Strategic Parental Investments in a Competitive Marriage Market*

V. Bhaskar† Wenchao Li‡ Junjian Yi§

April 4, 2023

Abstract

Marriage market competition, coupled with imperfect commitment to resource sharing, may distort parental investments composition. When sex ratio is male-biased, parents of boys will overinvest in housing, potentially crowding out human capital investment. We examine the effect of more male-biased local sex ratio upon investments in boys relative to girls, using nationally representative Chinese data and a novel instrument to overcome the usual difficulty that sex ratios are related to unobserved son preference. Parents with boys increase labor supply and migrate to improve investments. They strategically respond to the sex ratio imbalance, increasing housing investment while reducing educational investment.

Key words: Premarital investments; Imperfect commitment; Sex ratio imbalance; Public goods; Human capital investment; Human capital development

JEL Codes: D13; J12; J13; J24

*We are grateful to Shang-Jin Wei, Xiaobo Zhang, Dan Ackerberg, and Dilip Mookherjee for extremely useful comments.

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

‡Corresponding author: Wenchao Li. School of Economics and Management, Tongji University; E-mail: `wenchaoli.1022@gmail.com`;

§National School of Development, Peking University; E-mail: `junjian.yi@gmail.com`.

1 Introduction

China's male-biased sex ratio and the consequent severe marriage market competition faced by boys has long been a cause for concern. Wei and Zhang (2011) show that China's high savings rate is driven, in substantial part, by parents who save more in order to improve their son's competitiveness on the marriage market. This is supported by theoretical work showing that a male-biased sex ratio leads parents to increase total investments in their sons (Hoppe et al., 2009; Bhaskar and Hopkins, 2016). Bhaskar et al. (2023) suggest a more nuanced picture. They set out a model where prospective marriage partners cannot make binding commitments to share labor income at the time of marriage, so that housing provides a more credible commitment to share resources. The model predicts that if men face greater marriage market competition than women, but also have greater bargaining power post-marriage, this will induce their parents to overinvest in housing, potentially crowding out human capital investment.

The assumptions invoked by Bhaskar et al. (2023) are compelling in the case of China. Before marriage, prospective brides are in an enviable position due to excess of men in the marriage market. After marriage, the traditional power of husbands reasserts itself and divorce is prohibitively costly. Drawing data from a nationally representative Chinese household survey, we examine the effect of county-level sex ratios (defined as the ratio of boys to girls) on investment decisions by parents with a first-born son, using those with a first-born daughter in the same county as a comparison group. That is, our focus is on the coefficient on $B_{ic} \times R_c$, where B_{ic} is an indicator variable that takes value one if the first child of family i is a boy and R_c is the county-level sex ratio. The gender of the first child is plausibly random, since gender selections in China typically occur at second and higher order births (see Figure 1). We will discuss concerns regarding the endogeneity of the county-level sex ratio shortly.

Our results show that parents with boys are more likely to increase labor supply and in particular, to become migrant workers, as the sex ratio gets more male-biased. In China, migration substantially augments family income, thereby permitting larger investments in children. Therefore, our finding is consistent with Wei and Zhang (2011). Based on our estimate, a one standard deviation increase in the local sex ratio is associated with a 25 percent increase in the probability of having a migrant father for first-son families relative to first-daughter families.[1] We obtain qualitatively similar findings for the migration of the mother.

[1] A one standard deviation of county-level sex ratios is about 0.1 in 2010, which is also China's overall increase in the sex ratio during 2002–2010.

More importantly, parents strategically respond to the sex ratio imbalance in the composition of their investments. They invest more in housing but less in child education when they have a son who faces marriage pressure. Specifically, a one standard deviation increase in the sex ratio raises housing construction area for first-son families by 4.8 percent relative to first-daughter families, house ownership by 3.2 percent, and mortgage loans by 34.7 percent. The same increase in the sex ratio reduces annual per child education expenditure for first-son families by 11.8 percent relative to first-daughter families, and the probability of having an education fund by 12.8 percent.

An important challenge to our empirical work is that sex ratios are not randomly assigned and are plausibly related to regional variations in son preference, which may independently affect parental investments as a function of the sex of their child. In order to systematically deal with this issue, we set out a behavioral model of parental investments. This model shows that bias arises from county-level variations in son preference (rather than individual-level variation). Reassuringly, the theoretical model in Bhaskar et al. (2023) implies that unobserved heterogeneity in son preference across counties would cause a positive correlation in both dimensions of parental investments, whereas our empirical work finds differential effects of an increase in the sex ratio—it raises housing investments in boys (relative to girls), while reducing educational investments. Furthermore, in our empirical work, we include county fixed effects and also use several proxy variables for regional variations in son preference.

Since concerns about endogeneity may remain, we construct a novel instrument for the county sex ratio: the real level of penalties imposed on healthcare workers for abusing ultrasound machines to disclose the sex of fetuses. We show that penalties significantly reduce the local sex ratio imbalance. Prior work has used penalties imposed on parents violating the one-child policy as an instrument for analyzing the effect of sex ratios (Wei and Zhang, 2011; Ebenstein, 2011). These penalties on additional births make parents be more likely to practice sex selections, and therefore affect those who determine the outcomes of our interest. In contrast, we focus on the supply of sex-selection services and explore the quasi-experimental variation in sex ratios induced by the supply. Since the early 1980s, ultrasound machines were introduced for disease diagnosis, but were unintendedly used to determine fetus gender, followed by sex-selective abortions (Ebenstein et al., 2013). The acquisition of the machines related to the local level of healthcare infrastructure and the training of healthcare workers who operate them. It also related to local import regulations, since a fraction of the machines were imported (Zeng et al., 1993).

As such, our supply-side instrument for sex ratio is not significantly correlated with parental son preference, which is supported both by existing studies and by our findings.

Moreover, the instrument does not directly affect fertility and other parental decisions, so that the exclusion restriction is more plausibly satisfied. Our estimates using instrumental variables (IV) confirm our qualitative findings.

A final challenge to our empirical work is endogenous fertility. Parents whose first child is a girl are more likely to have a second child, than those whose first born is a boy.[2]. Consequently, the parents of a first-born girl are more likely to be resource constrained. By explicitly modeling fertility, we show that the key coefficient of interest remains consistent when we allow for a differential fertility response to the gender of the first child. More precisely, we show that gender-based stopping rules will affect the magnitude of parental investments in families with a first born boy relative to those with a first born girl—i.e. the coefficient on B_{ic}. However, these rules do not affect the coefficient on the interaction term—i.e. on $B_{ic} \times R_c$.

An alternative strategy for dealing with endogenous fertility is to condition upon fertility. Our explicit modeling of fertility allows us to examine the conditions under which such conditioning is valid. We show that our qualitative results are similar if we restrict the sample to families with one child. We also use the unrestricted sample to estimate a specification where the key explanatory variable is the *proportion* of boy children $\times$ regional sex ratio, with instruments for the proportion of boys, for family size, and for the county-level sex ratio.

Finally, we investigate the long-term implications of the strategic investment patters for human capital development. Since parental absences due to migration have an adverse effect on children, and since emphasis on housing investment rather than educational investment may hinder children's development, they may combine to result in adverse development consequences. We measure child human capital in terms of cognitive, non-cognitive, and health outcomes, and use the same empirical strategies as before. Our results suggest that the sex ratio imbalance inducing parents with a boy to migrate and to reduce educational investment, may undermine the human capital development of the boy.

The paper is organized as follows. We now discuss related prior work. Section 2 provides background information. Section 3 describes data and variables. Section 4 presents empirical strategies and results. Section 5 connects results with proposed mechanisms. Section 6 explores heterogeneity across rural and urban and the implications for child human capital. Section 7 concludes.

[2]This is due to both one-child policy exceptions and parental son preference.

1.1 Related literature

We have already discussed the most closely related work in the introduction. Two other literatures are relevant: imperfect commitment within the household and sex ratio imbalances. Limited commitment may arise because of divorce laws (Voena, 2015) or information asymmetry between spouses (Doepke and Tertilt, 2016).[3] Lundberg and Pollak (1993) pioneered the idea of separate-spheres bargaining. The implications of limited commitment have been studied in the context of intertemporal consumption (Mazzocco, 2007), private expenditures and time use (Lise and Yamada, 2018), underinvestment in childcare and child education (Gobbi, 2018), and fertility (Doepke and Kindermann, 2019). Chiappori and Mazzocco (2017) discuss household models with limited commitment. Browning et al. (2014) focus on models of full commitment.

The literature on sex ratio imbalances includes Chiappori et al. (2002), Ebenstein (2010), Bhaskar (2011), Ebenstein (2011), Wei and Zhang (2011), Bhaskar and Hopkins (2016), Cameron et al. (2019), Li et al. (2022), Li (2021), and Edlund et al. (2013).

Finally, our work analyzes how the human capital development of children is affected by parental migration and investments, and is thus related to recent work on the importance of human capital investment, especially in the early years of a child's life (Heckman, 2006; Cunha and Heckman, 2007; Heckman, 2007; Heckman et al., 2013). Given the large returns to education in China (Heckman, 2003, 2005), the welfare consequences of the sex ratio imbalance are likely to be large.

2 Background

Male-biased sex ratios and marriage market competition The sex ratio at birth has increased dramatically in China, from 1.12 boys per girl in 1990 to 1.2 in 2000, and has stabilized at this level subsequently. This imbalance is mainly achieved by sex-selective abortions, which are driven by traditional son preference (Ebenstein, 2010) and the limitations on family size imposed by the one-child policy (Li et al., 2011).

High sex ratios contribute to the current oversupply of men and intensify their marriage competition.[4] In China, marriage markets are essentially local, as the household registration

[3]Studies based on laboratory and field experiments find strong evidence for the importance of private information in the allocation of household resources in developing countries (Ashraf, 2009; Castilla and Walker, 2013; Castilla, 2015; De Laat, 2014; Schaner, 2015).

[4]In both theoretical and empirical studies, the relationship between regional sex ratio and competition in the marriage market is well documented (e.g., Becker, 1973, 1991; Chiappori et al., 2002; Edlund, 1999; Rao, 1993; Brown et al., 2011).

(*hukou*) system provides different types of permits for individuals to live and/or work, presenting a formidable obstacle to marriage migration (Davin, 2005; Wei and Zhang, 2011).[5] Since sex ratio imbalance is large and displays significant regional variation, and since marriage markets are local, variation in sex ratio measures variation in marriage market competition.

Random first-child gender Although sex selection is rife in China, the gender of the first child in a family is not selected. Parents do not practice sex selection for the first birth because of some exceptions of the policy.[6] Data from population censuses (1982, 1990, 2000, and 2010) reveal that high sex ratios in China are driven by imbalances in second and higher order births, while the sex ratio for first births is stable and falls in the biologically normal range (Figure 1). Of families in our study sample, 51.1 percent have a son as the first-born, implying a sex ratio of 1.04, which is in line with the norm in countries without any sex selections. 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, as revealed in a balance test (panel A, Appendix Table A1). In addition, we cannot reject the null that none of the characteristics are correlated with the first-child gender, as indicated by an F-test of joint significance (panel B).

Migration as a form of labor supply In China, migration is usually temporary and circular, and is an important part of aggregate labor supply (Zhao, 1999). At the start of each year, migrant workers temporarily leave their families behind in order to increase earnings. At the end of the year, they usually return home for the Spring Festival. This temporary migration is pervasive in rural as well as urban areas. It is difficult for migrant workers to assimilate with the local population as a result of the restrictions imposed by the hukou system. Migrant workers have limited access to benefits that are available to local residents, and their children are often denied access to local public schools (Zhao, 1999).

Housing as an investment for children In China, housing has traditionally been important in determining marriage prospects, which is also revealed in recent surveys.[7] Housing

[5]The census shows that more than 90 percent of rural residents and 62 percent of urban residents live in their county of birth; 89 percent of couples are from the same county. Migrants often get married before leaving hometown, since 82 percent of migrant couples in cities are from the same place.

[6]Before 2015, a second child was officially permitted if the first one was a girl, for households in most rural areas where son preference was stronger. This "1.5 children" policy was applicable to people who accounted for more than 60 percent of the Chinese population.

[7]According to a survey conducted in 2019, over 70 percent of respondents who were single and born after 1990 stated that they wanted a property before getting married; over 80 percent of the female respondents

investment by parents plays an important role for men in the marriage market (Pierson, 2010). Not only are houses often bequeathed, most marriage-age men are unable to accumulate enough wealth to afford a house on their own. Rosenzweig and Zhang (2014) find that the young and the old save more, relative to the middle-aged, driven by the need to buy houses. It should also be noted that intergenerational co-residence is common, especially in rural China (Xie and Jin, 2015; Rosenzweig and Zhang, 2014).[8] Consequently, the quality of the family home is important, since a newly-wed woman values her privacy. Wei and Zhang (2011) show that in urban areas, families are much less likely to have an unmarried adult son staying at home if they own a house; in rural areas, they are much less likely so if the house is of higher-quality; the marriage outcomes of a daughter are unrelated to the parents' house. Wrenn et al. (2019) have a similar finding. According to a nationally representative household survey, one of the most important motivations for migrant workers leaving home to increase earnings is to buy or build a house in hometown, even if their children are still young.

Lack of marital commitment Although women are in a favorable position prior to marriage because of the sex ratio imbalance, this advantage would not necessarily convert to a greater sharing of consumption after marriage. This is in part due to the difficulty in obtaining divorce and in securing alimony. Divorce and separation are infrequent in China, and often, women are not granted alimony (Civil Code of China: Part V, 2020). Asymmetric information also implies a lack of marital commitment, since the husband may command a greater share of his own earnings. Evidence shows that in China, married men are more likely to put their earnings into their own personal savings account when these choices are private information (Xing, 2017a,b). More anecdotally, a journalist reports that about 70 percent of married men in China have private savings (Liu, 2013). The prevalence of up-front compensatory bride prices is evidence for this lack of marital commitment (Becker, 1991; Zhang and Chan, 1999).

perceived owning a property as a prerequisite for marriage, as per a recent statistical report (Ma, 2020). A news report describes China's current marriage market situation: "to those men who are looking for a Chinese wife, housing is often a necessity. ... Men are then judged by the tenure, value and size of their home as a signal of eligibility" (Chai, 2020).

[8]More than 70 percent of young couples live with the groom's parents in the first few years after marriage.

