DISCUSSION PAPER SERIES

DP16864

Multidimensional pre-marital investments with imperfect commitment

Venkataraman Bhaskar, Wenchao Li and Junjian Yi

DEVELOPMENT ECONOMICS

CEPR

ISSN 0265-8003

Multidimensional pre-marital investments with imperfect commitment

Venkataraman Bhaskar, Wenchao Li and Junjian Yi

Discussion Paper DP16864
Published 06 January 2022
Submitted 05 January 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Multidimensional pre-marital investments with imperfect commitment

Abstract

We analyze a model of multidimensional parental investments under imperfect marital commitment. Parents choose to invest in their child's human capital and in housing, which is a public good within marriage. If men are in excess supply in the marriage market, but also enjoy greater bargaining power within marriage, their parents will overinvest in housing to credibly commit to share more resources with potential wives. This may crowd out human capital investment. To test the theory, we use nationally representative Chinese household survey data to examine the effect of a more male-biased local sex ratio upon investments in boys, relative to investments in girls.Parents with boys increase labor supply and migrate more to improve earnings and investments. They increase housing investment while reduce educational investments.

JEL Classification: J12, J13, J16, J18, J24, D10, O15, J61

Keywords: Premarital investment, Imperfect commitment, Sex-ratio imbalances, Public Goods, Human capital investment, human capital development

Venkataraman Bhaskar - v.bhaskar@austin.utexas.edu
University of Texas at Austin and CEPR

Wenchao Li - wenchaoli.1022@gmail.com
Tongji University

Junjian Yi - junjian.yi@gmail.com
Chinese University of Hong Kong

Multidimensional Premarital Investments with Imperfect Commitment*

V. Bhaskar† Wenchao Li‡ Junjian Yi§

January 5, 2022

Abstract

We analyze a model of multidimensional parental investments under imperfect marital commitment. Parents choose to invest in their child's human capital and in housing, which is a public good within marriage. If men are in excess supply in the marriage market, but also enjoy greater bargaining power within marriage, their parents will overinvest in housing to credibly commit to share more resources with potential wives. This may crowd out human capital investment. To test the theory, we use nationally representative Chinese household survey data to examine the effect of a more male-biased local sex ratio upon investments in boys, relative to investments in girls. Parents with boys increase labor supply and migrate more to improve earnings and investments; and they increase housing investment while reduce educational investment.

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

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

*We are grateful to Dan Ackerberg, three anonymous referees, and the editor (James Heckman) for very useful comments. We thank Sangwoo Choi for capable research assistance.

†University of Texas at Austin and CEPR; 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

What determine premarital investments in children when marriage market considerations are important? Much of the literature examines the question within the transferable utility paradigm of Becker (1973) and Shapley and Shubik (1971), and includes Cole et al. (2001), Iyigun and Walsh (2007), Chiappori et al. (2017), and Zhang (2021). This effectively assumes full commitment at the time of marriage. However, recent work has begun to examine the implications of imperfect commitment (e.g., Anderson and Bidner, 2015; Galichon et al., 2019). Indeed, some papers assume non-transferable utility, an extreme instance of imperfect commitment (e.g., Peters and Siow, 2002; Bhaskar and Hopkins, 2016).

This paper studies how imperfect commitment within marriage affects premarital investments in children undertaken by their parents, theoretically and empirically. Our empirical work relates to China, where the imperfect commitment assumption is particularly compelling. 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. The effects of imperfect commitment are magnified by the divergence in the relative bargaining powers of men and women, between the ex-ante stage (before marriage) and the ex-post stage (after marriage).

Our focus is on the *composition* of investments, i.e. its division between bequeathed physical capital (housing) and human capital. We assume that marital partners are unable to commit, at the time of marriage, to share future household resources in a pre-agreed fashion. If a man invests in human capital, which will increase his future labor earnings, the sharing of this between spouses is determined by ex post—rather than ex ante—bargaining, and the shares reflect the greater bargaining power of the man. If a man invests in housing, which is a public good and thus non-excludable, spouses jointly consume it without bargaining. Thus, investment in housing by the parents with a boy provides a credible commitment to share resources with the future spouse. In other words, 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 human capital. If the costs of investments exhibit sufficient complementarity, human capital investment in sons may decline as marriage market competition intensifies. We derive these results using a theoretical model with multidimensional investments; we show existence and uniqueness of a stable equilibrium, with intuitive comparative statics properties.

Our empirical work is based on data from a nationally representative Chinese household survey. We examine the effect of county-level sex ratios (defined as the ratio of men to women) on investment decisions for families with a first-born son, using families with a first-born daughter in the same county as a comparison group.The identification partly relies on the fact that the gender of the first child is plausibly random; gender selections in China typically occur at second and higher order births (see Figure 1).

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 increases. In China, migration substantially raises family income, thereby permitting larger investments in children. More importantly, the composition of investments is affected by the sex-ratio imbalance. Specifically, a 0.1 increase in the local sex ratio is associated with a 25.1 percent increase in the probability of having a migrant father, a 4.8 percent increase in housing construction area, and a 11.8 percent decline in annual education expenditure per child, for first-son families relative to first-daughter families.[1]

An important challenge to our empirical work is potential endogeneity of sex ratios—preference for sons in the Chinese culture may cause high sex ratios and also affect parental investments. Nonetheless, our theoretical model implies that unobserved heterogeneity in son preference across counties would cause a positive correlation in both dimensions of investments, in contrast with the differential effects of the sex ratio on investments between housing and child education. We also control for county fixed effects and compare first-son families with first-daughter families within a county, to reduce the confounding effects of unobserved preference that affects investments of the two types of families in a county in a similar manner.

More importantly, we construct a novel instrument, the real level of penalties imposed on healthcare workers for providing services that facilitate gender selections, which we show significantly reduces the local sex-ratio imbalance. Prior work has used penalties imposed on families violating the one-child policy as an instrument for analyzing the effects 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 the demand for sex-selection services. In contrast, we are the first to use a supply-side instrument, which does not directly affect fertility and other parental decisions, to explore the quasi-experimental variation in sex ratios induced by the supply of sex-selection services. We also provide evidence that the supply-side instrument is not significantly correlated with parental son preference. Our estimates using instrumental variables confirm our qualitative findings.

Another challenge to our empirical work is endogenous fertility—since parents whose first child is a girl are more likely to have a second child than those with a first boy,[2] they are more likely to be resource constrained. By explicitly modeling fertility, we are able to show

[1]Note that 0.1 is the overall increase in the sex ratio in China between 2002 to 2010. It also corresponds to one standard deviation of the 2010 county-level sex ratios.

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

that this does not bias our estimates of the main coefficient of interest. Furthermore, we obtain similar results if we either restrict the sample to one-child families, or control for fertility in the unrestricted sample, using suitable instruments for fertility and the number of boys in the latter specification.

1.1 Related literature

We have already discussed the most relevant literature on premarital investments, and so we will not repeat this discussion here. Most of this literature focuses on one-dimensional investments. An exception is Dizdar (2018), who examines how multidimensional investments give rise to the possibility of coordination failures and inefficiencies even in a transferable utility setting, although an efficient equilibrium exists.

Anderson and Bidner (2015) develop a model where marital payments—dowries or bride prices—are a more effective commitment device than education, thereby highlighting the role of imperfect commitment. Similarly in Nunn (2005), bride price serves as a credible commitment by men. Since our model has two types of real investments, and non-transferable utility, we are able to examine allocative distortions arising from marriage market competition.[3]

Our work also relates to the literature on decision-making within the family. The recent textbook, Browning et al. (2014), focuses mainly on full commitment. Lundberg and Pollak (1993) pioneered the idea of "separate-spheres" bargaining. The implications of the inability to commit have been examined in several contexts: 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 popular household models with limited commitment.

Our paper also contributes to the literature on sex-ratio imbalance in China and other societies (e.g., Chiappori et al., 2002; Ebenstein, 2010; Bhaskar, 2011; Ebenstein, 2011; Wei and Zhang, 2011; Bhaskar and Hopkins, 2016; Lin et al., 2016). Our empirical analysis builds upon Wei and Zhang (2011), who show that parents facing high sex ratios competitively save more to improve marriage market prospects of their sons.

Finally, our empirical work, which examines how the human capital development of children is affected by parental migration and investments, also relates 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).

[3] Anderson and Bidner (2015) do not examine welfare effects.

2 Multidimensional investments: Theory

We now set out a model of multidimensional premarital investments with imperfect commitment within marriage. We highlight the difference between physical capital (housing) and human capital. The distinguishing characteristic of housing is that it is a public good within marriage—both partners share the house equally. Human capital, on the other hand, gives rise to earnings that are individual specific. The share of a spouse in these earnings depends upon relative bargaining powers after marriage, and not on marriage market conditions.[4] Individuals realize that their attractiveness on the marriage market depends on the composition of their ex ante investment, given the lack of commitment. We show the existence of a unique, stable equilibrium, and examine how how equilibrium investments differ from the investments chosen by a utilitarian social planner. The model provides us with rich comparative statics—in particular, when men are in excess supply, and have large ex post bargaining power, the parents with boys will overinvest in housing, and may consequently reduce investments in their son's human capital.

2.1 Balanced sex ratio

Assume a continuum of men and a continuum of women of with equal measure. At the ex ante stage, the parents with a boy have to choose a vector of investments for their son, (x_B, y_B). x_B is investment in a private good, such as the son's human capital. y_B is investment in good which is public within marriage, such as the purchase of a house. Similarly, the parents with a girl choose a vector of investments (x_G, y_G). The costs of investment, in terms of foregone consumption utility for the parents, are given by a common function $c : \mathbb{R}^2_+ \to \mathbb{R}_+$.

Normalize the return on investments on each good to one, so that a unit of investment yields one unit of the good.[5]

Finally, we assume that returns to the private good are stochastic: each boy is subject to a zero-mean shock ε, so that the realized return on investment in a boy equals $x_B + \varepsilon$. Similarly, each girl is subject to a zero-mean shock η, so that the realized return equals $x_G + \eta$. Shocks for boys have a distribution function F and density f, and while those for

[4] Conceptually, our analysis can be applied to any situation with multidimensional investments, where the sharing rules for the returns differ across investment categories. The distinction between public and private goods is a leading example, but by no means the only one. For example, individuals may have greater bargaining power over labor income than over income derived from financial assets. Our model can be extended to such a context. Our model can also be adapted to account for marital transfers such as dowry and bride price, the sharing rules of which are determined ex ante.

[5] This is without loss of generality since the cost functions capture any non-linearity. Consequently, convexity of the cost function captures both diminishing returns to investment and decreasing marginal utility of parental consumption.

girls have a distribution function G and and density g.

Bhaskar and Hopkins (2016) show that when investments are one-dimensional such shocks ensure existence and uniqueness of a (quasi-symmetric) equilibrium in investment levels, under appropriate distributional assumptions. We extend the analysis to multidimensional investments, where the public good component gives rise to strategic interaction between the investments in the two sexes.

We assume that two parties who agree to marry cannot commit to a sharing rule of the returns to the parental investments in any private good, including the stochastic component. Instead, the shares in the private good are determined by ex post bargaining. We model this by assuming that a man has a share λ_B in the returns, with his partner securing the remaining share, $1-\lambda_B$. Similarly, a woman has share λ_G of the returns to in her investment in the private good good, the remainder going to her spouse. These simple sharing rules can be derived from cooperative or non-cooperative models of bargaining, as in Nash (1950) or Rubinstein (1982). As Shaked and Sutton (1984) have shown, the shares will not depend upon the outside options of the two parties, as implied by marriage market conditions, unless these outside options are binding.[6] The public goods consumption by the couple equals the sum $y := y_B + y_G$, with a man's payoff being $v_B(y)$ and a woman's payoff being $v_G(y)$. We assume:

Assumption 1 *1. The functions $c(.), v_B(.), v_G(.), f(.)$, and $g(.)$ are continuously twice differentiable. The cost function $c(.)$ is increasing and strictly convex, and $v_B(.)$ and $v_G(.)$ are increasing and strictly concave.*

2. $c_x(0,y) = 0$ and $\lim_{x\to\infty} c_x(x,y) = \infty$ for any y. $c_y'(x,0) = 0$ and $\lim_{y\to\infty} c_y(x,y) = \infty$ for any x.

