
	(0.160)	(0.149)	(0.172)
Number of children	0.054***	0.057***	0.053***
	(0.012)	(0.011)	(0.012)
R-squared	0.825	0.830	0.813
(3) IV, adding number of children & interaction			
First son	0.459***	0.455***	0.460***
	(0.116)	(0.112)	(0.124)
Number of children	-0.203***	-0.195***	-0.201***
	(0.012)	(0.012)	(0.012)
Number of children * Share of sons	0.491***	0.486***	0.487***
	(0.020)	(0.019)	(0.019)
R-squared	0.923	0.925	0.919
Observations	4,314	4,169	3,978

Notes: Data are from the 2010 CFPS survey. In column (3), education expenditure is measured for first-born children who are at least two years old. The instrument for having any son and the share of sons is the first-son dummy; see panel C, Table A4 for second-stage results. Estimations are weighted by the CFPS survey sampling weights. Standard errors are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table A5 Regressing sex ratios on variables for implementation of family planing policy

Dependent variable	Sex ratio		
	(1)	(2)	(3)
Policy-violation penalty	0.004***	0.004***	0.004***
	(0.000)	(0.000)	(0.000)
Quota of births	0.034***	0.031***	0.037***
	(0.005)	(0.006)	(0.006)
Policy-violation penalty * Minority	-0.004***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)
Quota of births * Minority	-0.025**	-0.019*	-0.027**
	(0.011)	(0.011)	(0.011)
Observations	4,314	4,169	3,978
R-squared	0.663	0.653	0.663
Other controls	Paternal migration estimation	House construction area estimation	Education expenditure estimation

Notes: Data are from the 2010 CFPS survey. Other controls include controls in the respective IV estimations plus province dummies (here both the dependent variable and key explanatory variables are at the county level). Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table A6 Stated purpose of migration remittances

Dependent variable	Migration purpose	
	For children's marriage	For children's education
	(1)	(2)
First son * Sex ratio (β_3)	0.179**	0.096
	(0.079)	(0.262)
Observations	1,071	1,071
R-squared	0.213	0.272
Model	OLS	OLS
Other controls?	YES	YES
County fixed effects?	YES	YES

Notes: Data are from the 2010 CFPS survey. The difference in the effect of sex imbalance between first-son and first-daughter families is reported in percentage points (β_3). Estimations are weighted by the CFPS survey sampling weights. Standard errors are given in parentheses.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table A7 Heterogenous effects of sex imbalance across families

Dependent variable	Paternal migration	House construction area, log sq.m	Education expenditure, thousand
	(1)	(2)	(3)
Benchmark: First son * Sex ratio (β_3)	0.235**	0.413**	-1.663**
A: Families with a first child above the age of 11			
First son * Sex ratio (β_3)	0.254**	0.846**	-0.265
	(0.119)	(0.392)	(1.073)
Observations	1,811	1,745	1,811
R-squared	0.162	0.265	0.369
Dependent variable mean	0.092	4.656	1.526
B: Families with a first child below the age of 11			
First son * Sex ratio (β_3)	0.284**	0.115	-2.651*
	(0.110)	(0.221)	(1.391)
Observations	2,503	2,424	2,167
R-squared	0.151	0.361	0.357
Dependent variable mean	0.102	4.646	1.492

Notes: Data are from the 2010 CFPS survey. In column (3), education expenditure is measured for first-born children who are at least two years old. Panel A includes a sample of families with a first child above the age of 11, and panel B includes a sample of families with a first child below the age of 11. The difference in the effect of sex imbalance between first-son and first-daughter families is reported in percentage points (β_3). Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table A8 Sex imbalance and children's human capital outcomes

Dependent variable	Cognitive skills		Non-cognitive skills		Health outcomes	
	Math ranking	Chinese ranking	Openness	Cooperation	Weight, z-score	Height, z-score
	(1)	(2)	(3)	(4)	(5)	(6)
First son * Sex ratio (β_3)	-0.734***	-0.567**	-0.498**	-0.572***	-0.907**	-0.179
	(0.237)	(0.246)	(0.250)	(0.200)	(0.412)	(0.605)
Observations	1,154	1,154	2,125	2,125	4,137	3,870
R-squared	0.618	0.641	0.405	0.457	0.265	0.261
Dependent variable mean	0.692	0.702	0.859	0.729	-0.505	-0.639
Percentage difference sex ratio+0.1	-10.6	-8.1	-5.8	-7.9	-18.0	-2.8
Model	OLS	OLS	OLS	OLS	OLS	OLS
Other controls?	YES	YES	YES	YES	YES	YES
County fixed effects?	YES	YES	YES	YES	YES	YES