3 Data and variables

3.1 Data sources

We use sex ratios at the county level, which are obtained from the 2010 population census. As discussed earlier, each county can be treated as a local marriage market, since restrictions on internal migration imposed by the Chinese state mean that marriage markets are essentially local. We use sex ratios for the age range 10–24.[9]

We draw data on parental labor supply and investments from the China Family Panel Studies (CFPS) survey, which is widely considered nationally representative due to its large sample size and scientifically stratified multi-stage sampling design. The survey covers 25 (out of 34) provinces,[10] representing 95 percent of the Chinese population (Xie, 2012). It contains datasets with comprehensive information at the individual (both adult and child) level and the family level. The family-level dataset contains details of household characteristics and activities such as migration, expenditures, investments, income, and wealth. Detailed information on the age, gender, schooling years, occupation, and working location of family members is available in the individual-level datasets. In particular, the child dataset offers information on educational investment for each child, as well as information on child development outcomes. The datasets are linked to each other by a set of identification numbers. The household identification number allows us to group parents and children by living unit. The individual identification numbers allow us to precisely identify parent-child relationships. Outcomes of interest are thus readily linked with potential covariates, enabling a systematic empirical analysis.

3.2 Variable definitions and sample

Outcome variables of our interest are: (i) parental labor supply and (ii) parental investments in housing and child education. Our interest in labor supply follows since the main way in which parents may increase total investments in their children is by earning more. As discussed in Section 2, migration represents an important part of labor supply in China. We construct three indicator variables for the migration of the father, the mother, and either parent. As Table 1 shows, about 9.8 percent of fathers and 2.5 percent of mothers in our

[9]We use this age range because parents' cognition on marriage market pressures mainly comes from the situation of cohorts closer to marriage age. Instead of knowing the exact local sex ratio, parents are likely to estimate it based on the experience of their relatives' or colleagues' marriage-age children in finding partners, or the prevailing marriage expenditure that indicates the level of competition. Using alternative age ranges does not meaningfully change the results pattern, as in Wei and Zhang (2011).

[10]"Provinces" include also provincial-level units. Hong Kong, Macao, Taiwan, Xinjiang, Tibet, Qinghai, Inner Mongolia, Ningxia, and Hainan are not included.

sample migrate; 11.1 percent of households have at least one migrant parent. A supplementary measure of labor supply is yearly working hours, of the father and the mother.[11] Information on parental labor supply is from the CFPS adult dataset.

To measure housing investment, we construct three variables: housing construction area, an ownership dummy, and mortgage debts. Housing construction area measures total living space. The ownership dummy indicates whether the house is owned by the family, including self-constructed ones. Mortgage debt measures the amount of financing of housing investment. Table 1 shows that the mean construction area is 126.2 square meters; 83.2 percent of the families in our sample own the house; the average mortgage is 5.4 thousand yuan. Information on housing investment is from the CFPS family dataset.

To measure child educational investment, we focus on the first-born older than two, and construct two variables. The first is yearly total expenditure on the child's education.[12] The second is an indicator variable for whether the family has put aside a specialized fund for the child's education. The average yearly education expenditure is 1.5 thousand yuan per child; 29.8 percent of families have an education fund. We also examine the ratio of annual education expenditure per child to the market value of current housing; the average ratio in our sample is 0.04. Information on child educational investment is from the CFPS child dataset.[13]

In our empirical work, we extract a sample of households from the nationwide baseline survey. We restrict attention to families in which the eldest child was under the age of 15, both parents were alive and at most 50 years old, and at least one parent participated in the adult survey.[14] Our sample contains 4,304 families.

[11] Working hours variables are constructed based on three survey questions from the adult dataset: the number of working months last year, the average number of working days per month, and the average number of working hours per day. Missing values arise when any of the three questions is not answered, which explains the smaller sample sizes. We have checked that the attrition of working hours data is not correlated with migration status (Appendix Table A3).

[12] This includes tuition fees, book and stationery costs, after-class tutoring expenses, accommodation fees, and commuting fees (we exclude living expenses).

[13] Admittedly, child educational investment can be nonmonetary. But in order to measure the two types of investments in the same form (e.g., in the form of foregone consumption utility of the parents), we focus on monetary investment in child education.

[14] The child age limit is to minimize the possibility that children have started work or participated actively in household decision-making. The parental age limit is to minimize the probability of parents' 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.

3.2.1 Preliminary analyses

We conduct two preliminary analyses before more formal empirical work. First, we plot housing investment and child education investment against the local sex ratio, in Figures 2 and 3 respectively. We observe that parents with a boy invest more in housing facing marriage pressure, both in rural and urban areas, but they do not appear to invest more in the boy's education; parents with a girl invest less in housing when girls are in the short side, but they invest no less in the girl's education.

Second, we compare parental investments in sons and daughters, by the level of local marriage market competition. Appendix Table A2 shows the summary statistics of main outcome variables by local sex ratio levels and first-child gender. We find that in counties where the sex ratio is high (above the third quartile), first-son families observe significantly more parental labor supply, more housing investment, and less child educational investment. In comparison, in counties where the sex ratio is somewhat balanced (below the first quartile), there are few significant differences between the two types of families (one exception is education expenditure—parents spend more on sons' education than daughters').

4 Empirical strategies and results

4.1 The econometric model

Consider the following model for the outcome y of household i in county c:

$$y_{ic} = \beta_c + \beta_0 B_{ic} + \beta_1 (B_{ic} \times R_c) + \alpha_c B_{ic} + \alpha_{ic} B_{ic} + \tilde{\epsilon}_{ic}. \tag{1}$$

For concreteness, let us take y_{ic} to denote a measure of total investments in a child; B_{ic} is an indicator variable that takes value one if and only if the first child is a boy; R_c is the county-level sex ratio; β_c is a county fixed effect that reflects factors that uniformly affect investments in the child, independent of gender; β_0 is a measure of the average difference between investments in boys and girls, over all China, reflecting (among other things) the average level of son preference in the country—henceforth we denote it as $\bar{\alpha}$; α_c is a zero-mean variable that affects how county-level variation in son preference differentially affects investments in boys in county c; α_{ic} is a zero-mean random variable, independent of α_c, that affects how household-level variation in son preference differentially affects investments in boys by household i. We assume that $\tilde{\epsilon}_{ic}$ is uncorrelated with the other right-hand side variables—a plausible assumption given that we are explicitly modelling household-level and county-level heterogeneity in son preference.

Our focus is on the coefficient β_1 on the interaction $B_{ic} \times R_c$, which measures how variations in the sex ratio *differentially* affect investments in boys relative to girls. Putting the unobserved α_c and α_{ic} in equation (1) in the error term, we have the following regression specification:

$$y_{ic} = \beta_c + \bar{\alpha} B_{ic} + \beta_1 (B_{ic} \times R_c) + \hat{\epsilon}_{ic}, \tag{2}$$

where

$$\hat{\epsilon}_{ic} = \alpha_c B_{ic} + \alpha_{ic} B_{ic} + \tilde{\epsilon}_{ic}. \tag{3}$$

In our empirical analyses, a vector of other influences on the outcomes are included as controls—the ages of parents, their years of schooling, their political affiliation, household location (rural or urban), ethnicity, the age of the first-born, and county-level age structure for both genders. These controls are allowed to differentially affect the outcomes for boys relative to girls. Note that our empirical strategies exploit spacial variation in sex ratios across counties. This does not represent a serious deficiency since there appears no significant variation in sex ratios within a county, across years in our study period.[15]

4.2 Dealing with endogenous sex ratio

An OLS regression of equation (2) leads to a biased estimate of the coefficient β_1 on the interaction $B_{ic} \times R_c$. While the first-son dummy B_{ic} is plausibly exogenous (as discussed in Section 2), the sex ratio R_c depends upon α_c, the county-level son-preference parameter. We model this as follows:

$$R_c = \gamma_0 + \gamma_1 \alpha_c + \gamma_2 Z_c + \eta_c, \tag{4}$$

where γ_1 measures how county-level son preference affects the sex ratio, and Z_c is a vector of other influences on the sex ratio. Therefore, α_c is positively correlated with R_c (through equation 4) and is a component of the error term (equation 3). This biases the estimated effects of sex ratio upon relative investments in boys. Note however, that although the household-specific son-preference parameter α_{ic} is also a component of the error term, this does not bias our coefficient of interest, since it is uncorrelated with the county sex ratio,[16] and since B_{ic} is exogenous.

We deal with the bias problem via two alternative strategies. First, we use proxy variables X_c for α_c, the county-level son preference. This leads us to the following regression

[15] As the CFPS survey data are biennial, and we use sex ratios for the age range 10–24 in a survey year, any variation in the temporal dimension comes from only two out of 15 birth years between waves.

[16] Recall that in our explicit model, household son preference is defined to be the component that remains after taking out the mean county-level preference; so by construction, it is independent of the county-level preference.

specification:

$$y_{ic} = \beta_c + \bar{\alpha} B_{ic} + \beta_1 (B_{ic} \times R_c) + \beta_2 (B_{ic} \times X_c) + \epsilon_{ic}. \tag{5}$$

The validity of this specification relies on the controls X_c picking up the variation in α_c so that the residual variation, ϵ_{ic}, is uncorrelated with the county sex ratio.

The most prominent proxies for son preference are fertility levels in the time before people get access to the option of selecting the sex of their children. Prior to the introduction of B-scan ultrasonography in the late 1980s, parents satisfy their preference for sons by giving more births. As a result, the degree of son preference was positively correlated with fertility during that time (Ebenstein, 2011). The 1982 census data contains information on county-level fertility a year before; we use both completed fertility, defined as the average fertility of women aged 35–64 in the county, and birth rate, defined as the number of births during the year over total population. We also proxy for the local degree of son preference using geographical region dummies (we consider three regions: east, middle, and west China), adult women's average schooling years, and traditional agricultural practices measured by latitude and longitude (Alesina et al., 2018). The specifications can control for the interactions between these proxies and first-child gender so that these proxies affect boys and girls differently.

The second strategy relies on finding an instrument for sex ratio that is uncorrelated with the county-level son-preference parameter α_c, which we turn to next.

4.3 An instrument for sex ratio

While high sex ratio is likely to have many important economic and social consequences, a difficulty with empirical work on its effects is the fact that, as a macro-level variable, it is never randomly assigned. Prior work, notably that of Wei and Zhang (2011) and Ebenstein (2011), has used the intensity of the one-child policy implementation—specifically, penalties imposed on violating parents—as an instrument. Since parents who are forced to have fewer children are more likely to select for boys, and since they are also who determine the outcomes of our interest, penalties for them might affect outcomes through channels other than influencing the sex ratio.

We instead use penalties for healthcare workers who facilitate sex selection by disclosing the sex of the fetus. To motivate this instrument, the following background is useful. Since the early 1980s, ultrasound machines were introduced for disease diagnosis, representing a remarkable improvement in public health technology across China (Ebenstein et al., 2013). Unintendedly, however, these machines were used by pregnant women to determine the gender of a fetus, followed by sex-selective abortions, resulting in a surge in the sex ratio at birth (Chen et al., 2013; Chu, 2001). In 1989, the Chinese government outlawed the use of ultrasound machines for prenatal gender disclosure for nonmedical purposes and legislated, permitting substantial penalties for healthcare workers violating the rule. Using county-level average penalties relative to average household income,[17] columns 1–3 of Table 2 show that higher penalties lead to less male-biased sex ratios.

These penalties were determined based on the adoption of ultrasound machines at local (county) level. Introduced for the purpose of medical procedures and diagnostic testing, ultrasound machines related to the local level of healthcare infrastructure and physician training. Counties with better developed health infrastructure, including hospitals and clinics, had been better equipped to acquire and maintain the machines. Since the machines require skilled medical professionals to operate them, their acquisition was linked to the level of healthcare workers training. Ultrasound adoption also related to local import regulations, since a fraction of the machines were imported (Zeng et al., 1993). Counties that were less heavily regulated on medical imports had better access to the machines, from different manufacturers and at more competitive prices. Therefore, using penalties for healthcare workers who used ultrasound machines to illegally practice sex screening as an instrument enables us to explore the quasi-experimental variation in sex ratios induced by the supply of sex-selection services.

Our supply-side instrument for the sex ratio is exogenous to the local son preference. Existing studies have provided evidence for this exogeneity, which is also supported by our findings. For instance, Ebenstein et al. (2013) show that the spread of ultrasound technology did not follow any clear geographic pattern across the country, and the availability of ultrasound was broadly independent of local mothers' characteristics. They also compare counties adopted ultrasound earlier and those adopted it later and find that, before adoption, they had similar sex ratios. Similarly, Chen et al. (2013) find no preexisting differential trends in sex ratio at birth in counties with and without access to the technology. These findings suggest that ultrasound adoption was uncorrelated with attitudes towards sons or fertility. Reassuringly, our results in columns 4 and 5 of Table 2 show that penalties for healthcare workers are not significantly correlated with pre-ultrasonography fertility (the most prominent proxies for son preference as we discuss earlier). Our finding of a negative relationship between penalties and sex ratios, in columns 1–3 of this table, is also reassuring, since if these penalties were designed to counter son preference, the relationship would be positive. Moreover, penalties for healthcare workers generate variation in the supply of sex-

[17] We collected data on penalties for healthcare workers who used ultrasound machines to illegally practice sex screening during the period 1990–2000, by searching through news reports, related case files, the announcements of the policy implementation from local governments' official websites, etc.

selection services, and do not directly relate to fertility or other decisions of parents. The exclusion restriction is thus more plausibly satisfied.

In estimating equation (2) based on the IV method, the interaction of the sex ratio and the first-son dummy is instrumented by the interaction of the penalty variable and the first-son dummy. The first-stage statistics, including t-statistics (about -5.5, Table 2) and F-statistics (47.1 to 82.4, Tables 3 and 4 below), indicate a strong predictive power of the instrument. As we show in the following, using this instrument gives us the same conclusions as using the OLS method.