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

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

$$\lambda_G(x_G + \eta) + (1-\lambda_B)(x_B + \varepsilon) + v_G(y_B + y_G). \tag{2}$$

[6] Since divorce is rare in China, one may assume that the outside options are unlikely to be binding, and therefore irrelevant in the calculation of the ex ante returns from investments.

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

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

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

$$\int \phi'(\varepsilon) f(\varepsilon) d\varepsilon = \int \frac{f(\varepsilon)}{g(\phi(\varepsilon))} f(\varepsilon) d\varepsilon =: \theta_B. \tag{3}$$

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

$$c_x(\hat{x}_B, y_B) = \lambda_B + (1 - \lambda_G)\theta_B. \tag{4}$$

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

$$c_x(\hat{x}_G, y_G) = \lambda_G + (1 - \lambda_B)\theta_G, \tag{5}$$

$$\theta_G := \int \frac{g(\eta)}{f(\phi^{-1}(\eta))} g(\eta) d\eta. \tag{6}$$

Consider next the public good. The key point is that free rider problem does not bedevil such investments, since they improve an individual's attractiveness on the marriage market. The marginal benefit from increasing $\hat{y}_B$ to $\hat{y}_B + \Delta$ is that for any ε, the boy is as attractive as type $\hat{\varepsilon}$ that satisfies

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

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

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

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

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

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

$$c_y(\hat{x}_B, y_B) = v'_B(\hat{y}_B + y_G) + \frac{1 - \lambda_G}{1 - \lambda_B}\theta_B v'_G(\hat{y}_B + y_G). \tag{7}$$

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

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

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

Equations (4), (5), (7), and (8) characterize the best response investments in the private good and the public good by both the sexes, that satisfy the first order conditions.[7] Our first result is the following proposition, which is proved in Appendix A.

Proposition 1 *Under assumption 1, there exists a unique quasi-symmetric equilibrium,*

[7]It is routine to verify that the second order conditions are also satisfied. We will simply assume that these conditions are sufficient, and that large deviations are unprofitable; Bhaskar and Hopkins (2016) have shown that this is indeed so for one-dimensional investments, under appropriate assumptions on costs and the distributions of shocks. We do not anticipate any difficulty in extending this argument to the two-dimensional case, but this would divert from the focus of the present paper.

which is stable.

Note that uniqueness and stability ensure that the comparative statics will be well defined and intuitive.

Welfare implications Our benchmark is that of a utilitarian social planner who chooses the levels of investments, but who cannot dictate the sharing rule, in order to maximize the ex-ante expected utility of the parent, before she observes the sex of her child. Since the child is equally likely to be a boy or a girl, the planner will give equal weight to their utilities.[8] The *utilitarian efficient* investment profile (x_B^{**}, y_B^{**}) satisfies the first order conditions:

$$c_x(x_B^{**}, y_B^{**}) = 1, \tag{9}$$

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

We say that there is *underinvestment* (resp. *overinvestment*) in the private good by one of the sexes, say men, if the marginal cost evaluated at the profile, $c_x(x_B^*, y_B^*) < 1$ (resp. > 1). Similarly, there is underinvestment in the public good by men, if the marginal cost $c_y(x_B^*, y_B^*)$ is less than the sum of the marginal benefits to the spouses. This notion differs from a simple comparison of the equilibrium investment level with the utilitarian level.[9]

How do equilibrium investments compare to the efficient ones? First, one has efficiency when the sexes are symmetric.

Proposition 2 *Suppose that $F = G$ and $\lambda_B = \lambda_G$, so that both sexes have the same distribution of shocks and the two bargaining powers, over labor income, are equal. Then investments in public goods and private goods are utilitarian efficient.*

The proof of this result follows from a verification that the first order conditions for equilibrium investments coincide with the conditions for utilitarian efficiency, under the two assumptions of the proposition. Consider the first order condition for investments in the public good by men, equation (7), and focus on the coefficient on $v_G'(y)$, the marginal utility of the spouse from the public good. When the shocks have the same distribution, a man of type ϵ is matched to a woman of the same type. Consequently, the ratio of the two densities in expression for θ_B, equation (3), equals one for every ϵ, and thus $\theta_B = 1$. Further, since $1 - \lambda_B = 1 - \lambda_G$, the ratio of the two terms equals one. Thus the weight on female marginal

[8]The implications of possible gender preference are discussed at the end of this section.

[9]This difference arises for the private good since the cost function is not separable—if it is so, and $c_{xy} = 0$, then the two notions coincide. For the public good, strategic interaction between the two sides of the market gives an additional reason for non-coincidence of the two notions.

utility equals one, and the first order condition reduces to Samuelson's efficiency condition for public goods. A similar argument applies to the condition for male investments in the private good, and for women's investments in both goods.

There are two notable features of this result. First, there is no trace of the free rider problem that normally bedevils investments in public goods. Even if demands for the public good differ greatly between the sexes, both sides will provide it. This follows from the fact that investments take place prior to matching. Consequently, if one side of the market, say women, value the public good more, an individual man has an incentive to provide it, in order to improve his attractiveness on the marriage market.

Second, the efficiency result may appear unexpected. Even if there is no equal sharing of income, with each partner having a larger bargaining power over the returns from the investments in human capital, this does not result in investment inefficiency, as long as both sexes are in the same position. Indeed, the proposition can be generalized as follows: boys' investments in both public and private goods are efficient as long as $\frac{1-\lambda_G}{1-\lambda_B}\theta_B = 1$; girls' investments are efficient as long as $\frac{1-\lambda_B}{1-\lambda_G}\theta_G = 1$. Of course, if the shock distributions are the same, $\theta_B = \theta_G = 1$, and if bargaining power over own labor incomes are equal, then $\frac{1-\lambda_G}{1-\lambda_B} = 1$. We now examine the role of differences in bargaining power between the sexes.

Proposition 3 *If men have more bargaining power over own income than women so that $\lambda_B > \lambda_G$, and if $F = G$, then men overinvest in the private good, and also overinvest in the public good, while women underinvest in both types of goods, relative to the utilitarian investment levels.*

This proof also follows from an examination of the first order conditions for equilibrium. When $F = G$, $\theta_B = \theta_G = 1$. Men have a greater incentive to invest in the private good, since they get a larger share of the return, and also a relatively larger share of the return of their partner. Thus the marginal benefit exceeds one. For women, the marginal benefit is less than one, and so they underinvest in the private good. More subtle is the incentive of men to overinvest in the public good. Men invest in the public good since it is an effective way to compete on the marriage market, more effective than investing the private good, where they cannot guarantee their partner a larger share. Effectively, investment in the public good acts as a commitment device for men.

Suppose now that marriage market competition favors women, while men have greater bargaining power within marriage. One would expect that the incentive of men to overinvest (and of women to underinvest) is magnified. To model a balanced marriage market that favors women, let us assume that the shocks for women are more dispersed than the shocks for men. Formally, we assume that distribution G is larger in the dispersive order than a distribution F, or $G \geq_d F$, i.e.:

$$g(G^{-1}(z)) \leq f(F^{-1}(z)) \text{ for all } z \in (0,1), \tag{11}$$

with the inequality being strict on a set of z values with positive measure; see Shaked and Shanthikumar (2007, pp148–9).[10]

If $G \geq_d F$, so that $f(\varepsilon)/g(\phi(\varepsilon)) \geq 1$ for all values of ε, and is strictly less on a set values of ε of positive measure. Thus $\theta_B > 1$, and $\theta_G < 1$. Observe that the incentive for men to invest in both goods is given by the *product* of θ_B and $1 - \lambda_G$. Thus increased marriage market competition interacts with bargaining power asymmetry to have a multiplier effect. We summarize this discussion in the following proposition.

Proposition 4 *Suppose that $G \geq_d F$, so that the shocks are more dispersed for women than for men, so that marriage market competition favors women. Suppose also that $\lambda_B > \lambda_G$, so that men's bargaining power is greater than women's. Both these factors cause men to overinvest in both goods and women to underinvest in both goods. The two forces, marriage market competition and bargaining power asymmetry, interact to have a larger multiplier effect on investments.*

Comparative statics We now examine the effects of a change in bargaining power of one of the sexes upon equilibrium investments. Since we have shown that the equilibrium is stable, we would expect that the comparative statics would be intuitive. Nonetheless, it is entirely possible that an increase in male bargaining power *reduces* the level of investment in the private good, while increasing investment in the public good. This arises due to fact that the cost function $c(x, y)$ is not separable in its two arguments. Consequently, if $c_{xy}(.)$ is large, and the marriage market competition is very important for males, their desire to be more attractive on the marriage market increases public good investment. This raises the marginal cost of private good investments, and reduces its level, even though the marginal benefit from such investments also increases. Similarly, women will invest less in the public good, and this may prompt greater private good investments on their part, despite a fall in the return. Observe that the reduction in women's public good investments will induce even greater public good investments by the men, due to the fact that the public good evaluation functions are strictly concave. Similarly, the increase in male public good investments will induce a further reaction in women's investments in the public good. This is summarized in

[10] For example, if F and G are both uniform distributions, where the support of G is a longer interval than that of F, then $G \geq_d F$. A second example is two normal distributions—the one with the higher variance is larger in the dispersive order.

the following proposition.

Proposition 5 *Suppose that marriage market competition favors women, so that $\theta_B > 1$ and $\theta_G < 1$. An increase in men's bargaining power λ_B increases public goods investments by men, an reduces that by women. If c_{xy} is large enough, equilibrium investments by men in the private good fall. Furthermore, if marriage market competition for men increases while that for women men falls, this also increases men's investments in the public good, and may reduce their investments in the private good.*

The proof of this proposition is presented in Appendix B. Apart from the analytical proof, we also present an illustrative numerical example, which illustrates the equilibrium values of investments by both sexes as λ_B changes. The intuition for this illustrates the role of imperfect commitment. As λ_B increases, with λ_G remaining fixed, men get a larger share of the total private good in marriage. But this makes men who invest more in the public good more attractive to women. Consequently, men invest more in the public good. Since c_{xy} is large, this increases the cost of investing in the private good, and men invest less in it, even though the return to such investments has increased.

The example also illustrates that as θ_B increases and θ_G falls, investments in the public good by men increases, while those in the private good may fall. That is, increased marriage market competition induces men to invest more in the public good, as this is a credible way to commit on the marriage market when their bargaining power over the private good is large.

2.2 Unbalanced sex ratios

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

Our modeling strategy incorporates the following innovation, which allows the sex ratio to affect investment incentives in a continuous fashion.[11] Suppose that the ratio of women to men is $r < 1$. We assume that the overall marriage market is composed of many local marriage markets, where the sex ratio varies. In some of these marriage markets, there is an excess of men, while in the others, the marriage market is balanced. A reduction in r, the aggregate ratio of women to men, increases the likelihood that an individual woman resides in a market where there is an excess of men. A simple way of modeling this is as follows. Fix $\hat{r} < 1$, and let this be sufficiently small so that the aggregate sex ratio, r, lies in the interval $(\hat{r}, 1]$. The sex ratio in a local market takes one of two values, $\hat{r}$ and 1, where the probability of the first value is $\rho(r)$. Since the aggregate sex ratio equals r, we deduce that $\rho(r) = \frac{1-r}{1-\hat{r}}$.