Notes: Data are from the 2010 CFPS survey. Human capital outcome is measured for the first-born child in a family. In columns (1)–(4), the sample excludes families in which the first child is below ten years old. The difference in the effect of sex imbalance between first-son and first-daughter families is reported in both percentage points (β_3) and percentages (β_3/dependent variable mean). Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

1139

Table A9 Parental migration and child development

	Father			Mother		
	At-home mean (1)	Migration mean (2)	Difference (3)	At-home mean (4)	Migration mean (5)	Difference (6)
A: Child's human capital outcomes						
School math exam ranking	0.683	0.646	0.037*	0.679	0.686	-0.007
School Chinese exam ranking	0.698	0.673	0.025	0.695	0.688	0.007
Openness	0.862	0.881	-0.019	0.863	0.883	-0.020
Cooperation	0.727	0.678	0.049*	0.723	0.650	0.073
Weight, kg	29.03	27.89	1.140*	28.97	26.43	2.540**
Height, m	1.286	1.259	0.027**	1.284	1.255	0.029
B: Child's time allocation on weekend, hours						
Homework and revision	2.006	1.718	0.288***	1.981	1.803	0.178
After-school tuition	0.399	0.129	0.270***	0.371	0.347	0.024
Extracurricular reading	0.720	0.604	0.116**	0.713	0.521	0.192**
Physical exercise	0.336	0.274	0.062*	0.332	0.252	0.080
Observations						2,245
C: Child's psychological well-being						
Happiness	0.465	0.369	0.096***	0.459	0.290	0.169***
Optimism about the future	0.409	0.398	0.011	0.410	0.323	0.087*
Relationship with others	0.341	0.280	0.061**	0.337	0.242	0.095*
Popularity	0.285	0.233	0.052**	0.281	0.226	0.055
Observations						2,259

Notes: Data are from the 2010 CFPS survey. For more information on variables in panel A, see notes to Table A8. In panels B and C, the sample excludes families in which the first child is below ten years old. Differences between non-migrant and migrant families are reported in columns (3) and (6); H_0 is difference=0 and H_1 is difference>0.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