4.4 Results

4.4.1 Parental labor supply

Table 3 reports results on parental labor supply. In this and the following tables, estimations are weighted by sampling weights provided in the survey data; standard errors are clustered at the county level. Panel A reports OLS results.[18] Panel B addresses potential endogeneity of sex ratios by additionally controlling for interactions between pre-ultrasonography fertility at the county level and the first-son dummy, as discussed in Section 4.2. Panel C reports IV results using the instrument in Section 4.3. Reassuringly, results in panel A experience no large changes when adding son-preference proxies (including those in panel B and alternative ones in Appendix Table A4); they are also broadly consistent with IV results.[19]

In column 1 of Table 3, the outcome is an indicator for father's migration. The coefficient β_1 on $B_{ic} \times R_c$ is estimated to be 0.247 (standard error 0.104) based on panel A, which is positive and statistically significant at the five percent level. That is, high sex ratio is much more likely to induce the father to work outside the hometown when the first child is a son relative to when the first child is a daughter. A one standard deviation increase in the sex ratio (about 0.1) raises the probability of having a migrant father by 2.5 percentage points for a first-born boy relative to a first-born girl. This represents a 25 percent difference relative to the baseline father-migration probability in our sample. Based on all estimates in the three panels, the effect size ranges between 2.5–3.2 percentage points. We obtain qualitatively similar findings for mother's migration and the migration of at least one parent, in columns 2 and 3.

Since migration is a main way for Chinese parents to increase earnings and investments,

[18]OLS estimates are similar to marginal effects from Probit estimations when the outcome is binary. Estimated coefficients on controls, which are untabulated for brevity, have the expected signs and sizes.

[19]Generally, IV estimates are moderately larger, perhaps due to measurement error. In the tests of exogeneity, the null hypothesis of no difference between OLS and IV estimates cannot be rejected at conventional levels.

this suggests that total investments in sons rise when their marriage market prospects worsen—consistent with the finding in Wei and Zhang (2011). We use the yearly working hours of parents as complementary measures of their labor supply. Columns 4 and 5 of Table 3 show that working time also increases with a rise in the sex ratio for parents with first sons relative to parents with first daughters, which is in line with findings based on migration.

4.4.2 Parental investments

Table 4 reports results on multidimensional parental investments. Again, OLS results in panel A are robust to adding proxy variables for local degree of son preference in panel B (and also in Appendix Table A5) and are consistent with IV results in panel C.

The first three columns in Table 4 report housing investment results. In column 1 where the outcome is (log) construction area, the estimated coefficient on the interaction term is 0.478 (standard error 0.198) based on panel A. Thus, parents with first-born sons invest in larger houses relative to those with first-born daughters, as the sex ratio becomes more biased towards males. A one standard deviation increase in the sex ratio raises housing construction area for families with sons by 4.8 percent relative to those with daughters. Columns 2 and 3 reveal similar patterns—2.7 percentage points or 3.2 percent for house ownership and 1.87 thousand yuan or 34.7 percent for mortgage loans.

Our most interesting findings are on parental investment on child education. Column 4 of Table 4 reports results for annual education expenditure per child. The estimated interaction-term coefficient is -1.775 (standard error 0.793) in panel A, negative and statistically significant at the five percent level. Accordingly, with a 0.1 increase in the sex ratio, annual education expenditure is 178 yuan less if the child is a boy. The economic magnitude is sizable—since the mean expenditure is 1,508 yuan per child in our sample, this represents a 11.8 percent reduction. Column 5 shows that a high sex ratio reduces the probability that parents with a first-born boy put aside a specialized fund for his education relative to parents with a first-born girl; the analogous estimate is 3.8 percentage points or 12.8 percent.[20] In column 6, we check the effect of sex ratio on the ratio of education expenditure per child to the market value of current housing, and find a significantly negative effect for boys relative to girls.

[20]Estimates for child educational investment might be biased if investment measures for children with migrant parents have systematic measurement error due to these parents' less control over resources devoted to their left-behind children. To address this potential issue, we control for the migration status of parents in the specification of education expenditure. Results are reported in columns 2–4, with column 1 being the baseline, of Appendix Table A6. We also use a subsample that excludes families with any migrant parents in column 5. Results are robust in all these additional checks.

These results suggest that as marriage market competition intensifies for men, parents with sons shift their investments towards housing and away from children's education.

5 Mechanisms

In this section, we reconcile the observed parental investment patterns using the theory developed in Bhaskar et al. (2023). We build a connection between the empirical findings and mechanisms entailed by the theory, by addressing concerns about the former that are related to unobserved son preference and endogenous fertility. We also discuss competing hypotheses.

5.1 A model of premarital investments with imperfect commitment

We reconcile our empirical patterns of strategic parental investments in a competitive marriage market, using a theoretical model of multidimensional premarital investments with imperfect commitment, developed in Bhaskar et al. (2023). The model predicts that, since the division of labor earnings within marriage is more difficult to commit to, while housing can be shared equally, parents of boys facing increased marriage market competition invest more in housing in order to credibly commit to share more resources with the boy's potential wife, possibly crowding out human capital investment. In China, marital partners are unable to make binding commitments to share their labor income with a prospective spouse, as discussed in Section 2. If parents invest in a son's education, which will increase his future labor earnings, the sharing of this between spouses is determined by after-marriage bargaining, and reflects the greater bargaining power of men. If parents with a son invest in housing, which is a public good and thus non-excludable, spouses share it equally. Thus, housing investment provides a credible commitment to share resources with the future spouse. That is, a man's attractiveness as a marital partner depends not only on the total, but also on the composition of investments. This creates an incentive for parents with sons to shift their investments towards housing and away from education as marriage market competition intensifies.

5.2 Concerns regarding unobserved son preference

Our empirical work uses variation in the county-level sex ratio to measure regional variation in marriage market competition. As a result, a key concern about the interpretation of our results is that the sex ratio may depend upon unobserved variation in the degree of son preference, that would also affect parental investments in a child-gender-specific manner. We allay this concern in several ways.

First, any unobserved heterogeneity in son preference, that is positively correlated with the sex ratio, would result in a positive correlation between the sex ratio and educational investment in sons (relative to daughters), rather than the negative relation we find. Indeed, the theory of Bhaskar et al. (2023) can be extended to allow for variation in the intensity of son preference, which shows that increased son preference would cause an increase in both dimensions of investments. Thus, if OLS results are biased by the positive correlation between son preference and the local sex ratio, this bias would attenuate the negative effect on educational investment. Our finding, that variations in the sex ratio have contradictory effects on the two components of investments, is reassuring.

Second, the economic/behavioral model of investment decisions we set out, equation (1), explicitly models how the decisions are affected by unobserved son preference at the county level and at the household level. We have shown that any bias in the estimates of the main coefficient of interest—that on interaction between the first-son dummy and county sex ratio—arises from county-level son preference (but not household level). Accordingly, we control for interactions between various proxy variables for county-level son preference and the first-son dummy to show the robustness of the findings.

Third, to address the concern that our proxies are insufficient, we develop a new instrument for the county-level sex ratio: the average real level of penalties for healthcare workers who illegally practiced gender screening and gender selection. The penalties are determined at the county level, and generate variation in the supply of sex-selection services. We find that higher penalties lead to less male-biased sex ratios. By affecting the supply of sex-selection services, the instrument does not directly affect parental decisions on fertility as well as their other decisions. Thus, the exclusion restriction is plausibly satisfied. It is also reassuring that penalties are uncorrelated with proxies for the degree of son preference. Our IV estimates give similar conclusions, for both dimensions of investments, thereby buttressing our confidence in the interpretation of results.

Finally, we probe into the potential sources of heterogeneity in son preference, and control for these sources in our specification. Variation in son preference may originate from variation in gender difference in earnings or old-age insurance considerations (e.g., Zhang, 2017). We allow these factors to affect son and daughter families differently in the regressions. As in panels A–C, Appendix Tables A8 and A9, controlling for these factors—either individually or collectively—does not meaningfully change our results.[21] The estimated coefficient on

[21] We use county-level gender wage differential and the fraction of households covered by social insurance as proxies. Using gender wage ratio does not make much difference from using wage differential.

the interaction, β_1, is not significantly different from the baseline estimate based on the p-value of Hausman's general specification test. We then consider a more comprehensive set of predictors of son preference based on the high-dimensional method (Belloni et al., 2014). We select variables that are the most strongly predictive of the county sex ratio,[22] and add their interactions with first son in the specification. Our results remain robust, as in panel D, after partialling out the effect of son-preference predictors.

5.3 Endogenous fertility

Another challenge to our interpretation centers around whether the investment patterns come from endogenous fertility and the resulting resources availability differences. The gender of the first child influences family size, as gender-dependent stopping rules for fertility are widespread, especially in developing countries. Parents whose first child is a girl are more likely to have a second child than those whose first-born is a boy. Ebenstein (2011) documents the phenomenon in the case of China. If parents have more children, they would have less resources to spend on the first-born, and thus, the first child's gender may have an effect on parental investments via fertility choices. Do our results—specifically, the estimated coefficient β_1 on $B_{ic} \times R_c$—reflect in part this effect?

To address this concern, consider the following model of fertility, f_{ic}.[23] We simplify the analysis by treating fertility as a continuous rather than discrete variable:

$$f_{ic} = \tilde{\delta}_c + \delta_1(\bar{\alpha} + \alpha_c + \alpha_{ic})(1 - B_{ic}) + \tilde{\omega}_{ic}, \tag{6}$$

where $\tilde{\delta}_c$ is a county-level fixed effect; $\tilde{\omega}_{ic}$ is the random error term. First-daughter families are likely to have greater fertility, to an extent that depends upon the overall son preference of the household, which is the sum of mean son preference ($\bar{\alpha}$), county-level son preference (α_c), and household-specific son preference (α_{ic}). The coefficient δ_1 is expected to be positive. Equation (6) can be rewritten as:

$$f_{ic} = \delta_c - \delta_1(\bar{\alpha} + \alpha_c + \alpha_{ic})B_{ic} + \omega_{ic}, \tag{7}$$

where $\delta_c = \tilde{\delta}_c + \delta_1(\bar{\alpha} + \alpha_c)$, and $\omega_{ic} = \tilde{\omega}_{ic} + \delta_1\alpha_{ic}$.

[22] The initial set of factors include local adult residents' age, schooling years, household registration status, political status, marital status, the number of siblings, ethnicity, annual income, social insurance, scores for a word and a math test, depression score, whether living with parents, average household wealth and income, and local industrial structure. For each factor, we compute the average, the average for men, the average for women, and gender difference at the county level. The final set consists of 161 variables.

[23] For concreteness, we assume that fertility is chosen first, and investments subsequently.

Since the household chooses investments conditional on the level of fertility, the behavioral equation (1) for investments is modified by adding a term, $\beta_3 f_{ic}$, to the right-hand side. Therefore, we have:

$$y_{ic} = \beta'_c + \bar{\alpha}' B_{ic} + \beta_1 (B_{ic} \times R_c) + \hat{\epsilon}'_{ic}, \tag{8}$$

where $\beta'_c = \beta_c + \beta_3 \delta_c$, $\bar{\alpha}' = (1 - \delta_1 \beta_3)\bar{\alpha}$, and

$$\hat{\epsilon}'_{ic} = (1 - \delta_1 \beta_3)\alpha_c B_{ic} + (1 - \delta_1 \beta_3)\alpha_{ic} B_{ic} + \tilde{\epsilon}_{ic} + \beta_3 \omega_{ic}. \tag{9}$$

We see that endogenous fertility choice affects the coefficient on B_{ic}—mean investment differences between boys and girls. However, it does not affect our main coefficient of interest, that on $B_{ic} \times R_c$. The error term (equation 9) has the same structure as that in the original specification, which does not explicitly model fertility and its implications (equation 3). By the same argument as before, the terms involving household-specific son preference α_{ic} are uncorrelated with R_c, the county sex ratio. So if we are able to control for county-level variation in son preference α_c, or alternatively, to instrument for county sex ratio, this resolves also the problem for the case where fertility choice is modeled.

More intuitively, changes in fertility affect parental investment decisions, giving rise to a channel where the gender of the first child affects investments. However, these effects via fertility are *independent* of marriage market conditions, arising instead from resource constraints. That is, endogenous fertility has implications for the estimated effect of first-child gender, by affecting household resource constraints. However, it does not affect our estimate of β_1.

It is worth spelling out the behavioral assumption underlying the model of fertility, equation (7). The assumption allows that parents with a first-born girl are more likely to have a second child, and this likelihood increases in their son preference. However, parents do not take into account future marriage prospects of their children while making fertility decisions, which is why the county-level sex ratio does not enter the fertility equation. We view this as a plausible assumption—even though parents are conscious of marriage market considerations at the time of making investment decisions for their child, these considerations are peripheral at the time they are choosing the number of children, when cultural preferences are more important. Evidence in support of this assumption is twofold. First, county-level sex ratios display a high degree of persistence, with little evidence of self-correction; second, the correlation between past county-level sex ratios and current fertility is weak (e.g., Scharping, 2013).

To summarize, our baseline specification is not affected by the endogeneity of fertility.

5.3.1 One-child families

We now examine the implications of restricting our sample to families with only one child, so that investment variables (total investments, housing investment, and educational investment) pertain to the single child. The key question is: do we bias our results by sample selection on the basis of an endogenous variable, fertility? Parents with a son are happy to stop, or required to do so by the one-child policy; but parents with a daughter choose whether to take advantage of policy exception and have a second child, depending upon their intensity of son preference. This implies that in the restricted sample, one-daughter families have less intense son preference than one-son families. To see this more formally, we return to the basic fertility equation (6), and restrict the sample to families with fertility below a specified level $\bar{f}$. This implies that families with one girl are those with $(\alpha, \tilde{\omega})$ values that satisfy:

$$\delta_1(\bar{\alpha} + \alpha_c + \alpha_{ic}) + \tilde{\omega}_{ic} \leq \bar{f}, \tag{10}$$

and families with one boy satisfy:

$$\tilde{\omega}_{ic} \leq \bar{f}. \tag{11}$$

In other words, in one-boy families, the distribution of son preference equals the unconditional distribution; in one-girl families, the values of α_c and α_{ic} are smaller (in the sense of first-order stochastic dominance) than the unconditional distributions.