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

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

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

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

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

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

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

Let us now define θ_{B+}, as follows:

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

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

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

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

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

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

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

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

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

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

$$c_y(x_B^*, y_B^*) = \begin{array}{c} \rho(r)\left[(1-\hat{r})v_B'(y_B^*) + \hat{r}v_B'(y_B^*+y_G^*) + \hat{r}\frac{1-\lambda_G}{1-\lambda_B}v_G'(y_B^*+y_G^*)\theta_{B+} + \frac{v_G'(y_B^*+y_G^*)}{1-\lambda_B}\hat{f}\bar{U}\right] \\ +(1-\rho(r))\left[v_B'(y_B^*+y_G^*) + \frac{1-\lambda_G}{1-\lambda_B}v_G'(y_B^*+y_G^*)\theta_B\right]. \end{array} \tag{14}$$

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

$$c_x(x_G^*, y_G^*) = \lambda_G + (1-\lambda_B)\left[\rho(r)\theta_{G+} + (1-\rho(r))\theta_G\right]. \tag{15}$$

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

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

$$c_y(x_G^*, y_G^*) = v_G'(y_B^*+y_G^*) + \frac{1-\lambda_B}{1-\lambda_G}v_B'(y_B^*+y_G^*)[\rho(r)\theta_{G+} + (1-\rho(r))\theta_G]. \tag{16}$$

An interior equilibrium, where both sexes invest positive amounts in the public good, is given by the profile of investments that satisfy the first order conditions (13), (14), (15), and (16). Existence, uniqueness and stability of equilibrium is confirmed by the following proposition.

Proposition 6 *For any $r \leq \hat{r}$,there exists a unique quasi-symmetric equilibrium, which is stable, provided that assumption 1 is satisfied.*

We omit the proof since it follows the same line of argument as that of proposition 1.

Welfare implications As before, our benchmark is utilitarian efficiency. For private goods, the return on investment equals 1, and this return is either shared if the individual marries, or accrues entirely to the individual if he remains single. Since the utilitarian planner puts equal weight on both partners, the first order condition for utilitarian efficiency remains $c_x(.) = 1$, for both men and women.

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

$$c_y(x_B^{**}, y_B^{**}) = r[v_B'(y_B^{**}+y_G^{**}) + v_G'(y_B^{**}+y_G^{**})] + (1-r)v_B'(y_B^{**}). \tag{17}$$

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

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

$$c_y(x_G^{**}, y_G^{**}) = v'_B(y_B^{*,*} + y_G^{**}) + v'_G(y_B^{**} + y_G^{**}). \tag{18}$$

As in the analysis of balanced marriage markets, we will say that a sex overinvests in a good, private or public, if the marginal cost of investment exceeds the first-best utilitarian criterion. We therefore compare the first order conditions at an equilibrium, (13), (14), (15), and (16), with the conditions for utilitarian efficiency. Recall that the efficiency condition for private good is that the marginal cost equals 1, while for the public good, these are given by equations (17) and (18). In order to focus on the effects of the sex ratio per se (rather than on differences in bargaining power or the distribution of shocks), let us assume that $\lambda_B = \lambda_G$, and that $F = G$.

Proposition 7 *Suppose that the sexes are otherwise symmetric, except the imbalance in the sex ratio, and that the common distribution of shocks has a density that is weakly increasing. An imbalance in the sex ratio, $r < 1$, leads men to overinvest in both public and private goods, and women to underinvest in both goods.*

The proof of this proposition follows from a comparison of the first-order conditions for equilibrium with those for utilitarian efficiency. Since bargaining powers are equal, $\lambda_B = \lambda_G$, when the shocks are identically distributed for the two sexes, $\theta_B = \theta_G = 1$. Furthermore, when the density of shocks is weakly increasing, it may be verified that $\theta_B^+ \geq 1$, while $\theta_G^+ \leq \hat{r}$. Finally, note that by definition, $\rho(r)\hat{r} + (1 - \rho(r)) = r$.

From the condition for men's investment in the private good, equation (13), we see that $c_x(x_B^*, y_B^*) > 1$, for two reasons. First, since the mass of women is smaller than that of men, women's qualities are effectively more dispersed, intensifying competition by men. Second, men have an additional incentive to compete in order to avoid being left single—this is captured by the term $\hat{f}\bar{U}$. Conversely, the first order condition for women's investment in the private good, equation (15), shows that women underinvest, since the qualities of men are effectively less dispersed.

Turning to the public good, we see from equation (14) that men overinvest in the public good in order to avoid being left single. From equation (16), we see that women invest less in the public good, for two reasons. First, they underweight the effect on their utility of their partner by a fraction r, even though efficiency dictates that they give this weight one. Second, there is a crowding out effect: Since men invest more in public goods, women have less incentives to invest.

Comparative statics of the sex ratio Our focus is on the comparative statics effects on the equilibrium investment with respect to r, the sex ratio. This can be derived by differentiating the system of equations given by the first order conditions (13), (14), (15), and (16), with respect to r. Appendix C presents the comparative statics results. These are relatively complex, and are difficult to interpret. Consequently, we have resorted to numerical methods, investigating the pattern of behavior for a large range of parameter values. We are particularly interested in the case where $\lambda_B > \lambda_G$, so that men have greater bargaining power over their own labor market income than women do, although we also explore the case where bargaining powers are equal. We find the following results:

- Boys' investments in the public good are always decreasing in r; girls' investments in the public good are always increasing in r. That is, if a sex faces increased marriage market competition, it invests more in the public good.[12]

- The effects of changes in r upon private good investments are ambiguous, for both sexes. While the direct effect of increased marriage market competition increases incentives for both types of investment. However, the direct effect may be larger for the public good, and the increased marginal costs from investing in the public good may reduce private good investments. This is the case when bargaining powers are asymmetric (i.e. $\lambda_B >> \lambda_G$) and when the cost function exhibits supermodularity, so that c_{xy} is large.

Appendix C presents one illustrative example when bargaining powers are asymmetric. It shows that the effects on private good investments by boys are ambiguous—these investments are decreasing in r for r values near 1, but are increasing in r for smaller values of r. In other words, for some range of r values, the pressure to increase public good investments arising from marriage market competition crowds out private good investments, thereby inducing a reduction in such investments. Similarly, girls' investments in the private good are decreasing in r for r values close to 1, but are increasing in r when the sex ratio is more unbalanced.

[12]While these are numerical results, this is true for all the parameter values we studied.

To summarize, our analysis of investments in unbalanced marriage markets shows that boys have a direct incentive to invest more, both in the private good and in the public good, so that they overinvest relative to the utilitarian benchmark. Conversely, girls underinvest in both goods relative to the efficient benchmark. These effects are more pronounced for public goods, where investments of the two sides are strategic substitutes. Consequently, boys' public good investments are increasing in the sex ratio, and girls' are decreasing. Since increased public good investments raise the marginal costs of private good investments, boys may invest less in the private good in response to an unfavorable sex ratio, while girls may invest more.

Son preference and investments A final observation: a minor modification of our model allows us to accommodate son preference, where parents value the welfare of boys more than that of girls. To do this, it suffices to multiply the expression for the welfare of boys in equation (1) by a weighting factor $\alpha > 1$. The comparative statics of stronger son preference, i.e. an increase in α, are straightforward—parents with boys would invest more in housing and in human capital, as compared to those with girls. This is of relevance for our empirical work: If there is unobserved heterogeneity in α across counties, it would cause a positive correlation in both dimensions of investments, in contrast with the effects of the sex ratio.

3 Empirical evidence from China

A major difficulty arises in testing the theory that we have set out. Our main predictions relate to the effects of a macro or region-level variable—marriage market competition—upon individual parental decisions, and it is hard to come up with good measures of regional variation in marriage market competition in developed economies. This problem does not arise in China, since the sex-ratio imbalance is large and displays significant regional variation. Furthermore, the restrictions imposed by the Chinese state on internal migration means that marriage markets are essentially local.[13]

3.1 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

[13]Instead of fully estimating the model, the data requirement of which is unpractically high, we follow the premarital investments literature to mainly focus on investment decisions. Related studies include Wei and Zhang (2011), Cameron et al. (2019), and Li et al. (2020); they presume a greater likelihood of being married for men with more wealth.

stabilized at this level subsequently. This imbalance is mainly due to 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). These high sex ratios contribute to the current oversupply of men and intensify their competition in the marriage market.[14] A growing literature examines the economic and social consequences of these imbalances, including on parental investments (Wei and Zhang, 2011) and crime (Edlund et al., 2013).

Random first-child gender Although sex selection is rife in China, the gender of the first child in a family is not selected. 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; see Figure 1. Parents are unlikely to practice gender selection for the first birth—51.1 percent of the families in our data 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.[15] The strongest evidence in favor of the randomness of first-child gender is that first-son and first-daughter families have similar predetermined parental and household characteristics in our data; in a balance test, none of the differences in the characteristics are statistically significant at conventional levels (panel A, Appendix Table A1). In addition, we regress the first-son dummy on the full set of variables used in our analysis and find an insignificant F-statistic for joint significance, indicating that we cannot reject the null that none of the variables are correlated with the first-child gender (panel B).

Marital frictions A key assumption underlying our analysis is that the sharing of consumption from different income sources reflects ex post bargaining power rather than marriage market conditions. This arises in part from the difficulty in obtaining divorce and in securing alimony. Divorce and separation are infrequent in China, and often, women are not granted alimony; see Civil Code of China: Part V (2020). Consequently, a woman who may be in a favorable position prior to marriage may lose this advantage on marrying.

Asymmetric information also implies that an income earner may command a greater share of his or her own earnings. For example, Doepke and Tertilt (2016) use laboratory and field experiments to show that private information plays an important role in the allocation

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

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

of household resources in developing countries. Xing (2017a,b) finds that married Chinese men are more likely to put their earnings into their own personal savings accounts when these choices are private information.[16]

Migration as a form of labor supply In China, migration is usually temporary and circular, and an important part of aggregate labor supply (Zhao, 1999). At the start of each year, migrant workers make temporary migration decisions, leaving their families behind in order to increase earnings. At the end of the year, they usually return home for the Spring Festival. This temporary and circular migration is pervasive in rural as well as urban areas. China's *hukou* (household registration) system provides different types of permits for individuals to live and/or work. These permits are restrictive, making it difficult for migrants to assimilate with the local population. Migrant workers have limited access to benefits that are available to local residents, and their children are often denied access to public schools (Zhao, 1999).

Premarital housing investment In China, housing has traditionally played an important role in determining marriage prospects. 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. Wrenn et al. (2019) have a similar finding. Interestingly, the marriage outcomes of a daughter are unrelated to the parents' house. Recent survey evidence documents the importance of housing in marital decisions.[17]

Housing investment by parents plays an important role in the marriage market (Pierson, 2010). Not only are houses often bequeathed (Xie and Jin, 2015), most marriage-age men are unable to accumulate enough wealth to afford a house on their own. It should also be noted that intergenerational co-residence is common, especially in rural China—more than 70 percent of young couples live with the groom's parents in the first few years after marriage. Consequently, the quality of the family home plays an important role, since a newly-wed woman values her privacy.

[16] More anecdotally, a journalist reports that about 70 percent of married men in China have personal savings (Liu, 2013).

[17] According to a survey conducted in 2019, over 70 percent of the single respondents who were born after 1990 stated that they wanted a property before getting married, and over 80 percent of the female respondents perceived owning a property as a prerequisite for marriage, as per a statistical report (Ma, 2020). A news report describes China's current marriage market situation that "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).