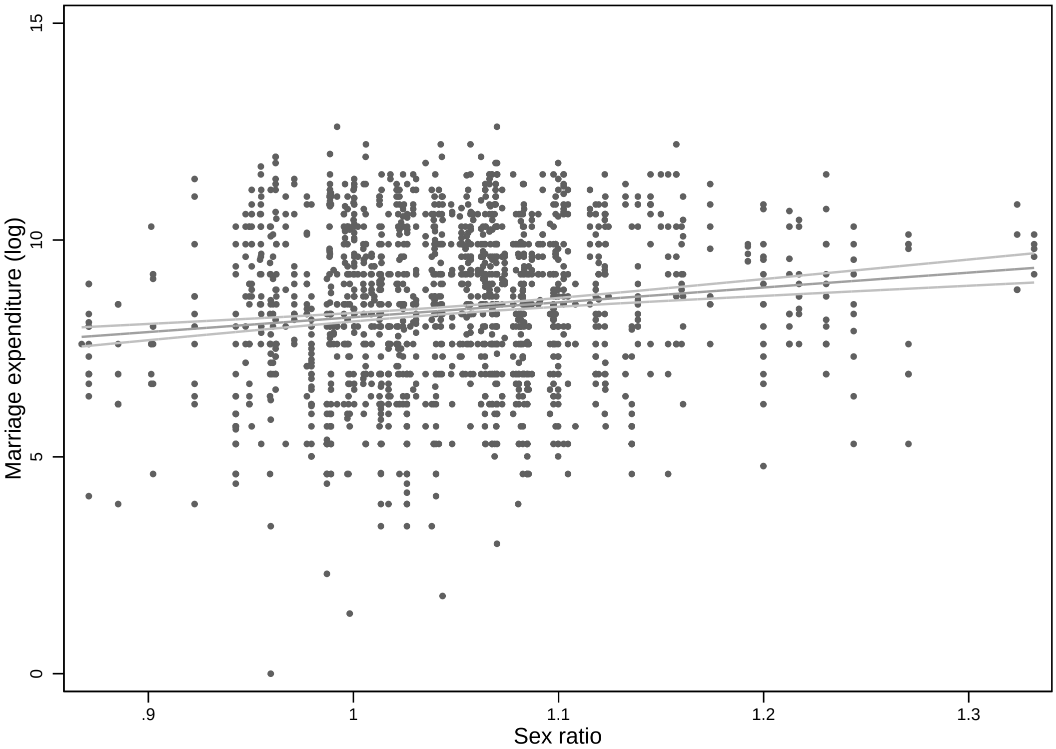

Figure A1 County-Level Sex Ratio and Marriage Expenditure in China

Notes: Data on county-level sex ratios are from the 2010 China Population Census. Data on marriage expenditures are from the 2010 CFPS survey.

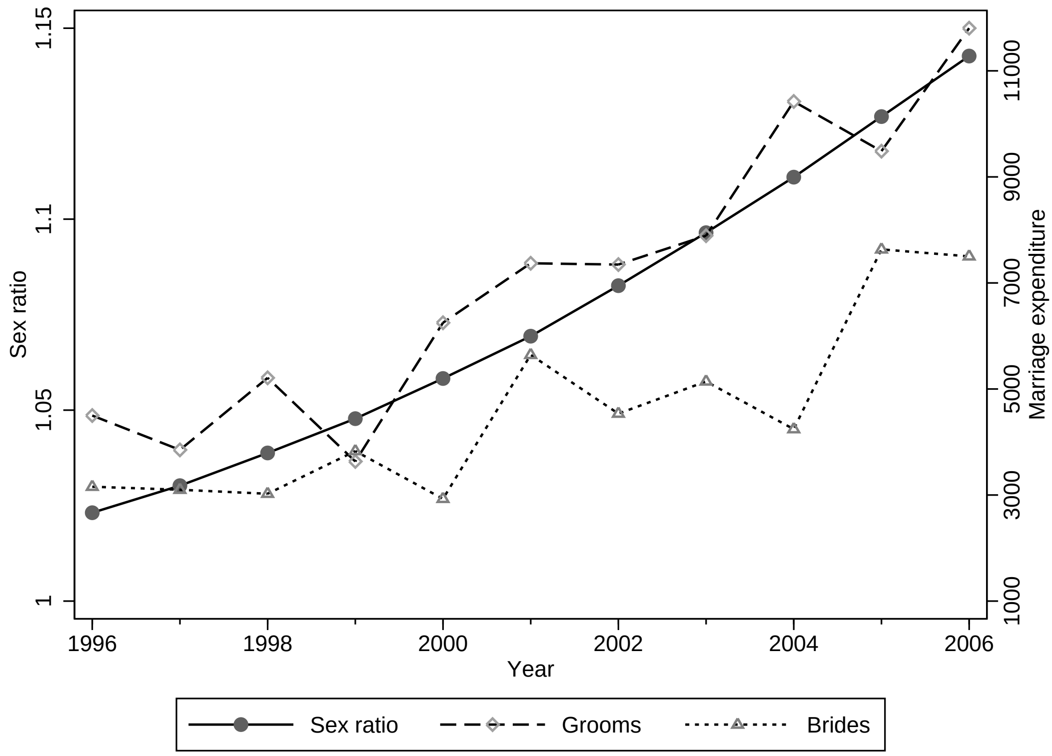

Figure A2 Trends in Sex Ratio and Marriage Expenditure in China

Notes: Data on sex ratios are projected from the 2010 China Population Census. For example, the sex ratio for the cohort between the ages of zero and 15 in 2006 is calculated using the cohort between the ages of four and 19 in 2010, since these two cohorts are supposed to be the same. Data on marriage expenditures are from Brown et al. (2011).