Assume that α_{ic} and α_c can be written in terms of the following linear projections:

$$\alpha_c = \theta_1 B_{ic} + \eta_c, \tag{12}$$

$$\alpha_{ic} = \theta_2 B_{ic} + \eta_{ic}, \tag{13}$$

where η_c and η_{ic} are the residual components of son preference that are uncorrelated with B_{ic}. Substituting for α_{ic} and α_c in our basic behavioral equation (1), gives us:

$$y_{ic} = \beta_c + (\bar{\alpha} + \theta_1 + \theta_2)B_{ic} + \beta_1(B_{ic} \times R_c) + \hat{\epsilon}_{ic}, \tag{14}$$

where

$$\hat{\epsilon}_{ic} = \eta_c B_{ic} + \eta_{ic} B_{ic} + \tilde{\epsilon}_{ic}. \tag{15}$$

Therefore, the estimated coefficient on $B_{ic} \times R_c$ is consistent under the same conditions—we are able to control for the county-level son preference or instrument for the sex ratio—but the estimated coefficient on B_{ic} is inconsistent. The argument is valid under the same behavioral assumption as before—the fertility decisions of parents with a first-born girl may depend upon their son preference, but not upon the future marriage prospects of their children. Tables 5 and 6 are based on the restricted sample of families with one child (about two thirds of the unrestricted sample); the estimated coefficients on $B_{ic} \times R_c$ are broadly consistent with those in our main estimations.

5.3.2 Controlling for fertility in the unrestricted sample

In the unrestricted sample, with some families having multiple children, one complication arises from the fact that they have to consider the marriage prospects also of subsequent children. Families with a first daughter often have a second child, and this child could be a boy whose marriage prospects are affected by the sex ratio imbalance. To isolate the effect of marriage market considerations, we use a variable that better represents sex-ratio pressure that parents face: the *proportion* of boys amongst children.[24] Since we measure child gender at the household level, we analyze investments by the household:

$$\check{y}_{ic} = \beta_c + \bar{\alpha}\check{B}_{ic} + \beta_1(\check{B}_{ic} \times R_c) + \check{\epsilon}_{ic}, \tag{16}$$

where $\check{y}_{ic}$ is household investments, and $\check{B}_{ic}$ is the proportion of boys amongst the children of household i.

We are interested in the coefficient on $\check{B}_{ic} \times R_c$. In Tables 7 and 8, panel A shows OLS results. As the sex ratio increases, families with a larger proportion of sons supply more labor, and invest more in housing and less in child education. Yet a new problem is introduced using equation (16), that the proportion of boys is correlated with son-preference parameters α_c and α_{ic} in the error term. We use the plausibly exogenous first-son dummy as an instrument, which strongly predicts the proportion of boys in the household (Appendix Table A7). IV results are in panel B, Tables 7 and 8, showing broadly consistent patterns.

Our main empirical analysis focuses on the gender of the first-born child, since it represents not only an exogenous, but also a consistent variable as a proxy for child-gender composition and thus sex-ratio pressure that parents face, despite potentially endogenous fertility. In our data, among families whose first-born is a girl, 25 percent go on to have a son and the proportion of sons amongst children is 0.12; among families with a first son, the proportion of sons is 0.94.

We go on to address the concern about the implications of endogenous family size for

[24] We also consider the number of boys in the household, or a dummy for having any boy. This yields qualitatively similar results.

investments, by conditioning on fertility in the unrestricted sample:

$$\check{y}_{ic} = \beta_c + \bar{\alpha}\check{B}_{ic} + \beta_1(\check{B}_{ic} \times R_c) + \varphi_0 f_{ic} + \varphi_1(\check{B}_{ic} \times f_{ic}) + \check{\epsilon}_{ic}. \tag{17}$$

An assumption underlying the specification is that if we condition upon total amount of resources allocated to a child, its division between housing and education does not depend upon the number of other children or upon their investments.

The endogeneity concern related to the county-level sex ratio, R_c, can be addressed using strategies discussed before—for instance, using the penalty variable as an instrument. The endogeneity concern related to the proportion of boys, $\check{B}_{ic}$, can be addressed using the first-son dummy as an instrument. A new concern is that fertility, f_{ic}, is also correlated with son-preference parameters α_c and α_{ic}, which enter the error term. To deal with this concern, we use the number of brothers the father of the household has, as an additional instrument for $\check{B}_{ic}$ and f_{ic} (Fan et al., 2018). If a father has more brothers, he faces less social pressure to have a son, since the sons of his brothers can continue the family name, resulting in this father having lower fertility and fewer sons (Appendix Table A7). In addition, the number of brothers the father has does not directly affect family investment decisions, since married brothers usually have independent financial pictures, and do not bear financial responsibility for one another.[25]

It is worth noting that one of the key assumptions that has underpinned our analysis so far is not required in the current specification: We do not need to assume that parents do not take into account marriage market conditions while making fertility decisions, since none of the instruments—the penalty for supplying sex-selection services, the first-son dummy, and the number of brothers the father has—are affected by the marriage prospects of children. OLS results conditioning on fertility are reported in panel C, Tables 7 and 8, and IV results are in panel D; they give broadly consistent patterns as before.

As a final remark, our instrument for the county-level sex ratio—the penalty for supplying sex-selection services—does not directly affect household fertility decisions, while the standard instrument in the literature—the penalty for requiring such services—may do so. To the extent that there might be residual concerns, it is reassuring that our results are robust to conditioning on fertility.[26]

[25] Another candidate for the additional instrument can be the standard one in the literature: the intensity of the one-child policy. Results remain similar if we use this instrument instead.

[26] It is also reassuring that both demand- and supply-side instruments give similar results.

5.4 Other hypotheses

We have allayed in Section 5.2 the concern that unobserved son preference is correlated with the sex ratio and also affects investments depending on child gender, in our interpretation of variation in the county sex ratio as capturing parents' marriage market consideration for children. Beyond son preference, there may be other forms of cross-county heterogeneity, that are associated with the local sex ratio and also affect parental investments in children. But since we compare first-son families with first-daughter families within a county, these confounding heterogeneities would not affect the interpretation of results as long as they affect investments in boys and girls uniformly. But if certain factors are correlated with the sex ratio and affect investments in boys and girls differently, they constitute competing hypotheses for our empirical patterns.

For instance, household structures may differ between high- and balanced-sex-ratio regions, and affect parental investment in a child-gender-specific way. In high-sex-ratio regions, a first-born girl is more likely to have a sibling relative to a first-born boy, while in balanced-sex-ratio regions, the first child, regardless of the gender, may be the only child. Parents with more children may devote less time to the labor market—in particular, are less likely to migrate; they also have less residual wealth to invest in real estate, as more children dilute household resources. Although in this respect, the hypothesis is consistent in part with our empirical findings, we have verified in Section 5.3 that our results are not driven by endogenous fertility. Also, results on educational investment show that when sex ratios are high, parents invest less on a first boy relative to a first girl, which contradicts this hypothesis as it suggests that first-son families have more per child resources available.

Household structures may also differ in term of whether grandparents are living together. Child gender may affect the likelihood of living with grandparents differently across high- and balanced-sex-ratio counties, and living with grandparents may influence household investment decisions. To reduce this concern, we control for the interaction between the proportion of households living with grandparents at the county level, and the first-son dummy. As in panel A, Appendix Tables A10 and A11, our results are not driven by the pattern of living with grandparents.

Moreover, high- and balanced-sex-ratio regions may differ in development levels and the spread of technologies. In particular, sex-selection technology is highly correlated with the local sex ratio (Zhang, 2017), and may have different influences across first-son and first-daughter families. We use county-level average household financial wealth (defined as the sum of liquid and illiquid assets) and average household income as proxies for the levels of development and technology. We add in the regressions interactions between these proxies and first son. As in panels B and C, Appendix Tables A10 and A11, considering these confounding factors does not change the results.

Next we consider status seeking as a potential explanation. Perhaps, status competition among son families is more intense than among daughter families in counties with high sex ratios. Indeed, Brown et al. (2011) find that grooms' families spend more on weddings as local competition for status intensifies. This hypothesis is in part compatible with our findings on parental investments, since a stronger desire to compete for status induces families with a son to engage in earning activities and housing investment more aggressively. This hypothesis speaks to overall investments and housing investment, yet could not provide an explanation for parents who face higher sex ratios investing relatively less in sons' education. In comparison, our proposed mechanisms have finer implications by distinguishing the types of investments, and can fully explain the empirical patterns, for both dimensions of investments.

One may also challenge the interpretation of housing as a form of parental investments for children. We have provided supportive evidence in Section 2 that parents see their property as investments in preparation for their sons' marriage, at the time when their sons are young. But alternatively, real estate investment might reflect the desire to get higher returns, and the returns may be correlated with the local sex ratio. It is difficult, however, to believe that any shocks that relate to housing investment returns apply to local families in a child-gender-dependent manner. Therefore, comparing investments between first-son and first-daughter families differences out the effect of such shocks. It is also possible that wealthier parents—those with a first son in high-sex-ratio regions, who are more likely to migrate and have fewer children—just happen to keep their wealth in the form of housing. But we have controlled for parental education levels in main estimations, and further added average household wealth and income in robustness checks, to account for any wealth effect. Thus, parents with sons attach greater importance to housing investment when the sex ratio gets higher, mainly because of marriage market considerations.

To sum up, while some other stories seem to rationalize part of our findings, a plausible interpretation is the theory by Bhaskar et al. (2023), that bequeathable housing of men is more attractive to potential brides than education, as spouses are unable to commit to an agreement regarding the after-marriage division of labor earnings.

6 Further analyses

6.1 Heterogeneity across rural and urban

We present the heterogenous effects across rural and urban households. We divide the sample into a rural and an urban subsample, and perform the same regression analysis using the two subsamples. Appendix Tables A12 and A13 show results on parental labor supply and parental investments, respectively, using the OLS method. Panel A is for rural families, panel B is for urban families, and panel C tests the difference in the effects between them.

For parental migration, columns 1–3 of Appendix Table A12 show that the effect is more prominent among rural households; the effect also exists among urban households, although appears to be smaller and less significant. Based on p-values from column-wise Hausman's general specification tests, the hypothesis of no statistically significant difference between rural and urban estimates for father's and at least one parent's migration cannot be rejected at conventional significance levels; for mother's migration, the hypothesis can be rejected at the five percent level.

For housing investment, columns 1–3 of Appendix Table A13 show that the effect on construction area and ownership is larger and more significant among rural households; the effect on mortgage is larger among urban households, although not statistically significant. Yet, the hypothesis of no significant difference between rural and urban estimates for housing investment cannot be rejected at the ten percent level or below. This suggests that despite any possible difference in housing investment modes or housing market structures across rural and urban areas, parents with sons in both areas invest more in housing than parents with daughters when the sex ratio is high.

For most other outcomes of interest—parental working hours (columns 4 and 5 of Appendix Table A12) and child educational investment (columns 4 and 5 of Appendix Table A13), similarly, the hypothesis of no statistically significant difference between rural and urban estimates cannot be rejected at conventional levels. But in column 6 of Appendix Table A13, we observe a stronger reducing effect on educational investment relative to housing value for rural families with sons than comparable urban families, and the effect difference is statistically significant at the five percent level.

6.2 Implications for child human capital development

What are the implications for children's human capital development of the parental labor supply and investment patterns that are driven by marriage market considerations for children? Educational investment has a direct effect on child development. Parental migration also has important implications, since the adverse effect of the absence of parents is well documented, both in China and in other countries (e.g., Zhang et al., 2014; Lyle, 2006).

To study these effects, we measure the human capital development in terms of cognitive, non-cognitive, and health outcomes of the first-born (cognitive and non-cognitive outcomes are for those above ten years old). Cognitive outcomes are measured by a child's ranking in the class' cumulative distribution function of mathematics and Chinese examination scores (closer to one means closer to the top and thus higher cognitive abilities). Non-cognitive outcomes are measured by indicator variables for the exhibition of interpersonal communication skills including reception and comprehension (a value of one implies a better performance).[27] Health outcomes are weight and height measured by z-scores transformed based on international child growth standards (a larger value implies a more satisfactory result).

We use the same empirical strategies as before and report results in Table 9. In the first four columns, the negative estimates for the interaction-term coefficient indicate that a high sex ratio adversely impacts both academic achievement and non-cognitive skills of boys relative to girls, and the adverse effects are significant economically and statistically in most cases. Columns 5 and 6 show adverse effects on health outcomes of children, although these are not always significant in the statistical sense. These results indicate that increased parental migration and reduced educational investment in families with a boy, as a result of sex ratio imbalance, may undermine the human capital development of the boy.

7 Conclusion

Our paper is among the few that consider multidimensional investments, by empirically studying investments made by parents in a marriage market where boys face severe competition due to sex ratio imbalances, but enjoy greater bargaining power after marriage. Using nationally representative Chinese data, our key finding is that the different components of investments may respond differently to marriage market competition. Male-biased sex ratios lead to increased parental migration, increased housing investment, and reduced educational investment for families with sons relative to families with daughters. Parents of boys direct their investments towards housing in order to credibly commit to share more resources with the boy's potential wife; they do not invest more in education for the boy, since he is unable to commit to future resource sharing from his labor income. We connect the empirical patterns to the proposed mechanisms by providing a variety of robustness analyses to

[27] After interviewing each child, interviewers were asked to evaluate interpersonal communication skills of the child, which were recorded on a 1–7 scale with a larger number meaning a better performance. We recode the record into a dummy, which is 1 if the record was above 5.

address the concerns related unobserved son preference and endogenous fertility. We also show evidence that parental absences due to migration and underinvestment in education may combine to result in adverse child development consequences.

References

Alesina, Alberto, Paola Giuliano, and Nathan Nunn, "Traditional agricultural practices and the sex ratio today," *PLOS One*, 2018, *13* (1), e0190510.

Ashraf, Nava, "Spousal control and intra-household decision making: An experimental study in the Philippines," *American Economic Review*, 2009, *99* (4), 1245–77.

Becker, Gary S., "A Theory of Marriage: Part I," *Journal of Political Economy*, 1973, pp. 813–846.

—, *A Treatise on the Family*, Harvard University Press, 1991.

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

Bhaskar, V., "Sex Selection and Gender Balance," *American Economic Journal: Microeconomics*, 2011, *3* (1), 214–244.

— **and Ed. Hopkins**, "Marriage as a Rat Race: Noisy Premarital Investments with Assortative Matching," *Journal of Political Economy*, 2016, *124* (4), 992–1045.

—**, Wenchao Li, and Junjian Yi**, "Multidimensional Premarital Investments with Imperfect Commitment," *Journal of Political Economy (forthcoming)*, 2023.

Brown, Philip H, Erwin Bulte, and Xiaobo Zhang, "Positional spending and status seeking in rural China," *Journal of Development Economics*, 2011, *96* (1), 139–149.