3.2 The econometric model

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

$$y_{ic} = \beta_c + \beta_0 B_{ic} + \beta_1 (B_{ic} \times R_c) + \alpha_c B_{ic} + \alpha_{ic} B_{ic} + \tilde{\epsilon}_{ic}. \tag{19}$$

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 investment 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.[18]

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 (19) in the error term, we have the following regression model:

$$y_{ic} = \beta_c + \bar{\alpha} B_{ic} + \beta_1 (B_{ic} \times R_c) + \hat{\epsilon}_{ic}, \tag{20}$$

where

$$\hat{\epsilon}_{ic} = \alpha_c B_{ic} + \alpha_{ic} B_{ic} + \tilde{\epsilon}_{ic}. \tag{21}$$

However, an OLS regression of equation (20) leads to a biased estimate of β_1. While B_{ic} is plausibly exogenous (as discussed in Section 3.1), R_c depends upon α_c, the county-level son-preference parameter. We model this as follows:

$$R_c = \gamma_0 + \gamma_1 \alpha_c + \gamma_2 Z_c + \eta_c, \tag{22}$$

where γ_1 measures how the 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

[18] In our actual estimations, a vector of other influences on the outcomes are included as baseline controls, which are assumed to differentially affect the outcomes for boys relative to girls: the ages of the parents, their years of schooling, their political affiliation, household location (rural or urban), ethnicity, the age of the first child, and county-level age structure for both genders.

(through equation 22) and is a component of the error term (equation 21). 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 the results, since it is uncorrelated with the county sex ratio, and since B_{ic} is exogenous.

The bias problem in the OLS estimation can be overcome 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 empirical specification:

$$y_{ic} = \beta_c + \bar{\alpha} B_{ic} + \beta_1 (B_{ic} \times R_c) + \beta_2 (B_{ic} \times X_c) + \epsilon_{ic}. \tag{23}$$

The validity of this specification relies, of course, 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 historical fertility. Prior to the introduction of B-scan ultrasonography and sex-selective abortions in the late 1980s, parents satisfy their preference for sons by giving more births; so historically, the degree of son preference was positively correlated with fertility (Ebenstein, 2011). We therefore use county-level historical fertility based on the 1982 census as proxies for unobserved son preference.[19] In robustness checks, we also use geographical region dummies (we consider three regions: east, middle, and west China) and traditional agricultural practice measured by latitude and longitude dummies (Alesina et al., 2018) to capture the local degree of son preference; 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 the county-level sex ratio that is uncorrelated with α_c, which we turn to next.

3.3 Potential endogeneity of sex ratios: A new instrument

We construct an instrument for the county-level sex ratio to address the endogeneity concern in equation (20): the average penalty for healthcare workers who illegally practiced gender screening and gender selection, relative to average household income.[20]

To motivate this instrument, the following background is useful. Since the late 1980s, sex selection in China has been accomplished mainly via B-scan ultrasonography of the fetus, to

[19] We use two variables for fertility in 1981: 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.

[20] We collect data on penalties for healthcare workers who practiced gender selections for nonmedical purposes 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.

determine its gender, followed by sex-selective abortions (Chen et al., 2013; Chu, 2001). As a result, China witnessed a surge in the sex ratio at birth in the 1980s (Chu, 2001). In 1989, the Chinese government outlawed prenatal gender disclosure for nonmedical purposes and legislated, permitting substantial penalties for healthcare workers violating the rule. These penalties were determined at local (county) level, and generate variation in the supply of sex-selection services. Using the penalty ratio as an instrument enables us to explore the quasi-experimental variation in sex ratios induced by the supply of sex-selection services.

Columns 1–3 of Table 1 show that the penalty ratio significantly reduces the local sex ratio—higher penalties lead to less biased sex ratios. This instrument for sex ratio is novel, and has not been used before. While the 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 families—as an instrument. These penalties affect the *demand* for sex-selection services, since parents who are forced to have fewer children are more likely to select for boys. In contrast, our instrument is supply-side, affecting the *supply* of sex-selection services, without directly affecting fertility and other decisions made by parents. The exclusion restriction is thus more plausibly satisfied. It is reassuring that both instruments give qualitatively similar results, and as we show later, that the instrumental variable (IV) results are also consistent with the OLS results.

The merit of the novel instrument also lies in the fact that, as a variable from the supply side of gender selections, it is less likely to be correlated with parental son preference. Columns 4 and 5 of Table 1 show that the penalty variable is not significantly correlated with historical fertility (the most prominent proxies for son preference as we discuss above), suggesting that our instrument is not largely correlated with parental son preference. The negative—rather than positive—correlation between the level of penalties and the sex ratio shown in columns 1–3, is also reassuring.

In estimating equation (20) by IV, the interaction of the sex ratio and the first-son dummy is instrumented by the interaction of the penalty ratio and the first-son dummy. The first-stage statistics, including t-statistics (about -5.5, Table 1) and robust F-statistics (47.1 to 82.4, Tables 3 and 4), indicate a strong predictive power of the instrument.

3.4 Data and variables

Sex ratios at the county level are obtained from the 2010 population census. In China, each county can be treated as a local marriage market, since the hukou system presents a formidable obstacle to marriage migration (Davin, 2005; Wei and Zhang, 2011). The census shows that more than 90 percent of rural residents and 62 percent of urban residents live in their county of birth and 89 percent of couples are from the same county. Of migrant couples in cities, 82 percent are from the same place, suggesting that migrants often get married before leaving their hometown. We use sex ratios for the age range 10–24.[21]

Other data are 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 645 communities in 25 (out of 34) provinces, representing 95 percent of the Chinese population (Xie, 2012).[22] It contains datasets with comprehensive information at several levels: individual (both adult and child), family, and community. The family-level dataset contains details of family activities and household characteristics such as migration, expenditures, investments, income, and wealth. Detailed information on the age, gender, years of education, occupation, and working location of family members is available in the individual-level dataset. The community-level dataset offers regional demographic and socioeconomic information. The datasets are linked across different levels by a set of identification numbers, and the household identification number allows us to group individuals by living unit. Parent-child relationships are precisely identified. Outcomes of interest are thus readily linked with potential covariates, enabling a systematic empirical analysis.

Our empirical analysis is based on a cross-sectional sample of households drawn from the 2010 nationwide CFPS baseline survey, which has the most comprehensive information on household activities. Thus we rely on cross-county variations in the sex ratio.[23] Specifically, we extract a sample of households from the family dataset in which the first-born child was 0–15 years old, both parents were alive and at most 50 years old, and at least one parent participated in the adult survey.[24] Our sample contains 4,304 families.

[21]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).

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

[23]The absence of a temporal dimension is not a serious deficiency since there is less variation in sex ratios across the survey waves within a county. Specifically, as we use sex ratios for the age range 10–24 in the survey year, and the CFPS surveys are biennial, any variation in the temporal dimension comes from only two out of 15 birth years across two waves.

[24]We restrict attention to families in which the eldest child was under the age of 15 to minimize the possibility that the children have started work or participated actively in household decision-making. We impose a constraint on the ages of parents to minimize the probability of their retirement or their ineffectiveness in making investment decisions due to, for example, health reasons. By placing age limits on both parents and

Outcome variables of our interest are: (i) parental labor supply and (ii) premarital investments including housing and child educational investment. Our interest in labor supply follows since the main way in which parents may increase total investments is by earning more. As discussed in Section 3.1, 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 2 shows, about 9.8 percent of fathers and 2.5 percent of mothers in our 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.[25]

To measure housing investment, we construct three variables: housing construction area, an ownership dummy, and mortgage debts. The ownership dummy indicates whether the family owns a house, including a self-constructed one. Housing construction area measures total living space, and mortgage debt measures the financing of housing investment. Table 2 shows that 83.2 percent of the families in our sample own a house; the mean construction area is 126.2 square meters; the average mortgage is 5.4 thousand yuan. 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.[26] 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 mean of which is 0.04.

3.5 Results

Table 3 reports results on parental labor supply. In this and the following tables, estimations are weighted by the CFPS survey sampling weights; standard errors are clustered at the county level.[27] Panel A reports OLS results with baseline controls. Panel B additionally controls for interactions between county-level historical fertility and the first-son dummy. Panel C reports IV results. Our discussion below is mainly based on panel A, because the

children, we maximize comparability across families.

[25]The 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 parental working hours data is not correlated with migration status.

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

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

results experience no large changes when adding proxy variables to address cross-county heterogeneity in son preference, and are broadly consistent with the IV results.[28]

In column 1 where 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), 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.

We obtain qualitatively similar findings for mother's migration and the migration of at least one parent, in columns 2 and 3. In addition, working time also increases with a rise in the sex ratio for parents with first sons relative to parents with first daughters, as shown in columns 4 and 5. Therefore, high sex ratios boost parental labor supply in families with first-born sons. Since migration is the main way for Chinese parents to increase earnings and investments, this suggests that total parental investments in sons rise when their marriage market prospects worsen.

In Table 4, the first three columns 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). 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 home owning parents with sons by 4.8 percent relative to those with daughters. Columns 2 and 3 reveal similar patterns, of 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 educational investment. 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), 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

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

estimate is 3.8 percentage points or 12.8 percent.[29] In column 6, we check the effect of sex ratio on the ratio of educational investment to housing investment in a family, and find a significantly negative effect.

As noted at the end of the theoretical model (see Section 2.2, page 17), variations in son preference cause both dimensions of investments to move together, whereas variations in the sex ratio lead to *contradictory* effects, by *reducing* educational investment even as they raise housing investment. Thus, if the OLS results are biased by the positive correlation between son preference and sex ratio, this would attenuate the estimated effects on educational investment.

Appendix Tables A3 and A4 present robustness results using alternative son-preference proxies. Appendix Tables A5 and A6 present heterogenous effects across rural and urban households, showing that the results are both supportive of the model implications (see Appendix D for details).

Parental migration and educational investment have important implications for child development. In particular, the adverse effect of parental absences on children is well documented, in China and in other countries (e.g., Zhang et al., 2014; Lyle, 2006). To study these effects, we measure child human capital development in terms of cognitive, non-cognitive, and health outcomes for the first-born (cognitive and non-cognitive outcomes are for those who are at least ten years old). Cognitive outcomes are based on class rankings in mathematics and Chinese examinations, measured by a child's position in the cumulative distribution function (closer to one means closer to the top). Non-cognitive outcomes are indicator variables for interpersonal communication skills including openness and cooperation, a value of one implying a better result.[30] Health outcomes are z-scores of weight and height transformed based on international child growth standards, a larger value implying a more satisfactory result.

In Appendix Table A7, the negative estimates for the interaction-term coefficient in the first four columns 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. Columns 5 and 6 show adverse effects on health outcomes,

[29] The estimates for child educational investment might be biased if investment measures for children with migrant parents have the issue of 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 regressions of education expenditure. Results are reported in columns 2–4 of Appendix Table A2, with column 1 being the baseline. We also use a subsample that excludes families with any migrant parents in column 5. Results are robust in all these additional checks.

[30] After interviewing each child, interviewers were asked to evaluate communication skills of the child, which were recorded from one to seven, a larger number meaning a better performance. The corresponding indicator variables are one if the record was five, six, or seven.

although these are not always statistically significant. The results indicate that as a result of the sex-ratio imbalance, increased parental migration and reduced educational investment in families with a first boy undermine the human capital development of the boy.

3.6 Endogenous fertility

It is well known that the gender of the first child influences family size; 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, then 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. So the question is, do our results—specifically, the coefficient β_1 on $B_{ic} \times R_c$—reflect in part this effect?