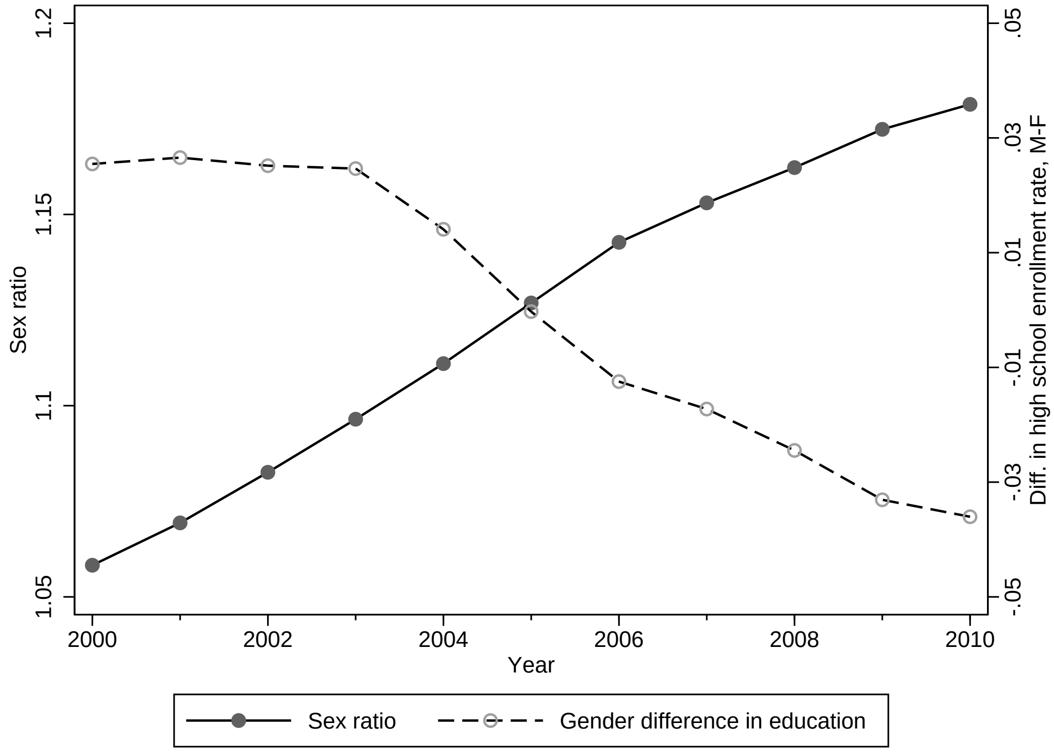

Figure A3 Trends in Sex Ratio and Gender Difference in Education in China

Notes: Data on sex ratios and high school enrollment rates are projected from the 2010 China Population Census. For example, the sex ratio for the cohort between the ages of zero and 15 in 2006 is calculated using the cohort between the ages of four and 19 in 2010, since these two cohorts are supposed to be the same. The correlation coefficient between sex ratio and gender difference in high school enrollment rate between 2000 and 2010 is -0.972 (with the 95 percent confidence interval: -0.993 to -0.893).