Browning, Martin, Pierre-André Chiappori, and Yoram Weiss, *Economics of the Family*, Cambridge University Press, 2014.

Cameron, Lisa, Xin Meng, and Dandan Zhang, "China's sex ratio and crime: Behavioural change or financial necessity?," *Economic Journal*, 2019, *129* (618), 790–820.

Castilla, Carolina, "Trust and reciprocity between spouses in India," *American Economic Review*, 2015, *105* (5), 621–24.

— **and Thomas Walker**, "Is ignorance bliss? The effect of asymmetric information between spouses on intra-household allocations," *American Economic Review*, 2013, *103* (3), 263–68.

Chai, Kangni, "How Marriage Affects Chinese Cities," *PEAKUrban*, 2020. June 16. https://www.peak-urban.org/blog/how-marriage-affects-chinese-cities.

Chen, Yuyu, Hongbin Li, and Lingsheng Meng, "Prenatal sex selection and missing girls in China: Evidence from the diffusion of diagnostic ultrasound," *Journal of Human Resources*, 2013, *48* (1), 36–70.

Chiappori, Pierre-Andre and Maurizio Mazzocco, "Static and intertemporal household decisions," *Journal of Economic Literature*, 2017, *55* (3), 985–1045.

Chiappori, Pierre-André, Bernard Fortin, and Guy Lacroix, "Marriage Market, Divorce Legislation, and Household Labor Supply," *Journal of Political Economy*, 2002, *110* (1), 37–72.

Chu, Junhong, "Prenatal sex determination and sex-selective abortion in rural central China," *Population and Development Review*, 2001, *27* (2), 259–281.

Civil Code of China: Part V, "Marriage and Family," *Article 1085, 1087*, 2020. https://www.chinajusticeobserver.com/a/is-alimony-always-granted.

Cunha, Flavio and James Heckman, "The Technology of Skill Formation," *American Economic Review*, 2007, *97* (2), 31–47.

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

Doepke, Matthias and Fabian Kindermann, "Bargaining over babies: Theory, evidence, and policy implications," *American Economic Review*, 2019, *109* (9), 3264–3306.

— **and Michele Tertilt**, "Asymmetric information in couples," *Unpublished Mimeo*, 2016.

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

— , "Estimating a dynamic model of sex selection in China," *Demography*, 2011, *48* (2), 783–811.

— **, Hongbin Li, and Lingsheng Meng**, "The impact of ultrasound technology on the status of women in China," *Working Paper*, 2013.

Edlund, Lena, "Son preference, sex ratios, and marriage patterns," *Journal of Political Economy*, 1999, *107* (6), 1275–1304.

— **, Hongbin Li, Junjian Yi, and Junsen Zhang**, "Sex ratios and crime: Evidence from China," *Review of Economics and Statistics*, 2013, *95* (5), 1520–1534.

Fan, Yi, Junjian Yi, Ye Yuan, and Junsen Zhang, "The glorified mothers of sons: Evidence from child sex composition and parental time allocation in rural China," *Journal of Economic Behavior & Organization*, 2018, *145*, 249–260.

Gobbi, Paula E, "Childcare and commitment within households," *Journal of Economic Theory*, 2018, *176*, 503–551.

Heckman, James, "China's investment in human capital," *Economic Development and Cultural Change*, 2003, *51* (4), 795–804.

— , "China's human capital investment," *China Economic Review*, 2005, *16* (1), 50–70.

— , "Skill formation and the economics of investing in disadvantaged children," *Science*, 2006, *312* (5782), 1900–1902.

— , "The Economics, Technology, and Neuroscience of Human Capability Formation," *Proceedings of the National Academy of Sciences*, 2007, *104* (33), 13250–13255.

— **, Rodrigo Pinto, and Peter Savelyev**, "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes," *American Economic Review*, 2013, *103* (6), 2052–2086.

Hoppe, Heidrun C, Benny Moldovanu, and Aner Sela, "The theory of assortative matching based on costly signals," *Review of Economic Studies*, 2009, *76* (1), 253–281.

Laat, Joost De, "Household allocations and endogenous information: The case of split migrants in Kenya," *Journal of Development Economics*, 2014, *106*, 108–117.

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

Li, Wenchao, "The "miseries" of sex imbalance: Evidence using subjective well-being data," *Journal of Development Economics*, 2021, *151*, 102634.

— **, Changcheng Song, Shu Xu, and Junjian Yi**, "High sex ratios and household portfolio choice in China," *Journal of Human Resources*, 2022, *57* (2), 465–490.

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

Liu, Yang, "70% of married men have private money: 15 best places to hide private money according to anecdotal records (in Chinese)," *Yangzi Evening Paper*, 2013. September 3. http://scitech.people.com.cn/n/2013/0903/c1057-22784523.html.

Lundberg, Shelly and Robert A. Pollak, "Separate Spheres Bargaining and the Marriage Market," *Journal of Political Economy*, 1993, pp. 988–1010.

Lyle, David S., "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*, 2006, *24* (2), 319–350.

Ma, Yihan, "Necessity of home ownership before marriage in China 2019, by age group," *Statista*, 2020. May 7. https://www.statista.com/statistics/1102317/china-necessity-of-property-ownership-before-marriage/.

Mazzocco, Maurizio, "Household Intertemporal Behaviour: A Collective Characterization and A Test of Commitment," *Review of Economic Studies*, 2007, *74* (3), 857–895.

Pierson, David, "China's Housing Boom Spells Trouble for Boyfriends," *Los Angeles Times*, 2010.

Rao, Vijayendra, "The Rising Price of Husbands: A Hedonic Analysis of Dowry Increases in Rural India," *Journal of Political Economy*, 1993, pp. 666–677.

Rosenzweig, Mark and Junsen Zhang, "Co-residence, life-cycle savings and inter-generational support in urban China," *NBER Working Paper*, 2014.

Schaner, Simone, "Do opposites detract? Intrahousehold preference heterogeneity and inefficient strategic savings," *American Economic Journal: Applied Economics*, 2015, *7* (2), 135–74.

Scharping, Thomas, *Birth Control in China 1949-2000: Population Policy and Demographic Development*, Routledge, 2013.

Voena, Alessandra, "Yours, Mine, and Ours: Do Divorce Laws Affect the Intertemporal Behavior of Married Couples?," *American Economic Review*, 2015, *105* (8), 2295–2332.

Wei, Shang-Jin and Xiaobo Zhang, "The competitive saving motive: Evidence from rising sex ratios and savings rates in China," *Journal of Political Economy*, 2011, *119* (3), 511–564.

Wrenn, Douglas H, Junjian Yi, and Bo Zhang, "House prices and marriage entry in China," *Regional Science and Urban Economics*, 2019, *74*, 118–130.

Xie, Yu, *The User's Guide of the China Family Panel Studies (2010)*, Beijing: Institute of Social Science Survey, Peking University, 2012.

_ **and Yongai Jin**, "Household Wealth in China," *Chinese Sociological Review*, 2015, *47* (3), 203–229.

Xing, Zhaoguo, "The evolution of Chinese rural families, centralization or individualization? Moral evaluation of private money as an assess point (in Chinese)," *Society*, 2017, *37* (5), 165–192.

_ , "Marriage stability, family economic control, and private money: An anthropological study of family hidden property (in Chinese)," *Journal of Guangxi University for Nationalities (Philosophy and Social Science Edition)*, 2017, *39* (2), 99–110.

Zeng, Yi, Ping Tu, Baochang Gu, Yi Xu, Bohua Li, and Yongping Li, "Causes and implications of the recent increase in the reported sex ratio at birth in China," *Population and Development Review*, 1993, pp. 283–302.

Zhang, Hongliang, Jere R. Behrman, C. Simon Fan, Xiangdong Wei, and Junsen Zhang, "Does Parental Absence Reduce Cognitive Achievements? Evidence from Rural China," *Journal of Development Economics*, 2014, *111*, 181–195.

Zhang, Junsen, "The evolution of China's one-child policy and its effects on family outcomes," *Journal of Economic Perspectives*, 2017, *31* (1), 141–60.

_ **and William Chan**, "Dowry and Wife's Welfare: A Theotrical and Empirical Analysis," *Journal of Political Economy*, 1999, *107* (4), 786–808.

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

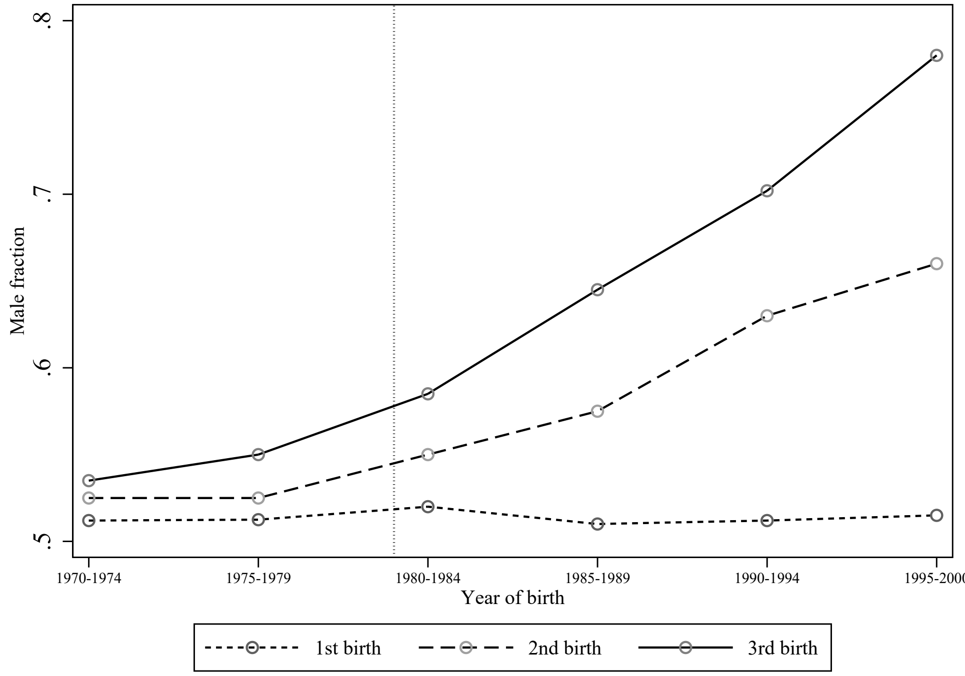

Figure 1 Male fraction of births by birth order in China

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.

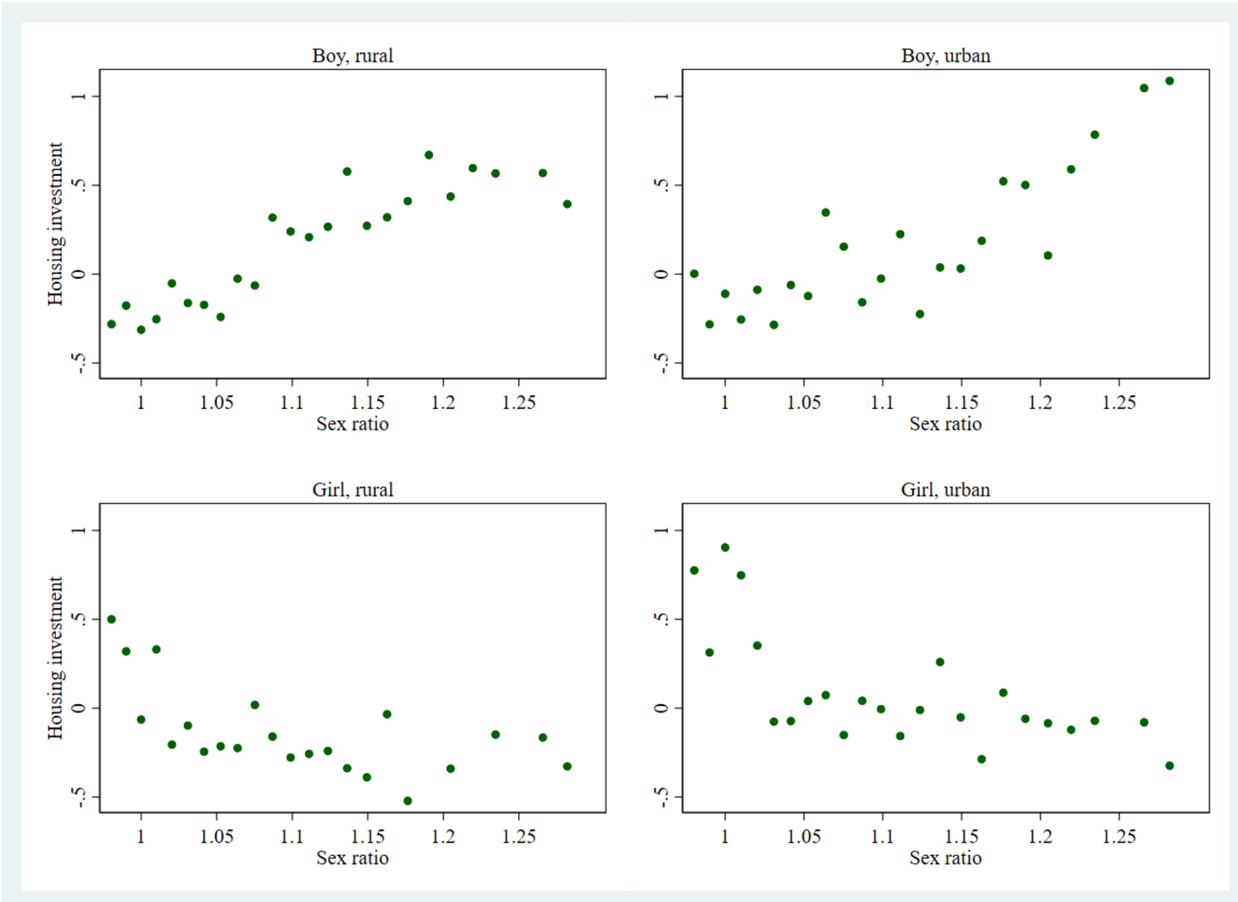

Figure 2 Housing investment and sex ratio by child gender and place of residence

Notes: This figure depicts housing investment made by parents with a first-born boy and parents with a first-born girl, against the county-level sex ratio (the ratio of boys to girls), in rural and urban China. Housing investment is the ratio of housing market value to household income; the variable is first standardized (by subtracting the mean and dividing by the standard deviation) and then averaged over counties within the 0.01 intervals of sex ratios.