To address this question, we consider a model of fertility, f_{ic}.[31] We simplify the analysis by treating fertility as a continuous rather than discrete variable:

$$f_{ic} = \tilde{\delta}_c + \delta_1(\bar{\alpha} + \alpha_c + \alpha_{ic})(1 - B_{ic}) + \tilde{\omega}_{ic}, \tag{24}$$

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 (24) can be re-written as:

$$f_{ic} = \delta_c - \delta_1(\bar{\alpha} + \alpha_c + \alpha_{ic})B_{ic} + \omega_{ic}, \tag{25}$$

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

Since the household chooses investments conditional on the level of fertility, the behavioral equation (19) for investments is modified by adding a term, $\beta_3 f_{ic}$, to the right-hand side. Therefore, we have:

$$y_{ic} = \beta'_c + \bar{\alpha}' B_{ic} + \beta_1(B_{ic} \times R_c) + \hat{\epsilon}'_{ic}, \tag{26}$$

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

$$\hat{\epsilon}'_{ic} = (1 - \delta_1\beta_3)\alpha_c B_{ic} + (1 - \delta_1\beta_3)\alpha_{ic} B_{ic} + \tilde{\epsilon}_{ic} + \beta_3\omega_{ic}. \tag{27}$$

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

We see that modelling fertility choice affects the coefficient on B_{ic}—mean investment differences between boys and girls. The error term (equation 27) has the same structure as that in the original specification, which does not explicitly model fertility and its implications (equation 21). 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 it, this resolves also the problem for the case where fertility choice is modelled.

More intuitively, changes in fertility affect 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 (25). 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, they do not take into account future marriage prospects 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 in 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 two-fold. First, the county-level sex ratio displays 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).

3.7 Conditioning on fertility

The above argument is the basis of our baseline specification where we do not control for fertility. Nonetheless, one might ask, what are the implications of controlling for family size in our estimations? Since family size is endogenous, and depends both upon first-child gender (B_{ic}) and the household's son preference ($\alpha_c + \alpha_{ic}$ in the error), would this induce a bias in the estimation of our coefficient of interest (β_1 on $B_{ic} \times R_c$)? We now present two different ways of addressing this question.

3.7.1 One-child families

First, we restrict our sample to families with only one child, so that investment variables (total investments, housing investment, and educational investment) pertain to the single child. If parents with a son are happy to stop (or required to do so by the one-child policy), but parents with a daughter choose to have a second child depending upon their intensity of son preference, then 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, (24), 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:

$$\delta_1(\bar{\alpha} + \alpha_c + \alpha_{ic}) + \tilde{\omega}_{ic} \leq \bar{f}, \tag{28}$$

and families with one boy satisfy:

$$\tilde{\omega}_{ic} \leq \bar{f}. \tag{29}$$

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:

$$\alpha_c = \theta_1 B_{ic} + \eta_c, \tag{30}$$

$$\alpha_{ic} = \theta_2 B_{ic} + \eta_{ic}, \tag{31}$$

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, (19), gives us:

$$y_{ic} = \beta_c + (\bar{\alpha} + \theta_1 + \theta_2)B_{ic} + \beta_1(B_{ic} \times R_c) + \hat{\epsilon}_{ic}, \tag{32}$$

where

$$\hat{\epsilon}_{ic} = \eta_c B_{ic} + \eta_{ic} B_{ic} + \tilde{\epsilon}_{ic}. \tag{33}$$

Therefore, the estimated coefficient on $B_{ic} \times R_c$ is consistent under the same conditions as before—i.e. we are able to control for the county-level son preference or instrument for the county-level sex ratio, while 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. In Tables 5 and 6 where we restrict our sample to families with one child (about 66 percent of the sample), the estimated coefficients on $B_{ic} \times R_c$ are broadly consistent with those in our main estimations.

3.7.2 Controlling for fertility in the unrestricted sample

In the unrestricted sample where some families have multiple children, we include family size as one of the regressors. The theoretical model of Section 2 can be extended for analyzing the investment choices of families with many children. The most important conclusion from such an extension is that the *composition* of investments exhibits a type of separability—if we condition upon total amount of resources allocated to a child, its division between human capital and housing does not depend upon the number of other children or upon their investments. More importantly, it leads to the following empirical specification for aggregate investments by the household:

$$\check{y}_{ic} = \beta_c + \bar{\alpha}\check{B}_{ic} + \beta_1(\check{B}_{ic} \times R_c) + \varphi_0 f_{ic} + \varphi_1(\check{B}_{ic} \times f_{ic}) + \check{\epsilon}_{ic}, \tag{34}$$

where $\check{y}_{ic}$ is household-specific investments, and $\check{B}_{ic}$ is the *proportion* of boys amongst the children of the household.[32] We are interested in the coefficient on $\check{B}_{ic} \times R_c$. The endogeneity concern related to the county sex ratio, R_c, can be addressed using strategies discussed before.

The new problem here is that both fertility and the proportion of boys are endogenous variables. In particular, both are correlated with the son-preference parameters α_c and α_{ic}, which enter the error term. While one may use the first-son dummy as an instrument, together with the penalty variable as an instrument for the county-level sex ratio, one needs a third instrument. We use the number of brothers the father of the household has as an additional instrument for $\check{B}_{ic}$ and f_{ic}, as in Fan et al. (2018). If the father has many brothers, he faces less pressure to have a son, since the sons of his brothers can also continue the family name, resulting in lower fertility and fewer sons for such a father (see Appendix Table A8). In addition, the number of brothers the father has does not directly affect family investment decisions, since married brothers usually have independent financial decisions, and do not bear responsibility for one another.[33]

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 the 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

[32] Since we are conditioning on fertility, $\check{B}_{ic}$ could also be taken to be the number of boys in the household.

[33] Another candidate for the third 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.

the number of brothers the father has—are affected by the marriage prospects of children. Results are reported in Appendix Tables A9 and A10, and are broadly consistent with the results in our main estimations.

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 intensity of the one-child policy, does so. To the extent that there might be residual concerns, it is reassuring that our results are robust to conditioning on fertility.

4 Conclusion

This paper studies how imperfect commitment in marriage affects premarital investments in children made by their parents. If boys are in excess supply in the marriage market but unable to commit to future resource sharing from their labor income, their parents may be induced to invest more in housing, as a way of overcoming the commitment problem. This may lead to a crowding out of human capital investment in boys. Using nationally representative Chinese data, we find that high sex ratios lead to increased parental migration, increased housing investment, and reduced educational investment for families with sons relative to families with daughters. Our empirical work controls for unobserved county-level heterogeneity and compares the effect of sex-ratio imbalance for first-son families with the effect for first-daughter families, at the county level. We also provide a variety of robustness analyses to address the concerns about potential endogeneity of sex ratios and other concerns.

This paper has important implications for China's future human capital development. Underinvestment in education and increased parental migration may combine to result in adverse development consequences for boys. The welfare consequences are likely to be large, given the large returns to education in China (Heckman, 2003, 2005). Finally, our paper is among the few that consider multidimensional investments, and our key finding is that the different components of investments may respond differently to marriage market competition.

References

Alesina, A., P. Giuliano, and N. Nunn (2018). Traditional agricultural practices and the sex ratio today. *PloS one 13*(1), e0190510.

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

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

Becker, G. S. (1991). *A Treatise on the Family.* Harvard University Press.

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

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

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

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

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

Chai, K. (2020). How marriage affects chinese cities. *PEAKUrban*. June 16. https://www.peak-urban.org/blog/how-marriage-affects-chinese-cities.

Chen, Y., H. Li, and L. Meng (2013). Prenatal sex selection and missing girls in china: Evidence from the diffusion of diagnostic ultrasound. *Journal of Human Resources 48*(1), 36–70.

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

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

Chiappori, P.-A., B. Salanié, and Y. Weiss (2017). Partner choice, investment in children, and the marital college premium. *American Economic Review 107*(8), 2109–67.

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

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

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

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

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

Dizdar, D. (2018). Two-sided investment and matching with multidimensional cost types and attributes. *American Economic Journal: Microeconomics 10*(3), 86–123.

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

Doepke, M. and M. Tertilt (2016). Asymmetric information in couples. *Unpublished mimeo*.

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

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

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

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

Fan, Y., J. Yi, Y. Yuan, and J. Zhang (2018). The glorified mothers of sons: Evidence from child sex composition and parental time allocation in rural china. *Journal of Economic Behavior & Organization 145*, 249–260.

Galichon, A., S. D. Kominers, and S. Weber (2019). Costly concessions: An empirical framework for matching with imperfectly transferable utility. *Journal of Political Economy 127*(6), 2875–2925.

Gobbi, P. E. (2018). Childcare and commitment within households. *Journal of Economic Theory 176*, 503–551.

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

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

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

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

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

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

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

Li, W., C. Song, S. Xu, and J. Yi (2020). High sex ratios and household portfolio choice in china. *Journal of Human Resources*, 1217–9245.

Lin, W., K. McEvilly, and J. Pantano (2016). Obesity and sex ratios in the us. *Review of Economics of the Household 14*(2), 269–292.

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

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

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

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

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

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

Nash, J. F. (1950). The bargaining problem. *Econometrica 18*, 155–162.

Nunn, N. (2005). A model explaining simultaneous payments of dowry and brideprice. *Unpublished manuscript*.

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

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

Rao, V. (1993). The rising price of husbands: A hedonic analysis of dowry increases in rural india. *Journal of Political Economy*, 666–677.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica 50*(1), 97–109.

Scharping, T. (2013). *Birth Control in China 1949-2000: Population policy and demographic development*. Routledge.

Shaked, A. and J. Sutton (1984, November). Involuntary unemployment as a perfect equilibrium in a bargaining model. *Econometrica 52*(6), 1351–1364.

Shaked, M. and J. G. Shanthikumar (2007). *Stochastic orders.* Springer Science & Business Media.

Shapley, L. S. and M. Shubik (1971). The assignment game I: The core. *International Journal of Game Theory 1*(1), 111–130.

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

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

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

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

Xing, Z. (2017a). The evolution of chinese rural families, centralization or individualization? moral evaluation of private money as an assess point (in chinese). *Society 37*(5), 165–192.

Xing, Z. (2017b). 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) 39*(2), 99–110.

Zhang, H. (2021). An investment-and-marriage model with differential fecundity. *Journal of Political Economy (forthcoming)*.

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

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

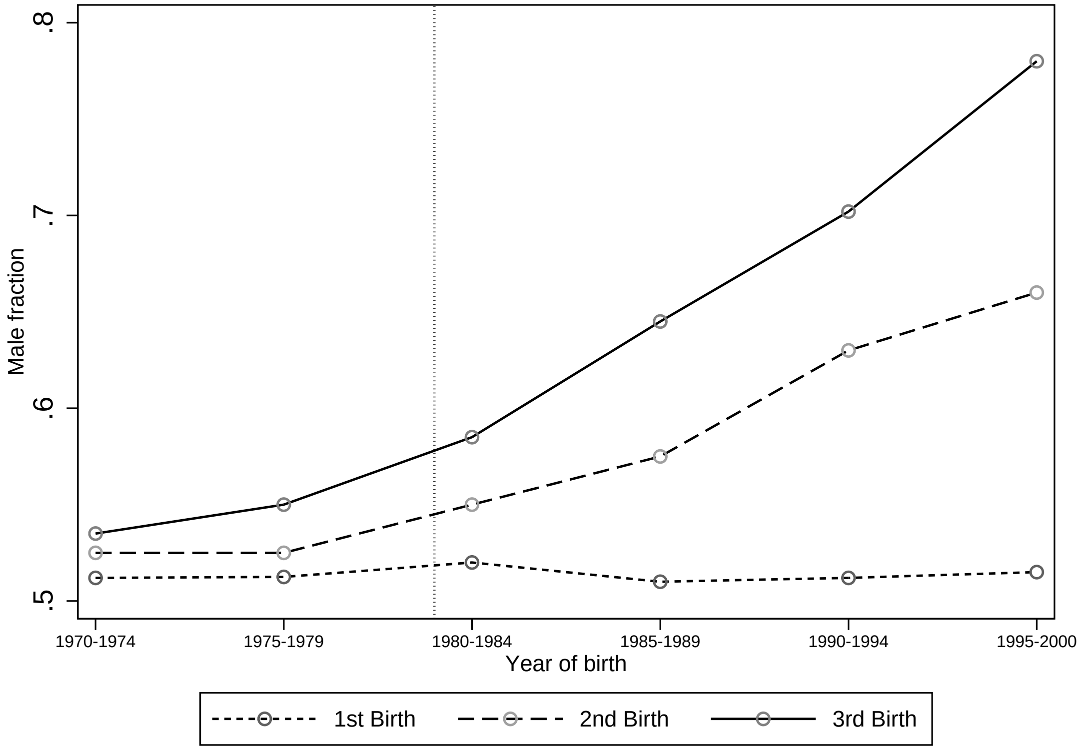


Figure 1 Male fraction of births by birth order in China

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.