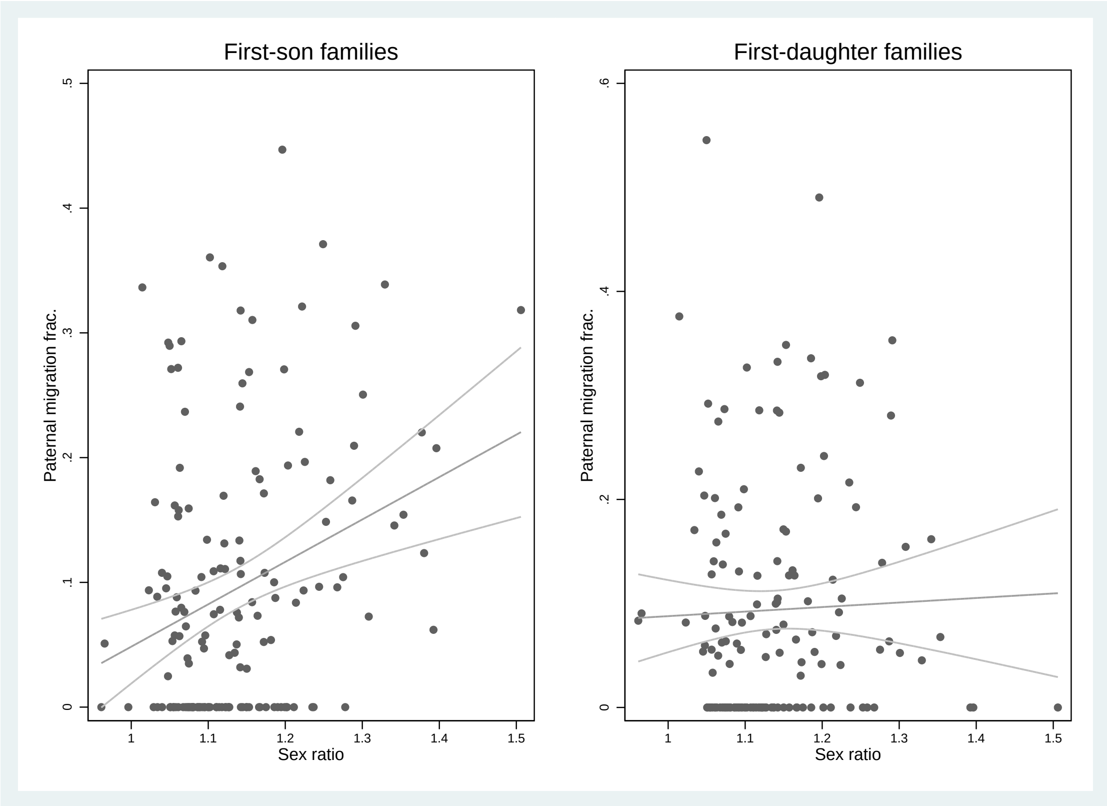

Figure A4 County-Level Sex Ratio and Father's Migration in China

Notes: Data on county-level sex ratios are from the 2010 China Population Census. Data on father's migration are from the 2010 CFPS survey. The figure shows that the probability of having a migrant father increases with the sex ratio for families with a first-born son; this probability does not change with the sex ratio for families with a first-born daughter.