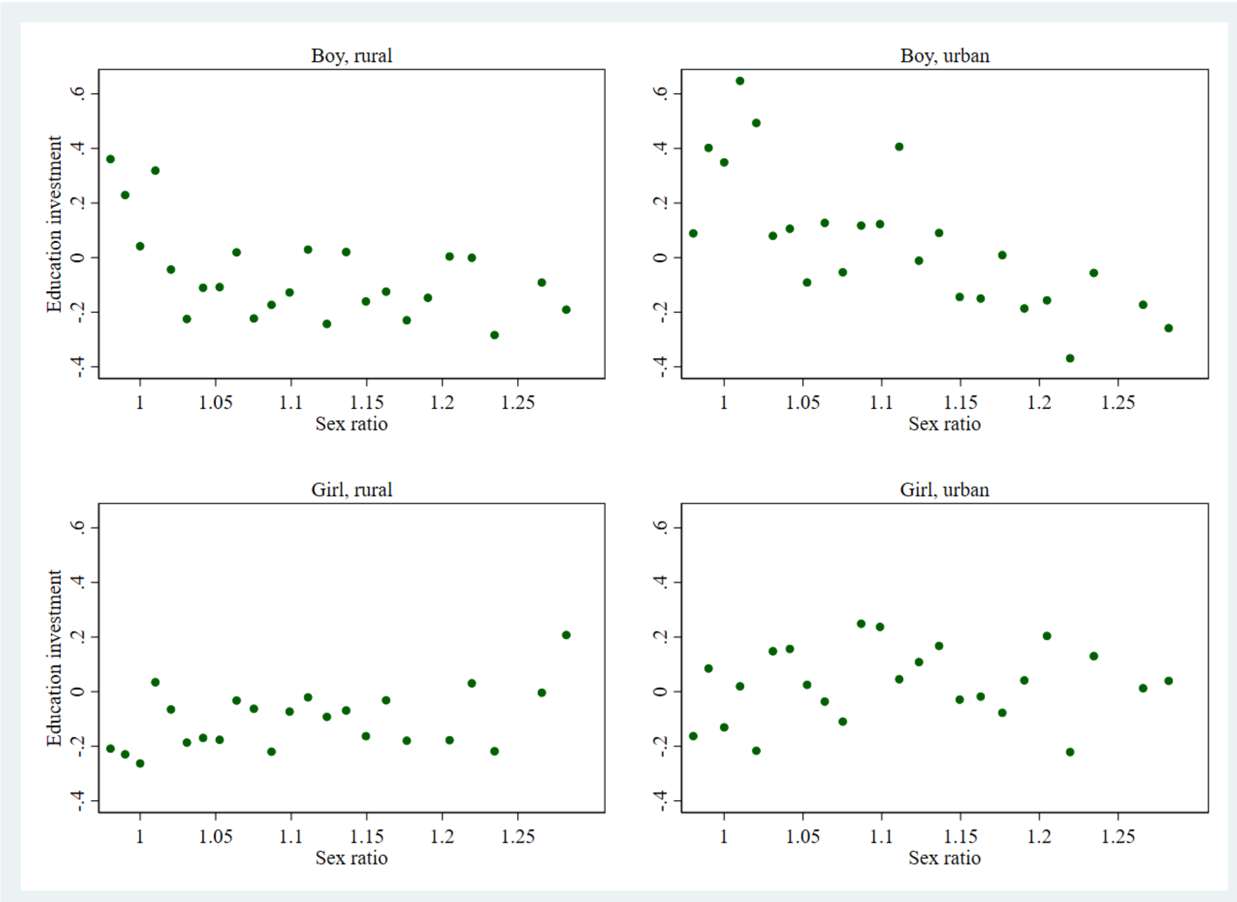

Figure 3 Education investment and sex ratio by child gender and place of residence

Notes: This figure depicts educational investment made by parents with a first-born boy and parents with a first-born girl, against the county-level sex ratio (the ratio of boys to girls), in rural and urban China. Educational investment is the ratio of annual educational expenditure per child to household income; the variable is first standardized (by subtracting the mean and dividing by the standard deviation) and then averaged over counties within the 0.01 intervals of sex ratios.

Table 1 Summary statistics of main outcome variables

	Mean	Std. dev.	Min.	Max.	Obs.
Parental labor supply					
Migration: father	0.098	0.298	0	1	4,304
Migration: mother	0.025	0.157	0	1	4,304
Migration: at least one parent	0.111	0.314	0	1	4,304
Working hours, yearly: father	2,466	947.1	148	5,400	1,527
Working hours, yearly: mother	2,416	901.8	240	5,400	972
Housing investment					
Housing construction area, sq.m	126.2	86.35	8	1,000	4,159
Housing ownership	0.832	0.374	0	1	4,304
Housing mortgage, thousand	5.402	33.98	0	750	4,304
Child educational investment					
Education expenditure, thousand	1.508	2.624	0	40	3,969
Having an education fund	0.298	0.457	0	1	3,969
Education expenditure/housing value	0.040	0.153	0	2.800	3,139

Notes: Data are from the CFPS baseline survey. The sample includes families in which the eldest child was under the age of 15, both parents were alive and at most 50 years old, and at least one parent participated in the adult survey. Child educational investment is measured for the first-born older than two. "Education expenditure/housing value" has a smaller number of observations because of missing housing value. Statistics are weighted by the survey sampling weights.

Table 2 Sex ratio, penalties for healthcare workers, and pre-ultrasonography fertility

Dependent variable	2010 Sex ratio			1990s Real penalty	
	(1)	(2)	(3)	(4)	(5)
1990s Real penalty	-0.182***	-0.175***	-0.150***		
	(0.033)	(0.032)	(0.028)		
	[-5.532]	[-5.450]	[-5.345]		
1981 Completed fertility				-0.026	
				(0.076)	
				[-0.341]	
1981 Birth rate					-0.001
					(0.002)
					[-0.883]
# Counties	155	155	155	151	151
R-squared	0.401	0.486	0.631	0.216	0.218
Other controls		Yes	Yes	Yes	Yes
Province FE			Yes	Yes	Yes

Notes: In this table, all variables are at the county level. "Real penalty" is defined as the average penalty for healthcare workers who illegally practiced gender screening and gender selection, relative to average household income. Columns 1–3 are the regression results of sex ratio in 2010 on real penalty in the 1990s; column 1 includes only penalty; column 2 further adds other county-level controls, including average age, average schooling years, industrial structure, and average wage; column 3 further adds province fixed effects. Columns 4 and 5 are the regression results of real penalty in the 1990s on fertility measures in 1981; "Completed fertility" is the average fertility of women aged 35–64; "Birth rate" is the number of births during the year over total population (with a multiplier 1000); other county-level controls include population, average age, the proportion with a junior high school degree, the proportion of ethnic minorities, and industrial structure. Standard errors given in parentheses are heteroskedasticity-robust. In square brackets are t-statistics.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 3 Parental labor supply

Dependent variable	Migration			Working hours, log	
	Father	Mother	At least one parent	Father	Mother
	(1)	(2)	(3)	(4)	(5)
A: OLS results					
First son * Sex ratio	0.247**	0.114*	0.286***	0.575***	0.434
	(0.104)	(0.064)	(0.102)	(0.187)	(0.383)
Observations	4,304	4,304	4,304	1,527	972
R-squared	0.112	0.065	0.119	0.176	0.270
Percentage difference sex ratio+1sd	25.1	45.4	25.8	5.8	4.3
B: Controlling for pre-ultrasonography fertility					
First son * Sex ratio	0.265**	0.107	0.291***	0.547***	0.122
	(0.111)	(0.067)	(0.111)	(0.205)	(0.343)
Observations	4,209	4,209	4,209	1,477	946
R-squared	0.115	0.067	0.121	0.184	0.294
First son * Pre-ult. fertility	Yes	Yes	Yes	Yes	Yes
C: IV results					
First son * Sex ratio	0.320**	0.183**	0.408***	0.639**	0.457
	(0.135)	(0.077)	(0.148)	(0.279)	(0.474)
Observations	4,304	4,304	4,304	1,527	972
R-squared	0.112	0.065	0.118	0.176	0.270
First-stage F-statistic	48.6	48.6	48.6	53.9	82.4
Dependent variable mean	0.098	0.025	0.111	7.725	7.701
Baseline controls	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes

Notes: In panel A, the difference in the effect of sex ratios between first-son and first-daughter families is reported in percentages in the middle. In panel B, regressions additionally control for interactions of “1981 Completed fertility” and “1981 Birth rate” as defined in notes to Table 2, with the first-son dummy (observations are fewer as four out of 155 counties have missing fertility variables). In panel C, the potentially endogenous variable is “First son * Sex ratio” and the instrument is “First son * Real penalty,” where “Real penalty” is defined in notes to Table 2. Estimations are weighted by the 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 Multidimensional parental investments

Dependent variable	Housing investment			Child edu. investment		Edu. exp.
	Constr. area, log sq.m	Ownership	Mortgage, thousand	Edu. exp., thousand	Having an edu. fund	/Housing value
	(1)	(2)	(3)	(4)	(5)	(6)
A: OLS results						
First son * Sex ratio	0.478**	0.269**	18.72*	-1.775**	-0.382**	-0.187**
	(0.198)	(0.113)	(10.43)	(0.793)	(0.167)	(0.093)
Observations	4,159	4,304	4,304	3,969	3,969	3,139
R-squared	0.276	0.182	0.182	0.317	0.149	0.119
Percentage difference sex ratio+1sd	4.8	3.2	34.7	-11.8	-12.8	-47.2
B: Controlling for pre-ultrasonography fertility						
First son * Sex ratio	0.452**	0.242*	22.03**	-1.256*	-0.413**	-0.197**
	(0.199)	(0.124)	(10.47)	(0.677)	(0.175)	(0.096)
Observations	4,066	4,209	4,209	3,884	3,884	3,075
R-squared	0.284	0.179	0.187	0.317	0.147	0.119
First son * Pre-ult. fert.	Yes	Yes	Yes	Yes	Yes	Yes
C: IV results						
First son * Sex ratio	0.590**	0.280*	25.49*	-1.845*	-0.743***	-0.258***
	(0.274)	(0.152)	(14.31)	(1.104)	(0.229)	(0.099)
Observations	4,159	4,304	4,304	3,969	3,969	3,139
R-squared	0.276	0.182	0.182	0.317	0.148	0.119
First-stage F-statistic	51.3	48.6	48.6	50.0	50.0	47.1
Dependent variable mean	4.651	0.832	5.402	1.508	0.298	0.040
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: In panel A, the difference in the effect of sex ratios between first-son and first-daughter families is reported in percentages in the middle. In panel B, regressions additionally control for interactions of "1981 Completed fertility" and "1981 Birth rate" as defined in notes to Table 2, with the first-son dummy (observations are fewer as four out of 155 counties have missing fertility variables). In panel C, the potentially endogenous variable is "First son * Sex ratio" and the instrument is "First son * Real penalty," where "Real penalty" is defined in notes to Table 2. Estimations are weighted by the 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 5 One-child families: Parental labor supply

Dependent variable	Migration			Working hours, log	
	Father	Mother	At least one parent	Father	Mother
	(1)	(2)	(3)	(4)	(5)
A: OLS results					
First son * Sex ratio	0.237**	0.155**	0.298***	0.480**	0.429
	(0.104)	(0.076)	(0.106)	(0.221)	(0.491)
Observations	2,851	2,851	2,851	1,166	828
R-squared	0.148	0.090	0.165	0.210	0.284
B: Controlling for pre-ultrasonography fertility					
First son * Sex ratio	0.289***	0.147*	0.333***	0.422*	0.007
	(0.110)	(0.079)	(0.117)	(0.254)	(0.493)
Observations	2,784	2,784	2,784	1,128	806
R-squared	0.154	0.092	0.169	0.223	0.312
First son * Pre-ult. fertility	Yes	Yes	Yes	Yes	Yes
C: IV results					
First son * Sex ratio	0.341**	0.255**	0.473***	0.725**	0.597
	(0.157)	(0.104)	(0.179)	(0.300)	(0.566)
Observations	2,851	2,851	2,851	1,166	828
R-squared	0.148	0.089	0.164	0.209	0.284
Dependent variable mean	0.088	0.023	0.101	7.740	7.695
Baseline controls	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes

Notes: The sample is restricted to one-child families. Estimations are weighted by the 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 6 One-child families: Multidimensional parental investments

Dependent variable	Housing investment			Child edu. investment		Edu. exp.
	Constr. area, log sq.m	Ownership	Mortgage, thousand	Edu. exp., thousand	Having an edu. fund	/Housing value
	(1)	(2)	(3)	(4)	(5)	(6)
A: OLS results						
First son * Sex ratio	0.501*	0.212*	20.918*	-1.765**	-0.349	-0.177**
	(0.268)	(0.120)	(11.645)	(0.876)	(0.245)	(0.088)
Observations	2,762	2,851	2,851	2,521	2,521	2,035
R-squared	0.332	0.196	0.200	0.345	0.182	0.148
B: Controlling for pre-ultrasonography fertility						
First son * Sex ratio	0.433*	0.165	24.565**	-1.834*	-0.422*	-0.194**
	(0.262)	(0.123)	(12.089)	(0.936)	(0.253)	(0.095)
Observations	2,697	2,784	2,784	2,464	2,464	1,993
R-squared	0.343	0.194	0.201	0.344	0.179	0.150
First son * Pre-ult. fert.	Yes	Yes	Yes	Yes	Yes	Yes
C: IV results						
First son * Sex ratio	0.543	0.346**	29.920*	-1.908	-0.675**	-0.241**
	(0.352)	(0.175)	(18.184)	(1.302)	(0.322)	(0.105)
Observations	2,762	2,851	2,851	2,521	2,521	2,035
R-squared	0.332	0.196	0.200	0.345	0.180	0.147
Dependent variable mean	4.638	0.834	6.765	1.795	0.347	0.033
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The sample is restricted to one-child families. Estimations are weighted by the 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 7 Addressing endogenous fertility: Parental labor supply