Table 1 County-level sex ratio, penalty ratio, and historical fertility

Dependent variable	2010 Sex ratio			1990s Penalty ratio	
	(1)	(2)	(3)	(4)	(5)
1990s Penalty ratio	-0.182***	-0.175***	-0.150***		
	(0.033)	(0.032)	(0.028)		
1981 Completed fertility				-0.026	
				(0.076)	
1981 Birth rate					-0.001
					(0.002)
T-statistic	-5.5	-5.5	-5.3	-0.3	-0.9
# 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: "Penalty ratio" 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 county-level sex ratio in 2010 on penalty ratio in the 1990s; column 1 includes only penalty ratio; column 2 further adds other county-level controls, including average age, average schooling years, the proportions in different industries, and average wage; column 3 further adds province fixed effects. Columns 4 and 5 are the regression results of county-level penalty ratio in the 1990s on fertility measures in 1981; "Completed fertility" is the average fertility of women aged 35–64 in the county; "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 the proportions in different industries. Standard errors given in parentheses are heteroskedasticity-robust.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 2 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 2010 CFPS survey. The sample includes all households from the family dataset in which the first-born child was 0–15 years old, 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 CFPS survey sampling weights.

Table 3 Main results: 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: Historical fertility as proxies for son preference					
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 * Proxies	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 between county-level historical fertility—"1981 Completed fertility" and "1981 Birth rate" as defined in notes to Table 1—and 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 * Penalty ratio," where "Penalty ratio" is defined in notes to Table 1. Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 4 Main results: Premarital 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**	18.724*	-1.775**	-0.382**	-0.187**
	(0.198)	(0.113)	(10.428)	(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: Historical fertility as proxies for son preference						
First son * Sex ratio	0.452**	0.242*	22.032**	-1.256*	-0.413**	-0.197**
	(0.199)	(0.124)	(10.467)	(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 * Proxies	YES	YES	YES	YES	YES	YES
C: IV results						
First son * Sex ratio	0.590**	0.280*	25.493*	-1.845*	-0.743***	-0.258***
	(0.274)	(0.152)	(14.305)	(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 between county-level historical fertility—"1981 Completed fertility" and "1981 Birth rate" as defined in notes to Table 1—and 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 * Penalty ratio," where "Penalty ratio" is defined in notes to Table 1. Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 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: Historical fertility as proxies for son preference					
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 * Proxies	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: Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table 6 One-child families: Premarital 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: Historical fertility as proxies for son preference						
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 * Proxies	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: Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Online Appendices

A Proof of Proposition 1

Consider the best response on the boys' side to a profile of investments on the girls' side, (x_G, y_G). Since the first order conditions for the boys' investments are unaffected by x_G, we may write this as a pair of functions $\hat{x}_B(y_G)$ and $\hat{y}_B(y_G)$. Thus, $\hat{y}_B(y_G)$ is the best response investment in the public good, when both types of investments are chosen optimally by the boy. Similarly, we may define best responses on the girls' side, $\hat{y}_G(y_B)$. Let ζ denote the composition of the functions $\hat{y}_G$ and $\hat{y}_B$ so that $\zeta(u) = \hat{y}_G(\hat{y}_B(u))$. The fixed points of ζ correspond to quasi-symmetric equilibria. More precisely, y_G is a fixed point of ζ if and only if $y_G = \hat{y}_G(\hat{y}_B(y_G))$, so that the pair $(\hat{y}_B(y_G), y_G)$ are mutual best responses, with the private good investments being given by $\hat{x}_B(\hat{y}(y_B))$ and $x_G = \hat{x}_G(y_B)$.

Observe that ζ is continuous and differentiable on the positive reals. Since $c_y(x, 0) = 0$, by assumption 1, $\zeta(0) > 0$. Also, since $v'_B(y)$ and $v'_G(y)$ are decreasing, while $c_y(x, y) \to \infty$, $\zeta(y) < y$ for y sufficiently large. Thus there exists a fixed point of ζ, which we denote by y^*_B. Let us denote this quasi-symmetric equilibrium by $((x^*_B, y^*_B), (x^*_G, y^*_G))$.

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

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

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

Similarly,

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

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

B Comparative statics in balanced marriage markets

We totally differentiate the first order conditions for optimal investment, to obtain the following system:

$$c^B_{xx}(.)\frac{dx_B}{d\lambda_B}+c^B_{xy}(.)\frac{dy_B}{d\lambda_B}=1 \quad \text{(A1)}$$

$$c^B_{xy}(.)\frac{dx_B}{d\lambda_B}+c^B_{yy}(.)\frac{dy_B}{d\lambda_B}=v^{''}_B(\hat{y})(\frac{dy_B}{d\lambda_B}+\frac{dy_G}{d\lambda_B})+\frac{1-\lambda_G}{1-\lambda_B}\theta_B v^{''}_G(\frac{dy_B}{d\lambda_B}+\frac{dy_G}{d\lambda_B})+\frac{1-\lambda_G}{(1-\lambda_B)^2}\theta_B v^{'}_G \quad \text{(A2)}$$

$$c^G_{xx}(.)\frac{dx_G}{d\lambda_B}+c^G_{xy}(.)\frac{dy_G}{d\lambda_B}=-\theta_G \quad \text{(A3)}$$

$$c^G_{xy}(.)\frac{dx_G}{d\lambda_B}+c^G_{yy}(.)\frac{dy_G}{d\lambda_B}=v^{''}_G(\hat{y})(\frac{dy_B}{d\lambda_B}+\frac{dy_G}{d\lambda_B})+\frac{1-\lambda_B}{1-\lambda_G}\theta_G v^{''}_B(\frac{dy_B}{d\lambda_B}+\frac{dy_G}{d\lambda_B})-\frac{\theta_G}{(1-\lambda_G)^2}v^{'}_B \quad \text{(A4)}$$

The equations (A1)–(A4) can be put into a matrix form as follows: $\mathbf{n}$ x $\mathbf{q}=\mathbf{a}$, where $\mathbf{n}$ is a 4×4 matrix and $\mathbf{q}=(\frac{dx_B}{d\lambda_B},\frac{dy_B}{d\lambda_B},\frac{dx_G}{d\lambda_B},\frac{dy_G}{d\lambda_B})$.

We first calculate inv($\mathbf{n}$) and factor out the determinant of $\mathbf{n}$ in order to make the final expression simple. Denote by det the determinant of $\mathbf{n}$. We will present the final answer in the form

$$\mathbf{q}=\frac{1}{det}\hat{\mathbf{q}}$$

, where $\hat{\mathbf{q}}=\mathbf{n}^{-1}a\times det$. det can be simplified in the following way. Define $H_i=(c^i_{xy})^2-c^i_{xx}c^i_{yy}$, where $i=G,B$. Note that H_i is the determinant of Hessian of the c_i function. In the simplification, we factor out H_i.

$$det=A_0+B_0\frac{1}{1-\lambda_G}+C_0\frac{1}{1-\lambda_B}, \quad \text{(A5)}$$

where

$$A_0=(H_B+c^B_{xx}v^{''}_B)(H_G+c^G_{xx}v^{''}_G)-c^B_{xx}v^{''}_B c^G_{xx}v^{''}_G$$
$$B_0=H_B c^G_{xx}v^{''}_B\theta_G(1-\lambda_B)$$
$$C_0=H_G c^B_{xx}v^{''}_G\theta_B(1-\lambda_G)$$

Next, we calculate $\hat{\mathbf{q}}\equiv(\hat{q}_1,\hat{q}_2,\hat{q}_3,\hat{q}_4)=\frac{1}{det}(\frac{dx_B}{d\lambda_B},\frac{dy_B}{d\lambda_B},\frac{dx_G}{d\lambda_B},\frac{dy_G}{d\lambda_B})$

$$\hat{q}_i=A_i+B_i\frac{1}{1-\lambda_G}+C_i\frac{1}{1-\lambda_B}+D_i\frac{1}{(1-\lambda_B)^2}, \quad \text{(A6)}$$

where $i = 1, 2, 3, 4$ and

$$
\begin{aligned}
A_1 &= (v_B^{''} - c_{yy}^B)H_G - [c_{xx}^G c_{yy}^B v_G^{''} + c_{xy}^G c_{xy}^B v_B^{''}\theta_G] \\
B_1 &= c_{xx}^G c_{xy}^B v_B^{'} v_B^{''}\theta_G - (1-\lambda_B)c_{xx}^G c_{yy}^B v_B^{''}\theta_G \\
C_1 &= (H_G - c_{xy}^G c_{xy}^B\theta_G)v_G^{''}\theta_B(1-\lambda_G) + c_{xx}^G\theta_G\theta_B(c_{xx}^B v_G^{''} + v_G^{'} v_B^{''}) \\
D_1 &= (H_G + c_{xx}^G v_G^{''})c_{xy}^B v_G^{'}\theta_B(1-\lambda_G)
\end{aligned}
$$

$$
\begin{aligned}
A_2 &= c_{xy}^B H_G + c_{xx}^G c_{xy}^B v_G^{''} + c_{xx}^B c_{xy}^G v_B^{''}\theta_G \\
B_2 &= c_{xx}^G c_{xy}^B v_B^{''}\theta_G(1-\lambda_B) - c_{xx}^B c_{xx}^G v_B^{'} v_B^{''}\theta_G \\
C_2 &= c_{xx}^B c_{xx}^G\theta_B\theta_G(v_B^{''} v_G^{'} - v_G^{''} v_B^{'}) + c_{xx}^B c_{xy}^G\theta_B\theta_G v_G^{''}(1-\lambda_G) \\
D_2 &= -v_G^{'}\theta_B(c_{xx}^B H_G + c_{xx}^B c_{xx}^G v_G^{''})
\end{aligned}
$$

$$
\begin{aligned}
A_3 &= (c_{yy}^G - v_G^{''})H_B\theta_G + [c_{xy}^G c_{xy}^B v_G^{''} + c_{yy}^G c_{xx}^B v_B^{''}\theta_G] \\
B_3 &= v_B^{''}\theta_G(1-\lambda_B)(c_{xy}^G c_{xy}^B - H_B\theta_G) - c_{xy}^G v_B^{'}\theta_G(H_B + c_{xx}^B v_B^{''}) \\
C_3 &= c_{xx}^B c_{yy}^G\theta_B\theta_G v_G^{''}(1-\lambda_G) - c_{xx}^B c_{xy}^G\theta_B\theta_G(v_B^{''} v_G^{'} + v_G^{''} v_B^{'}) \\
D_3 &= -c_{xx}^B c_{xy}^G v_G^{''} v_B^{'}\theta_B(1-\lambda_G)
\end{aligned}
$$

$$
\begin{aligned}
A_4 &= -c_{xy}^G H_B\theta_G - [c_{xx}^G c_{xy}^B v_G^{''} + c_{xy}^G c_{xx}^B v_B^{''}\theta_G] \\
B_4 &= (H_B + c_{xx}^B v_B^{''})c_{xx}^G v_B^{'}\theta_G - c_{xx}^G c_{xy}^B v_B^{''}\theta_B(1-\lambda_B) \\
C_4 &= c_{xx}^B c_{xx}^G\theta_B\theta_G(v_B^{''} v_G^{'} + v_G^{''} v_B^{'}) - c_{xx}^B c_{xy}^G\theta_B\theta_G v_G^{''}(1-\lambda_G) \\
D_4 &= c_{xx}^B c_{xx}^G v_G^{''} v_G^{'}\theta_B(1-\lambda_G)
\end{aligned}
$$

B.1 Numerical results

We now present a numerical example, which shows the parameter range where $\frac{dx_B}{d\lambda_B} < 0$. We assume the following functional forms and parameter values.