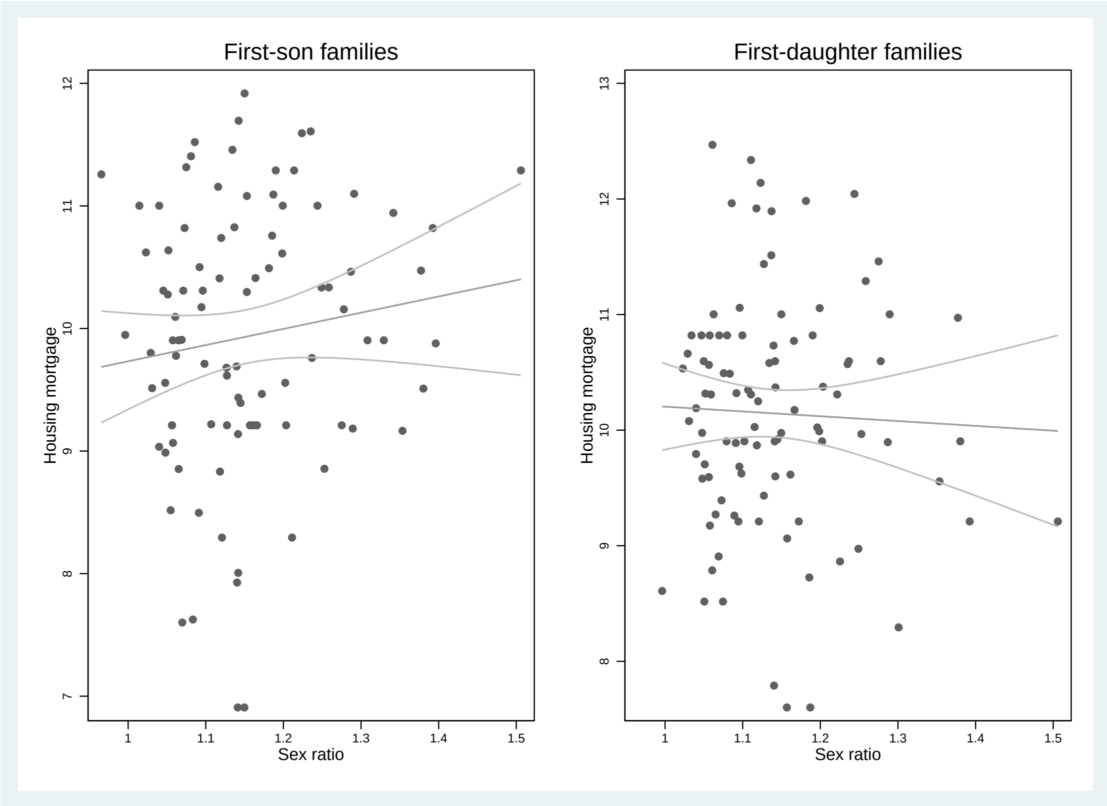

Figure A5 County-Level Sex Ratio and Housing Mortgage in China

Notes: Data on county-level sex ratios are from the 2010 China Population Census. Data on housing mortgages are from the 2010 CFPS survey. The figure shows that housing mortgage increases with the sex ratio for families with a first-born son; housing mortgage does not change with the sex ratio for families with a first-born daughter.

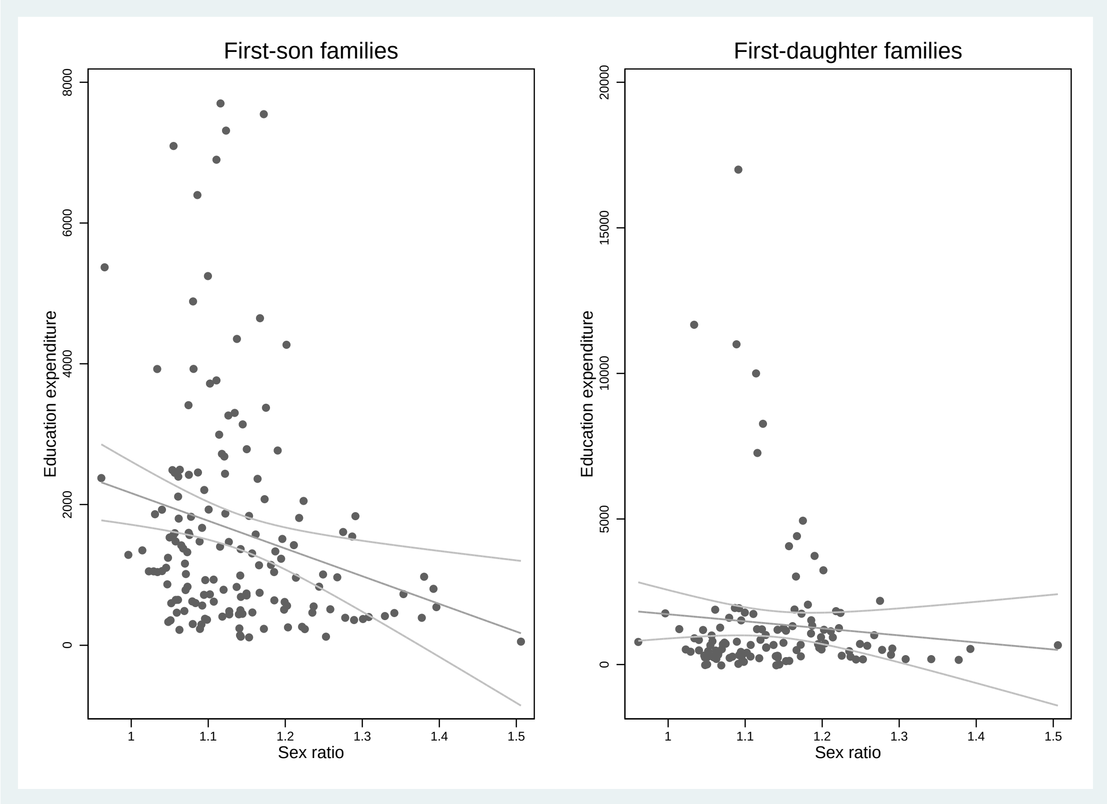

Figure A6 County-Level Sex Ratio and Education Expenditure in China

Notes: Data on county-level sex ratios are from the 2010 China Population Census. Data on education expenditures are from the 2010 CFPS survey. The figure shows that education expenditure decreases with the sex ratio for families with a first-born son; education expenditure does not change with the sex ratio for families with a first-born daughter.