Dependent variable	Migration			Working hours, log	
	Father	Mother	At least one parent	Father	Mother
	(1)	(2)	(3)	(4)	(5)
A: OLS results					
Prop. sons * Sex ratio	0.310***	0.182**	0.365***	0.499**	0.443
	(0.099)	(0.077)	(0.099)	(0.224)	(0.429)
Observations	4,304	4,304	4,304	1,527	972
R-squared	0.113	0.068	0.121	0.174	0.269
B: IV results					
Prop. sons * Sex ratio	0.402**	0.233**	0.514***	0.758**	0.607
	(0.173)	(0.099)	(0.190)	(0.330)	(0.551)
Observations	4,304	4,304	4,304	1,527	972
R-squared	0.112	0.067	0.120	0.172	0.260
C: OLS results, conditioning on fertility					
Prop. sons * Sex ratio	0.313***	0.183**	0.366***	0.489**	0.443
	(0.099)	(0.077)	(0.099)	(0.228)	(0.429)
Observations	4,304	4,304	4,304	1,527	972
R-squared	0.114	0.068	0.121	0.178	0.269
D: IV results, conditioning on fertility					
Prop. sons * Sex ratio	0.357*	0.228**	0.478**	0.695**	0.351
	(0.204)	(0.108)	(0.213)	(0.323)	(0.731)
Observations	3,772	3,772	3,772	1,304	818
R-squared	0.054	0.068	0.072	0.191	0.053
Baseline controls	Yes	Yes	Yes	Yes	Yes
Prop. sons * Controls	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes

Notes: Panels A and B are based on equation (16). Panels C and D are based on equation (17). Estimations are weighted by the 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 8 Addressing endogenous fertility: Multidimensional parental investments

Dependent variable	Housing investment			Child edu. investment		Edu. exp.
	Constr. area, log sq.m	Ownership	Mortgage, thousand	Edu. exp., thousand	Having an edu. fund	/Housing value
	(1)	(2)	(3)	(4)	(5)	(6)
A: OLS results						
Prop. sons * Sex ratio	0.566***	0.284**	17.946	-0.824	-0.349*	-0.312*
	(0.211)	(0.121)	(11.282)	(0.990)	(0.188)	(0.162)
Observations	4,159	4,304	4,304	3,969	3,969	3,139
R-squared	0.275	0.182	0.183	0.246	0.150	0.110
B: IV results						
Prop. sons * Sex ratio	0.715**	0.347*	32.418*	-3.681*	-0.951***	-0.533**
	(0.328)	(0.191)	(17.534)	(1.922)	(0.294)	(0.207)
Observations	4,159	4,304	4,304	3,969	3,969	3,139
R-squared	0.274	0.180	0.183	0.243	0.146	0.108
C: OLS results, conditioning on fertility						
Prop. sons * Sex ratio	0.562***	0.278**	17.71	-0.586	-0.366*	-0.295**
	(0.212)	(0.118)	(11.31)	(0.934)	(0.186)	(0.149)
Observations	4,159	4,304	4,304	3,969	3,969	3,139
R-squared	0.275	0.183	0.184	0.266	0.155	0.126
D: IV results, conditioning on fertility						
Prop. sons * Sex ratio	0.694**	0.440**	16.41	-2.686	-0.981***	-0.429*
	(0.323)	(0.216)	(20.80)	(2.102)	(0.294)	(0.228)
Observations	3,637	3,772	3,772	3,505	3,505	2,763
R-squared	0.268	0.167	0.130	0.264	0.146	0.118
Baseline controls	Yes	Yes	Yes	Yes	Yes	
Prop. sons * Controls	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Panels A and B are based on equation (16). Panels C and D are based on equation (17). Estimations are weighted by the 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 9 Child human capital development

Dependent variable	Cognitive		Non-cognitive		Health	
	Math ranking	Chinese ranking	Openness	Cooperation	Weight, z-score	Height, z-score
	(1)	(2)	(3)	(4)	(5)	(6)
A: OLS results						
First son * Sex ratio	-0.653***	-0.567***	-0.423**	-0.536**	-0.760*	-0.222
	(0.208)	(0.215)	(0.204)	(0.259)	(0.455)	(0.486)
Observations	1,147	1,147	2,117	2,117	4,127	3,860
R-squared	0.238	0.267	0.248	0.291	0.219	0.210
Percentage difference sex ratio+1sd	-9.4	-8.1	-4.9	-7.4		
B: Controlling for pre-ultrasonography fertility						
First son * Sex ratio	-0.627***	-0.520**	-0.447**	-0.577**	-0.857*	0.175
	(0.215)	(0.229)	(0.209)	(0.268)	(0.458)	(0.423)
Observations	1,122	1,122	2,068	2,068	4,034	3,773
R-squared	0.222	0.250	0.245	0.296	0.221	0.219
First son * Pre-ult. fert.	Yes	Yes	Yes	Yes	Yes	Yes
C: IV results						
First son * Sex ratio	-0.474*	-0.493*	-0.377	-0.084	-0.864	-0.588
	(0.267)	(0.279)	(0.259)	(0.397)	(0.627)	(0.875)
Observations	1,147	1,147	2,117	2,117	4,127	3,860
R-squared	0.237	0.267	0.248	0.289	0.219	0.209
First-stage F-statistic	46.6	46.6	42.7	42.7	49.4	48.9
Dependent variable mean	0.692	0.701	0.859	0.728	-0.502	-0.638
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Human capital outcomes are measured for the first-born, and cognitive and non-cognitive outcomes are for those who are at least ten years old. In panel A, the difference in the effect of sex ratios between first-son and first-daughter families is reported in percentages in the middle. In panel B, regressions additionally control for interactions of "1981 Completed fertility" and "1981 Birth rate" as defined in notes to Table 2, with the first-son dummy (observations are fewer as four out of 155 counties have missing fertility variables). In panel C, the potentially endogenous variable is "First son * Sex ratio" and the instrument is "First son * Real penalty," where "Real penalty" is defined in notes to Table 2. Estimations are weighted by the 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.

Online Appendices

A Additional tables

Table A1 Balance test: First-son versus first-daughter families

	Mean (Std. dev.)				
	All	First-son families	First-daughter families	Difference (2)-(3)	P-value
	(1)	(2)	(3)	(4)	(5)
A: Summary statistics					
First son	0.511				
	(0.500)				
County-level sex ratio (m/f)	1.079	1.079	1.078	0.001	0.783
	(0.103)	(0.102)	(0.104)		
County-level penalty ratio	1.009	1.010	1.008	0.002	0.864
	(0.348)	(0.344)	(0.352)		
Region of residence (urban=1)	0.465	0.473	0.456	0.017	0.313
	(0.499)	(0.499)	(0.498)		
First-child age	8.692	8.583	8.805	-0.222	0.198
	(4.490)	(4.508)	(4.469)		
Ethnicity (minority=1)	0.159	0.160	0.159	0.001	0.906
	(0.366)	(0.367)	(0.365)		
Father's age	36.10	36.01	36.20	-0.190	0.335
	(5.987)	(5.921)	(6.055)		
Father's schooling years	7.954	8.025	7.879	0.146	0.365
	(4.244)	(4.237)	(4.250)		
Father's political status (CCP=1)	0.096	0.094	0.098	-0.004	0.765
	(0.295)	(0.292)	(0.298)		
Mother's age	34.24	34.09	34.40	-0.310	0.125
	(6.123)	(6.104)	(6.141)		
Mother's schooling years	6.754	6.702	6.809	-0.107	0.436
	(4.564)	(4.528)	(4.602)		
Mother's political status (CCP=1)	0.027	0.032	0.022	0.010	0.150
	(0.162)	(0.175)	(0.147)		
Observations	4,304	2,178	2,126		
B: F-test of joint significance					
F-statistic and p-value			1.310		0.226

Notes: Data are from the CFPS baseline survey. In panel A, columns 1–3 show the unconditional means for all families, families with a first son, and families with a first daughter, respectively; standard deviations are reported below means; column 4 shows the difference in means across columns 2 and 3, and column 5 shows the p-value for the difference in means. None of the differences are statistically significant at conventional levels. Panel B reports the F-statistic and p-value for joint significance test in regressing the first-son dummy on all other variables in the table.

Table A2 Summary statistics of main outcome variables by sex ratio levels

	High sex ratio			Balanced sex ratio		
	First son	First daughter	Diff. (1)-(2)	First son	First daughter	Diff. (4)-(5)
	(1)	(2)	(3)	(4)	(5)	(6)
Migration: father	0.133	0.072	0.062**	0.097	0.119	-0.023
	(0.340)	(0.258)		(0.295)	(0.325)	
Migration: mother	0.051	0.022	0.029**	0.023	0.026	-0.003
	(0.221)	(0.147)		(0.151)	(0.161)	
Migration: at least one parent	0.150	0.075	0.075**	0.113	0.134	-0.021
	(0.357)	(0.264)		(0.317)	(0.341)	
Working hours, yearly: father	2,606	2,489	117.7	2,353	2,540	-187.4*
	(878.7)	(828.0)		(894.4)	(916.8)	
Working hours, yearly: mother	2,516	2,283	232.7	2,381	2,430	-49.10
	(934.4)	(887.5)		(900.1)	(1,001)	
Housing construction area, sq.m	126.4	114.0	12.39*	127.1	125.2	1.899
	(78.64)	(69.75)		(95.69)	(86.02)	
Housing ownership	0.884	0.822	0.062*	0.823	0.840	-0.018
	(0.321)	(0.383)		(0.382)	(0.366)	
Housing mortgage, thousand	8.528	3.329	5.199**	7.180	5.819	1.361
	(41.75)	(27.12)		(34.42)	(39.30)	
Education expenditure, thousand	1.498	1.741	-0.243	1.934	1.460	0.474***
	(2.793)	(3.221)		(2.961)	(2.600)	
Having an education fund	0.279	0.388	-0.109**	0.331	0.308	0.024
	(0.449)	(0.488)		(0.471)	(0.462)	
Observations	470	442		577	566	

Notes: Data are from the CFPS baseline survey. Columns 1 and 2 show the unconditional means for families with a first son and families with a first daughter, respectively, in counties with a male-biased sex ratio (larger than the third quartile, 1.12); standard deviations are reported below means; column 3 shows the difference in means across columns 1 and 2. Columns 4 and 5 show the unconditional means for families with a first son and families with a first daughter, respectively, in counties with a balanced sex ratio (smaller than the first quartile, 1.02); standard deviations are reported below means; column 6 shows the difference in means across columns 4 and 5. Statistics are weighted by the survey sampling weights.

Table A3 Attrition of parental working hours data and migration status

Dependent variable	Father working time missing=1		Mother working time missing=1	
	(1)	(2)	(3)	(4)
Father migration	-0.047	-0.039		
	(0.041)	(0.040)		
	[-1.157]	[-0.981]		
Mother migration			-0.018	-0.023
			(0.047)	(0.048)
			[-0.387]	[-0.469]
Observations	4,304	4,304	4,304	4,304
R-squared	0.001	0.007	0.000	0.002
Other controls	Yes	Yes	Yes	Yes

Notes: In columns 1 and 2, the outcome variable is a dummy for father working hours missing; in columns 3 and 4, the outcome variable is a dummy for mother working hours missing. Estimations are weighted by the survey sampling weights. Standard errors given in parentheses are clustered at the county level. In square brackets are t-statistics.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table A4 Alternative proxies for local son preference: Parental labor supply

Dependent variable	Migration			Working hours, log	
	Father	Mother	At least one parent	Father	Mother
	(1)	(2)	(3)	(4)	(5)
A: Regional FE					
First son * Sex ratio	0.242**	0.080	0.268**	0.650***	0.239
	(0.106)	(0.065)	(0.104)	(0.201)	(0.304)
Observations	4,304	4,304	4,304	1,527	972
R-squared	0.040	0.018	0.043	0.051	0.068
Baseline controls	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes
First son * Regional FE	Yes	Yes	Yes	Yes	Yes
B: Latitude and longitude FE					
First son * Sex ratio	0.235**	0.066	0.246**	0.530**	0.291
	(0.103)	(0.066)	(0.099)	(0.229)	(0.253)
Observations	4,304	4,304	4,304	1,527	972
R-squared	0.046	0.025	0.050	0.068	0.082
Baseline controls	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes
Lat. lon. FE	Yes	Yes	Yes	Yes	Yes
First son * Lat. lon. FE	Yes	Yes	Yes	Yes	Yes
C: Average education of adult women					
First son * Sex ratio	0.253**	0.115*	0.291***	0.573***	0.479
	(0.102)	(0.064)	(0.101)	(0.187)	(0.397)
Observations	4,304	4,304	4,304	1,527	972
R-squared	0.112	0.065	0.119	0.177	0.273
Baseline controls	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes
Mother schooling years mean	Yes	Yes	Yes	Yes	Yes
First son * Mother sch. yr. mean	Yes	Yes	Yes	Yes	Yes

Notes: Estimations are weighted by the 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 A5 Alternative proxies for local son preference: Multidimensional parental investments

Dependent variable	Housing investment			Child edu. investment		Edu. exp.
	Constr. area, log sq.m	Ownership	Mortgage, thousand	Edu. exp., thousand	Having an edu. fund	/Housing value
	(1)	(2)	(3)	(4)	(5)	(6)
A: Regional FE						
First son * Sex ratio	0.428**	0.304**	19.324*	-2.407***	-0.319*	-0.184**
	(0.211)	(0.146)	(10.899)	(0.817)	(0.187)	(0.084)
Observations	4,159	4,304	4,304	3,969	3,969	3,139
R-squared	0.042	0.040	0.032	0.156	0.059	0.035
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes	Yes
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes
First son * Regional FE	Yes	Yes	Yes	Yes	Yes	Yes
B: Latitude and longitude FE						
First son * Sex ratio	0.431**	0.163	22.281**	-1.999**	-0.323*	-0.164**
	(0.200)	(0.131)	(9.963)	(0.768)	(0.191)	(0.071)
Observations	4,159	4,304	4,304	3,969	3,969	3,139
R-squared	0.119	0.057	0.041	0.173	0.068	0.052
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes	Yes
Lat. lon. FE	Yes	Yes	Yes	Yes	Yes	Yes
First son * Lat. lon. FE	Yes	Yes	Yes	Yes	Yes	Yes
C: Average education of adult women						
First son * Sex ratio	0.486**	0.268**	19.444*	-1.746**	-0.392**	-0.181*
	(0.197)	(0.114)	(10.628)	(0.808)	(0.166)	(0.093)
Observations	4,159	4,304	4,304	3,969	3,969	3,139
R-squared	0.276	0.182	0.183	0.317	0.150	0.120
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mother schooling years mean	Yes	Yes	Yes	Yes	Yes	Yes
First son * M. sch. yr. mean	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Estimations are weighted by the 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 Results on education expenditure accounting for parental migration