- $c(x,y) = \frac{\gamma}{2}(x+y)^2 + \frac{1-\gamma}{2}(x^2+y^2)$
- $v_G(y = v_B(y) = ay - by^2$
- $\lambda_G = 0.5$
- a = 1, b = 0.1
- $\gamma = \{0.2, 0.5\}$

Each graph in Appendix Figure A1 represents the level of equilibrium investments as a function of $\lambda_B \in [0.1, 0.9]$. Note that we have interior solutions for all $\lambda_B \in [0.1, 0.9]$. Graphs on the left column are set for $\gamma = 0.5$ and those on the right for $\gamma = 0.2$. Each row is for different values of θ_B, θ_G. It is easy to see that x_B decreases in λ_B after some point of λ_B. The table below summarizes those critical values of λ_B after which $\frac{dx_B}{d\lambda_B} < 0$.

	$\gamma = 0.5$	$\gamma = 0.2$
$(\theta_B = 1, \theta_G = 1)$	0.642	0.81
$(\theta_B = 2, \theta_G = 0.5)$	0.535	0.773

For example, when $\theta_B = 1, \theta_G = 1, \gamma = 0.5$, $\frac{dx_B}{d\lambda_B} < 0$ for $\lambda_B \in (0.642, 0.9]$. Note that the result is for $\lambda_B \in [0.1, 0.9]$. The range is chosen to guarantee interior solutions.

Appendix Figure A2 shows the equilibrium investments as a function of θ_B while fixing $\theta_B\theta_G$ to 1 with $\lambda_B = 0.9, \lambda_G = 0.5$ and $\gamma = 0.5$.

Note that we have interior solutions for all $\theta_B \in [1, 3]$. We can see x_B is decreasing in θ_B for some value of θ_B. Solving analytically for $\frac{dx_B}{d\theta_B}$ gives us that $\frac{dx_B}{d\theta_B} < 0$ for $\theta_B \in [1, 1.96049]$ and increasing otherwise.

C Comparative statics in unbalanced marriage markets

We totally differentiate the first-order conditions for equilibrium investments, equations (13), (14), (15), and (16), with respect to r, to obtain the following system of four equations with four unknowns. The second derivative of each function is evaluated at the equilibrium values. To simplify the notation, we denote the utility derived from public good investment of a boy when he remains single by $m_B = v_B(y_B^*)$.

$$
\begin{aligned}
c_{xx}^B \frac{dx_B}{dr} + c_{xy}^B \frac{dy_B}{dr} &= \rho^{'}(r)[(1-\hat{r})(1-\lambda_B) + (1-\lambda_G)(\hat{r}\theta_{B+} - \theta_B) + \hat{f}\bar{U}], \\
c_{xy}^B \frac{dx_B}{dr} + c_{yy}^B \frac{dy_B}{dr} &= (v_B^{''} + \frac{1-\lambda_G}{1-\lambda_B} v_G^{''}\theta_B)(\frac{dy_B}{dr} + \frac{dy_G}{dr}) \\
&\quad + \rho^{'}(r)[(1-\hat{r})[m_B^{'} - v_B^{'}] + \frac{1-\lambda_G}{1-\lambda_B} v_G^{'}(\hat{r}\theta_{B+} - \theta_B) + \frac{v_G^{'}}{1-\lambda_B}\hat{f}\bar{U}] \\
&\quad + \rho(r)[(1-\hat{r})[m_B^{''}\frac{dy_B}{dr} - v_B^{''}(\frac{dy_B}{dr} + \frac{dy_G}{dr})] \\
&\quad + (\frac{1-\lambda_G}{1-\lambda_B} v_G^{''}(\hat{r}\theta_{B+} - \theta_B) + \frac{v_G^{''}}{1-\lambda_B}\hat{f}\bar{U})(\frac{dy_B}{dr} + \frac{dy_G}{dr})], \\
c_{xx}^G \frac{dx_G}{dr} + c_{xy}^G \frac{dy_G}{dr} &= (1-\lambda_B)\rho^{'}(r)(\theta_{G+} - \theta_G), \\
c_{xy}^G \frac{dx_G}{dr} + c_{yy}^G \frac{dy_G}{dr} &= v_G^{''}(\frac{dy_B}{dr} + \frac{dy_G}{dr}) \\
&\quad + \frac{1-\lambda_B}{1-\lambda_G}[v_B^{''}(\frac{dy_B}{dr} + \frac{dy_G}{dr})[\rho(r)\theta_{G+} + (1-\rho(r)\theta_G)] + v_B^{'}\rho^{'}(r)(\theta_{G+} - \theta_G)].
\end{aligned}
\tag{A7}
$$

The above equations can be written in a matrix form as follows:

$$
\mathbf{N}\mathbf{q} = \rho'(r)\mathbf{a}, \tag{A8}
$$

where $\mathbf{N}$ is a 4×4 matrix and $\mathbf{q} = (\frac{dx_B}{dr}, \frac{dy_B}{dr}, \frac{dx_G}{dr}, \frac{dy_G}{dr})$. We write $\mathbf{q}$ in the form

$$
\mathbf{q} = \rho'(r)\mathbf{N}^{-1}\mathbf{a} \equiv \frac{\rho'(r)}{det(N)}\hat{\mathbf{q}}\mathbf{a}, \tag{A9}
$$

where $\hat{\mathbf{q}} := det(N)\mathbf{N^{-1}}$. So

$$
det(\mathbf{N}) = A_0 + B_0\frac{1-\lambda_G}{1-\lambda_B} + C_0\frac{1-\lambda_B}{1-\lambda_G}, \tag{A10}
$$

where

$$
\begin{aligned}
A_0 &= H_B H_G + c_{xx}^B H_G[\rho(1-\hat{r})(m_B^{''} - v_B^{''}) + v_B^{''}] \\
&+[c_{xx}^G H_B + \rho c_{xx}^B c_{xx}^G (1-\hat{r}) m_B^{''}] v_G^{''}, \\
B_0 &= c_{xx}^B H_G v_G^{''}[(1-\rho)\theta_B - \rho(\hat{f}\bar{U} + \hat{r}\theta_{B+})], \\
C_0 &= c_{xx}^G [H_B + c_{xx}^B \rho m_B^{''}(1-\hat{r})] v_B^{''}[(1-\rho)\theta_G + \rho\theta_{G+}].
\end{aligned} \tag{A11}
$$

The expression for $\hat{q}$ takes the following form:

$$
\begin{aligned}
\hat{q}_1 = & \frac{1}{(1-\lambda_B)(1-\lambda_G)}[c_{xx}^G c_{xy}^B v_B^{'}(\theta_G - \theta_{G+})(1-\lambda_B)\hat{q}_{11} - [c_{xx}^G \hat{q}_{12} + (c_{xy}^G)^2(1-\lambda_G)\hat{q}_{13}]\hat{q}_{14} \\
& + c_{xy}^B \hat{q}_{15}\hat{q}_{16} c_{xy}^B c_{xy}^G (\theta_G - \theta_{G+})(1-\lambda_B)(1-\lambda_G)\hat{q}_{17}],
\end{aligned} \tag{A12}
$$

where

$$
\begin{aligned}
\hat{q}_{11} &= \hat{f}\bar{U}\rho v_G^{''} + (1-\lambda_B)v_{BB}^{''}((1-\rho) + \rho\hat{r}) + (1-\lambda_G)v_G^{''}((1-\rho)\theta_B + \rho\hat{r}\theta_{B+}). \\
\hat{q}_{12} &= (1-\lambda_B)^2(c_{yy}^B - \rho m_B^{''}(1-\hat{r}))v_B^{''}((1-\rho)\theta_G + \rho\theta_{G+}). \\
&+(1-\lambda_G)[\hat{f}\bar{U}\rho c_{yy}^G v_G^{''} + (1-\lambda_B). \\
&\quad [c_{yy}^G \rho(1-\hat{r})(m_B^{''} - v_B^{''}) - \rho m_B^{''}(1-\hat{r})v_G^{''} + c_{yy}^G(v_B^{''} - c_{yy}^B) + c_{yy}^B v_G^{''}]] \\
&+(1-\lambda_G)^2 c_{yy}^G v_G^{''}((1-\rho)\theta_B + \rho\hat{r}\theta_{B+}) \\
\hat{q}_{13} &= -\hat{f}\bar{U}v_G^{''} + (1-\lambda_B)[c_{yy}^B - \rho(1-\hat{r})(m_B^{''} - v_B^{''})] \\
&-(1-\lambda_G)v''_G[(1-\rho)\theta_B + \rho\hat{r}\theta_{B+}]. \\
\hat{q}_{14} &= \hat{f}\bar{U} + (1-\lambda_B)(1-\hat{r}) + (1-\lambda_G)(\hat{r}\theta_{B+} - \theta_B). \\
\hat{q}_{15} &= (1-\lambda_B)c_{xx}^G v_B^{''}[(1-\rho)\theta_G + \rho\theta_{G+}] + (1-\lambda_G)[H_G + c_{xx}^B v_G^{''}]. \\
\hat{q}_{16} &= \hat{f}\bar{U}v_G^{'} + (1-\lambda_B)(1-\hat{r})(m_B^{'} - v_B^{'}) + (1-\lambda_G)v_G^{'}(\hat{r}\theta_{B+} - \theta_B). \\
\hat{q}_{17} &= \hat{f}\bar{U}v_G^{''} + (1-\lambda_B)v_B^{''}[(1-\rho) + \rho\hat{r}] + (1-\lambda_G)v_G^{''}[(1-\rho)\theta_B + \rho\hat{r}\theta_{B+}].
\end{aligned} \tag{A13}
$$

$$
\hat{q}_2 = c_{xx}^B c_{xy}^G(\theta_G - \theta_{G+})\hat{q}_{21}[1 - \frac{v_G^{'}}{1-\lambda_G}] + \hat{q}_{22}[\frac{c_{xx}^B}{(1-\lambda_B)(1-\lambda_G)}\hat{q}_{23} - \frac{c_{xy}^B}{1-\lambda_G}\hat{q}_{24}], \tag{A14}
$$

where

$$
\begin{aligned}
\hat{q}_{21} &= \hat{f}\bar{U}v_G^{''} + (1-\lambda_G)v_G^{''}[(1-\rho)\theta_B + \rho\hat{r}\theta_{B+}] + (1-\lambda_B)v_B^{''}[(1-\rho) + \rho\hat{r}. \\
\hat{q}_{22} &= (1-\lambda_G)[c_{xx}^G v_G^{''} - H_G] - (1-\lambda_B)c_{xx}^G v_B^{''}[(1-\rho)\theta_G + \rho\theta_{G+}]. \\
\hat{q}_{23} &= \hat{f}\bar{U}v_G^{'} + (1-\lambda_G)v_G^{''}(\hat{r}\theta_{B+} - \theta_B) + (1-\lambda_B)(1-\hat{r})(m_B^{'} - v_B^{'}). \\
\hat{q}_{24} &= \hat{f}\bar{U} + (1-\lambda_G)(\hat{r}\theta_{B+} - \theta_B) + (1-\lambda_B)(1-\hat{r}).
\end{aligned}
\tag{A15}
$$

$$
\begin{aligned}
\hat{q}_3 = &\frac{1}{(1-\lambda_B)(1-\lambda_G)}[(\theta_G - \theta_{G+})(1-\lambda_b)\hat{q}_{31} + c_{xy}^G v_B^{'}(\theta_G - \theta_{G+})\hat{q}_{32} \\
&-c_{xy}^G \hat{q}_{33}[c_{xy}^B(1-\lambda_B)\hat{q}_{34} + c_{xx}^B\hat{q}_{35}],
\end{aligned}
\tag{A16}
$$