Dependent variable	Education expenditure, thousand				
	(1)	(2)	(3)	(4)	(5)
First son * Sex ratio	-1.775**	-1.797**	-1.784**	-1.805**	-1.979*
	(0.793)	(0.801)	(0.794)	(0.802)	(1.035)
Father migration		0.088			
		(0.120)			
Mother migration			0.089		
			(0.243)		
At least one parent migration				0.111	
				(0.122)	
Observations	3,969	3,969	3,969	3,969	3,475
R-squared	0.317	0.317	0.317	0.317	0.321
Baseline controls	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes

Notes: Column 1 repeats the baseline estimation. Columns 2–4 additionally control for a dummy for father migration, a dummy for mother migration, and a dummy for at least one parent migration, respectively. Column 5 excludes in the sample families with any migrant parents. Estimations are weighted by the 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 A7 Endogenous regressors: Proportion of sons and number of children

Dependent variable	Proportion of sons				Number of children	
	(1)	(2)	(3)	(4)	(5)	(6)
First son	0.785***	0.786***	0.787***	0.787***	-0.206***	-0.201***
	(0.014)	(0.014)	(0.013)	(0.013)	(0.025)	(0.023)
	[57.02]	[57.08]	[59.21]	[59.58]	[-8.239]	[-8.948]
Number of brothers			-0.011***	-0.010***	-0.142***	-0.132***
of the father			(0.003)	(0.003)	(0.012)	(0.011)
			[-4.011]	[-3.686]	[-12.10]	[-12.25]
Observations	4,304	4,304	3,772	3,772	3,772	3,772
R-squared	0.797	0.800	0.797	0.800	0.096	0.261
Other controls		Yes		Yes		Yes

Notes: Columns 1 and 2 are the regression results of the proportion of sons amongst the children of the household on the first-son dummy; columns 3 and 4 further control for the number of brothers of the father. Columns 5 and 6 are the regression results of the number of children on the first-son dummy and the number of brothers of the father. Columns 1, 3, and 5 include only the key regressors; columns 2, 4, and 6 further add other family-level controls. Standard errors given in parentheses are clustered at the county level. In square brackets are t-statistics.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table A8 Controlling for son preference sources: Parental labor supply

Dependent variable	Migration			Working hours, log	
	Father	Mother	At least one parent	Father	Mother
	(1)	(2)	(3)	(4)	(5)
A: Controlling for gender wage differential					
First son * Sex ratio	0.254**	0.116*	0.294***	0.535***	0.439
	(0.098)	(0.063)	(0.096)	(0.185)	(0.385)
	[0.589]	[0.660]	[0.590]	[0.439]	[0.787]
R-squared	0.112	0.065	0.119	0.179	0.270
First son * Gender wage diff.	Yes	Yes	Yes	Yes	Yes
B: Controlling for social insurance coverage					
First son * Sex ratio	0.251**	0.113*	0.288***	0.576***	0.437
	(0.101)	(0.066)	(0.100)	(0.183)	(0.386)
	[0.624]	[0.632]	[0.695]	[0.975]	[0.824]
R-squared	0.112	0.065	0.119	0.177	0.270
First son * Social insurance cov.	Yes	Yes	Yes	Yes	Yes
C: Controlling for both					
First son * Sex ratio	0.255**	0.115*	0.294***	0.538***	0.439
	(0.098)	(0.064)	(0.096)	(0.188)	(0.383)
	[0.575]	[0.870]	[0.600]	[0.513]	[0.777]
R-squared	0.112	0.066	0.119	0.179	0.270
First son * Gender wage diff.	Yes	Yes	Yes	Yes	Yes
First son * Social insurance cov.	Yes	Yes	Yes	Yes	Yes
D: Controlling for predictors selected by high-dimensional method					
First son * Sex ratio	0.245**	0.111*	0.279***	0.557***	0.433
	(0.100)	(0.065)	(0.099)	(0.200)	(0.322)
	[0.916]	[0.724]	[0.639]	[0.785]	[0.990]
R-squared	0.112	0.066	0.119	0.177	0.275
First son * High-dim. pre.	Yes	Yes	Yes	Yes	Yes
Observations	4,304	4,304	4,304	1,527	972
Baseline controls	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes

Notes: Panel A adds the interaction between county-level gender wage differential and the first-son dummy. Panel B adds the interaction between county-level social insurance coverage and the first-son dummy. Panel C adds the above two interactions collectively. Panel D adds the interaction between county-level sex-ratio predictors selected by high-dimensional method and the first-son dummy. In square brackets are p-values of Hausman's general specification test for the equality of the coefficient to the baseline. Estimations are weighted by the 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 A9 Controlling for son preference sources: Multidimensional parental investments

Dependent variable	Housing investment			Child edu. investment		Edu. exp.
	Constr. area, log sq.m	Ownership	Mortgage, thousand	Edu. exp., thousand	Having an edu. fund	/Housing value
	(1)	(2)	(3)	(4)	(5)	(6)
A: Controlling for gender wage differential						
First son * Sex ratio	0.489**	0.267**	19.21*	-1.799**	-0.380**	-0.185*
	(0.198)	(0.114)	(10.466)	(0.798)	(0.166)	(0.097)
	[0.620]	[0.836]	[0.673]	[0.779]	[0.908]	[0.785]
R-squared	0.276	0.182	0.182	0.317	0.149	0.119
First son * Gender wage diff.	Yes	Yes	Yes	Yes	Yes	Yes
B: Controlling for social insurance coverage						
First son * Sex ratio	0.502**	0.267**	19.71*	-1.837**	-0.375**	-0.184*
	(0.197)	(0.113)	(10.606)	(0.775)	(0.164)	(0.096)
	[0.597]	[0.788]	[0.598]	[0.567]	[0.686]	[0.616]
R-squared	0.277	0.182	0.183	0.318	0.150	0.119
First son * Social insur. cov.	Yes	Yes	Yes	Yes	Yes	Yes
C: Controlling for both						
First son * Sex ratio	0.500**	0.266**	19.65*	-1.831**	-0.376**	-0.184*
	(0.197)	(0.113)	(10.495)	(0.771)	(0.164)	(0.098)
	[0.616]	[0.791]	[0.632]	[0.628]	[0.770]	[0.712]
R-squared	0.277	0.182	0.183	0.318	0.150	0.119
First son * Gender wage diff.	Yes	Yes	Yes	Yes	Yes	Yes
First son * Social insur. cov.	Yes	Yes	Yes	Yes	Yes	Yes
D: Controlling for predictors selected by high-dimensional method						
First son * Sex ratio	0.482**	0.260**	19.17*	-1.727**	-0.376**	-0.176**
	(0.196)	(0.122)	(10.683)	(0.729)	(0.181)	(0.082)
	[0.943]	[0.698]	[0.760]	[0.719]	[0.857]	[0.484]
R-squared	0.276	0.182	0.183	0.317	0.150	0.120
First son * High-dim. pre.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,159	4,304	4,304	3,969	3,969	3,139
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Panel A adds the interaction between county-level gender wage differential and the first-son dummy. Panel B adds the interaction between county-level social insurance coverage and the first-son dummy. Panel C adds the above two interactions collectively. Panel D adds the interaction between county-level sex-ratio predictors selected by high-dimensional method and the first-son dummy. In square brackets are p-values of Hausman's general specification test for the equality of the coefficient to the baseline. Estimations are weighted by the 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 A10 Testing competing hypotheses: Parental labor supply

Dependent variable	Migration			Working hours, log	
	Father	Mother	At least one parent	Father	Mother
	(1)	(2)	(3)	(4)	(5)
A: Controlling for proportion of households living with grandparents					
First son * Sex ratio	0.248**	0.114*	0.286***	0.577***	0.470
	(0.104)	(0.064)	(0.103)	(0.186)	(0.382)
	[0.734]	[0.819]	[0.963]	[0.904]	[0.504]
R-squared	0.112	0.065	0.119	0.177	0.272
First son * Prop. with grandparents	Yes	Yes	Yes	Yes	Yes
B: Controlling for average financial wealth					
First son * Sex ratio	0.246**	0.114*	0.285***	0.576***	0.448
	(0.100)	(0.063)	(0.098)	(0.185)	(0.373)
	[0.944]	[0.943]	[0.944]	[0.939]	[0.811]
R-squared	0.112	0.065	0.119	0.177	0.273
First son * Average financial wealth	Yes	Yes	Yes	Yes	Yes
C: Controlling for average household income					
First son * Sex ratio	0.267**	0.111*	0.307***	0.538***	0.368
	(0.104)	(0.064)	(0.104)	(0.189)	(0.421)
	[0.368]	[0.762]	[0.762]	[0.762]	[0.357]
R-squared	0.112	0.065	0.119	0.177	0.272
First son * Average household income	Yes	Yes	Yes	Yes	Yes
Observations	4,304	4,304	4,304	1,527	972
Baseline controls	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes

Notes: Panel A adds the interaction between the proportion of households living with grandparents at the county level and the first-son dummy. Panel B adds the interaction between county-level average financial wealth and the first-son dummy. Panel C adds the interaction between county-level average household income and the first-son dummy. In square brackets are p-values of Hausman's general specification test for the equality of the coefficient to the baseline. Estimations are weighted by the 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 A11 Testing competing hypotheses: Multidimensional parental investments

Dependent variable	Housing investment			Child edu. investment		Edu. exp.
	Constr. area, log sq.m	Ownership	Mortgage, thousand	Edu. exp., thousand	Having an edu. fund	/Housing value
	(1)	(2)	(3)	(4)	(5)	(6)
A: Controlling for proportion of households living with grandparents						
First son * Sex ratio	0.471**	0.275**	18.13*	-1.783**	-0.382**	-0.186**
	(0.196)	(0.112)	(10.764)	(0.797)	(0.166)	(0.094)
	[0.676]	[0.608]	[0.575]	[0.813]	[0.967]	[0.825]
R-squared	0.276	0.182	0.183	0.317	0.149	0.119
First son * Prop. w. grandp.	Yes	Yes	Yes	Yes	Yes	Yes
B: Controlling for average financial wealth						
First son * Sex ratio	0.478**	0.270**	18.75*	-1.756**	-0.382**	-0.187**
	(0.199)	(0.117)	(10.257)	(0.738)	(0.167)	(0.090)
	[0.967]	[0.945]	[0.941]	[0.916]	[0.961]	[0.957]
R-squared	0.276	0.182	0.182	0.318	0.149	0.119
First son * Ave. fin. wealth	Yes	Yes	Yes	Yes	Yes	Yes
C: Controlling for average household income						
First son * Sex ratio	0.456**	0.250**	18.41*	-1.826*	-0.399**	-0.173**
	(0.210)	(0.120)	(10.893)	(0.939)	(0.168)	(0.086)
	[0.678]	[0.615]	[0.909]	[0.851]	[0.671]	[0.447]
R-squared	0.276	0.182	0.182	0.317	0.150	0.120
First son * Ave. hh. inc.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,159	4,304	4,304	3,969	3,969	3,139
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes:Panel A adds the interaction between the proportion of households living with grandparents at the county level and the first-son dummy. Panel B adds the interaction between county-level average financial wealth and the first-son dummy. Panel C adds the interaction between county-level average household income and the first-son dummy. In square brackets are p-values of Hausman's general specification test for the equality of the coefficient to the baseline. Estimations are weighted by the 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 A12 Rural and urban families: Parental labor supply

Dependent variable	Migration			Working hours, log	
	Father	Mother	At least one parent	Father	Mother
	(1)	(2)	(3)	(4)	(5)
A: Rural families					
First son * Sex ratio	0.369**	0.220*	0.393**	0.078	1.034
	(0.168)	(0.112)	(0.173)	(0.706)	(0.774)
Observations	2,423	2,423	2,423	530	237
R-squared	0.138	0.088	0.135	0.274	0.634
Dependent variable mean	0.141	0.037	0.159	7.657	7.631
B: Urban families					
First son * Sex ratio	0.140*	0.014	0.170**	0.664***	0.537
	(0.083)	(0.033)	(0.083)	(0.212)	(0.462)
Observations	1,881	1,881	1,881	997	735
R-squared	0.146	0.072	0.141	0.268	0.284
Dependent variable mean	0.050	0.011	0.056	7.763	7.726
C: Test of coefficient equality between rural and urban					
P-value	0.209	0.039	0.226	0.384	0.498
Baseline controls	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes

Notes: Panel A reports results for rural families and panel B, urban families. Panel C reports p-values of Hausman's general specification tests that the estimated coefficient in each column of panels A and B is equal. Estimations are weighted by the 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 A13 Rural and urban families: Multidimensional parental investments

Dependent variable	Housing investment			Child edu. investment		Edu. exp.
	Constr. area, log sq.m	Ownership	Mortgage, thousand	Edu. exp., thousand	Having an edu. fund	/Housing value
	(1)	(2)	(3)	(4)	(5)	(6)
A: Rural families						
First son * Sex ratio	0.542**	0.337**	16.889*	-1.669***	-0.316	-0.348***
	(0.241)	(0.161)	(9.419)	(0.624)	(0.275)	(0.129)
Observations	2,311	2,423	2,423	2,239	2,239	1,784
R-squared	0.297	0.216	0.673	0.349	0.167	0.189
Dependent variable mean	4.714	0.888	2.792	0.870	0.230	0.056
B: Urban families						
First son * Sex ratio	0.196	0.282*	23.547	-2.082	-0.328	-0.054
	(0.262)	(0.161)	(22.792)	(1.380)	(0.214)	(0.037)
Observations	1,848	1,881	1,881	1,730	1,730	1,355
R-squared	0.367	0.211	0.136	0.313	0.213	0.135
Dependent variable mean	4.581	0.767	8.403	2.244	0.375	0.020
C: Test of coefficient equality between rural and urban						
P-value	0.286	0.805	0.791	0.782	0.970	0.012
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
First son * Controls	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Panel A reports results for rural families and panel B, urban families. Panel C reports p-values of Hausman's general specification tests that the estimated coefficient in each column of panels A and B is equal. Estimations are weighted by the 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.