where

$$
\begin{aligned}
\hat{q}_{31} &= (1-\lambda_B)^2[H_B + c_{xx}^B\rho m_B^{''}(1-\hat{r})]v_B^{''}[(1-\rho)\theta_G + \rho\theta_{G+}] \\
&\quad -(1-\lambda_G)(\hat{f}\bar{U}\rho c_{xx}^B c_{yy}^G v_G^{''} - (1-\lambda_B)[H_B(v_G^{''} - c_{yy}^G) \\
&\quad +\rho m_B^{''}(1-\hat{r})(v_G^{''} - c_{yy}^G) - c_{yy}^G((1-\rho) + \rho\hat{r}v_B^{''})] \\
&\quad -(1-\lambda_G)^2[c_{xx}^B c_{yy}^G v_G^{''})((1-\rho)\theta_B + \rho\hat{r}\theta_{B+})]. \\
\hat{q}_{32} &= \hat{f}\bar{U}c_{xx}^B v_G^{''} \\
&\quad +(1-\lambda_B)[(c_{xy}^B)^2 + c_{xx}^B(v_B^{''} - c_{yy}^B) + \rho c_{xx}^B(1-\hat{r})(m_B^{''} - v_B^{''})] \\
&\quad +(1-\lambda_G)c_{xx}^B v_G^{''}((1-\rho)\theta_B + \rho\hat{r}\theta_{B+}). \\
\hat{q}_{33} &= (1-\lambda_B)v_B^{''}((1-\rho)\theta_G + \rho\theta_{G+}) + (1-\lambda_G)v_G^{''}. \\
\hat{q}_{34} &= \hat{f}\bar{U} + (1-\lambda_B)(1-\hat{r}) + (1-\lambda_G)(\hat{r}\theta_{B+} - \theta_B). \\
\hat{q}_{35} &= \hat{f}\bar{U}v_G^{'} + (1-\lambda_B)(1-\hat{r})(m_B^{'} - v_B^{'}) + (1-\lambda_G)v'_G(\hat{r}\theta_{B+} - \theta_B).
\end{aligned}
\tag{A17}
$$

$$
\hat{q}_4 = (\theta_G - \theta_{G+})(1-\lambda_B)(\frac{c_{xx}^G v_B^{'}}{1-\lambda_G} - c_{xy}^G)\hat{q}_{41} + c_{xx}^G\hat{q}_{42}[c_{xy}^B\hat{q}_{43} - \frac{c_{xx}^B c_{xx}^G}{1-\lambda_B}\hat{q}_{44}],
\tag{A18}
$$

where

$$
\begin{aligned}
\hat{q}_{41} &= -\frac{1-\lambda_G}{1-\lambda_B}c_{xx}^B v_G^{''}[(1-\rho)\theta_B + \rho\hat{r}\theta_{B+}]. \\
\hat{q}_{42} &= -v_G^{''} - \frac{1-\lambda_B}{1-\lambda_G}v_B^{''}[(1-\rho)\theta_G + \rho\theta_{G+}]. \\
\hat{q}_{43} &= \hat{f}\bar{U} + (1-\lambda_B)(1-\hat{r}) + (1-\lambda_G)(\hat{r}\theta_{B+} - \theta_B). \\
\hat{q}_{44} &= \hat{f}\bar{U}v_G^{'} + (1-\lambda_B)(1-\hat{r})(m_B^{'} - v_B^{'}) + (1-\lambda_G)v_G^{'}(\hat{r}\theta_{B+} - \theta_B).
\end{aligned}
\tag{A19}
$$

These expressions are extremely complex, and to get some insight, we make the following simplifying assumptions. We assume that the distribution of shocks are uniformly distributed and are the same for the two sexes, so that $F = G$. This implies that $\theta_B = \theta_G = 1$, while $\theta_{B+} = 1$ and $\theta_{G+} = \hat{r}$. Second, we assume a cost function is sufficiently supermodular, so that $c_{xy} > 0$ and is large enough; however, we do not want to make to too large, since we would not have interior solutions for investments by the two sexes. Accordingly, we assume the following cost function:

$$c(x, y) = \frac{\gamma}{3}(x + y)^3 + \frac{1 - \gamma}{2}(x^2 + y^2),$$

with $\gamma = 0.29$ the chosen value.

We also assume quadratic utility for the public good:

$$v_G(y) = v_B(y) = ay - by^2,$$

where $a = 1, b = 0.1$ are the numerical values.

The bargaining power parameters are taken to be $\lambda_B = 0.8$, $\lambda_G = 0.2$. The sex ratio in the market with imbalances, i.e. $\hat{r}$ is set to 0.8. Finally, we take $\hat{f}U = 5$.

Note that we have interior solutions for all feasible values of r, i.e. those in the interval $[\hat{r}, 1]$.

Observe that the investments in the public good for boys are always decreasing in r, i.e. increased marriage market competition induces more investment in the public good, in line with our empirical results. The effects on private good investments by boys are ambiguous—these investments are decreasing in r for r values near 1, but are increasing in r for smaller values of r.

Investments by girls follow the opposite pattern from those by boys. Investments in the public good are consistently increasing in r. Appendix Figure A3 depicts the separate graph of x_G, girls' investments in the private good. These are decreasing in r for r values close to 1, but are increasing in r when the sex ratio is more unbalanced.

D Heterogenous effects across rural and urban families

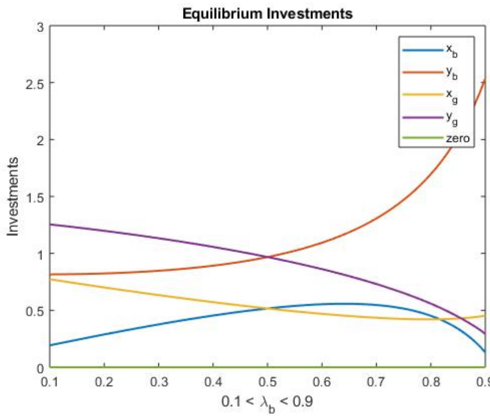
Below we present the heterogenous effects of sex-ratio imbalance across rural and urban households, and check whether the results are both supportive of the implications of our model. We divide the sample into a rural subsample and an urban subsample, and perform the same regression analysis using the two subsamples. Appendix Tables A5 and A6 present results on parental labor supply and premarital investments, respectively, based on 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 rural and urban families.

For parental migration, columns 1–3 of Appendix Table A5 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.

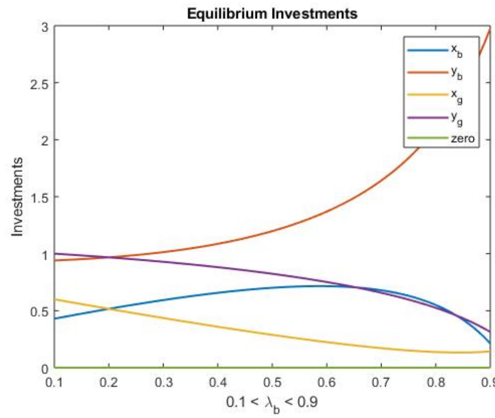
For housing investment, columns 1–3 of Appendix Table A6 show that the effect on construction area and ownership is larger and more significant for rural households; the effect on mortgage is larger for urban households, although not statistically significant. The hypothesis of no significant difference between rural and urban estimates for housing investment cannot be rejected at the ten percent level or below. This suggests that the possible difference in housing investment modes across rural and urban areas does not largely affect our analysis.

For most other outcomes of interest—parental working hours (columns 4 and 5 of Appendix Table A5) and child educational investment (columns 4 and 5 of Appendix Table A6), similarly, the hypothesis of no statistically significant difference between rural and urban estimates cannot be rejected at the ten percent level or below. Therefore, empirical results based on rural and urban households are both supportive of the implications of our model. But in column 6 of Appendix Table A6 where the outcome variable is the ratio of education expenditure per child to the market value of current housing, the hypothesis of no significant difference can be rejected at the five percent level, showing a significantly stronger reducing effect on educational investment relative to housing investment for rural families with sons.

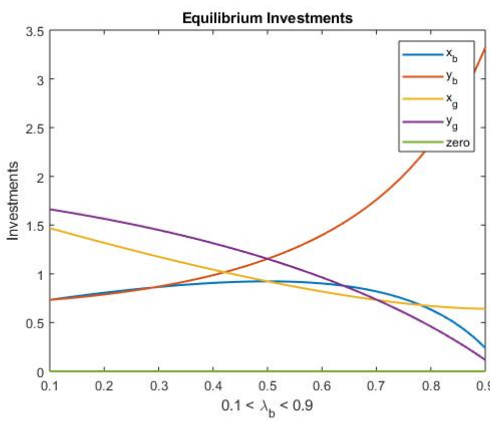
E Additional figures and tables



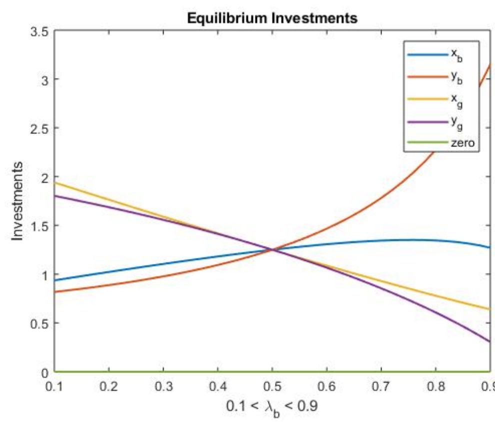
(a) $\gamma = 0.5, \theta_B = 1, \theta_G = 1$



(b) $\gamma = 0.2, \theta_B = 1, \theta_G = 1$



(c) $\gamma = 0.5, \theta_B = 2, \theta_G = 0.5$



(d) $\gamma = 0.2, \theta_B = 2, \theta_G = 0.5$

Figure A1 Equilibrium investments as a function of λ_B

Notes: This figure is generated in Online Appendix B. Each graph depicts the level of equilibrium investments as a function of $\lambda_B \in [0.1, 0.9]$.

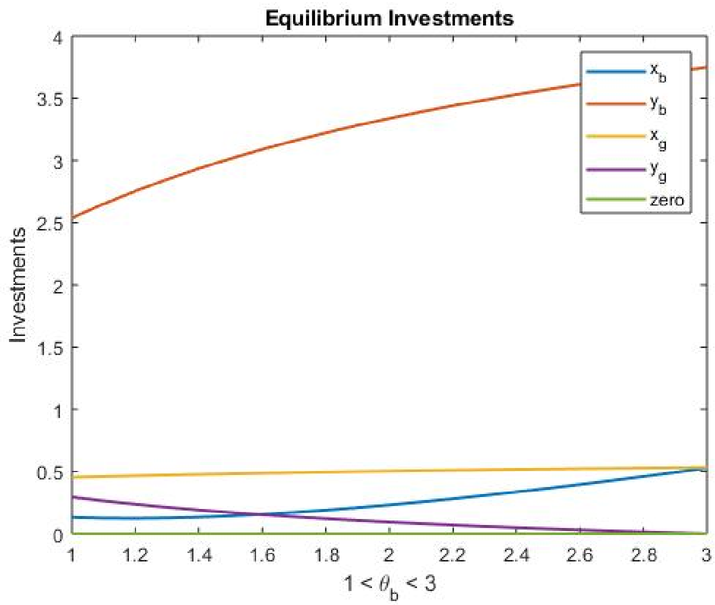


Figure A2 Equilibrium investments as a function of θ_B

Notes: This figure is generated in Online Appendix B. It depicts the level of equilibrium investments while fixing $\theta_B\theta_G$ to 1 with $\lambda_B = 0.9, \lambda_G = 0.5$ and $\gamma = 0.5$.

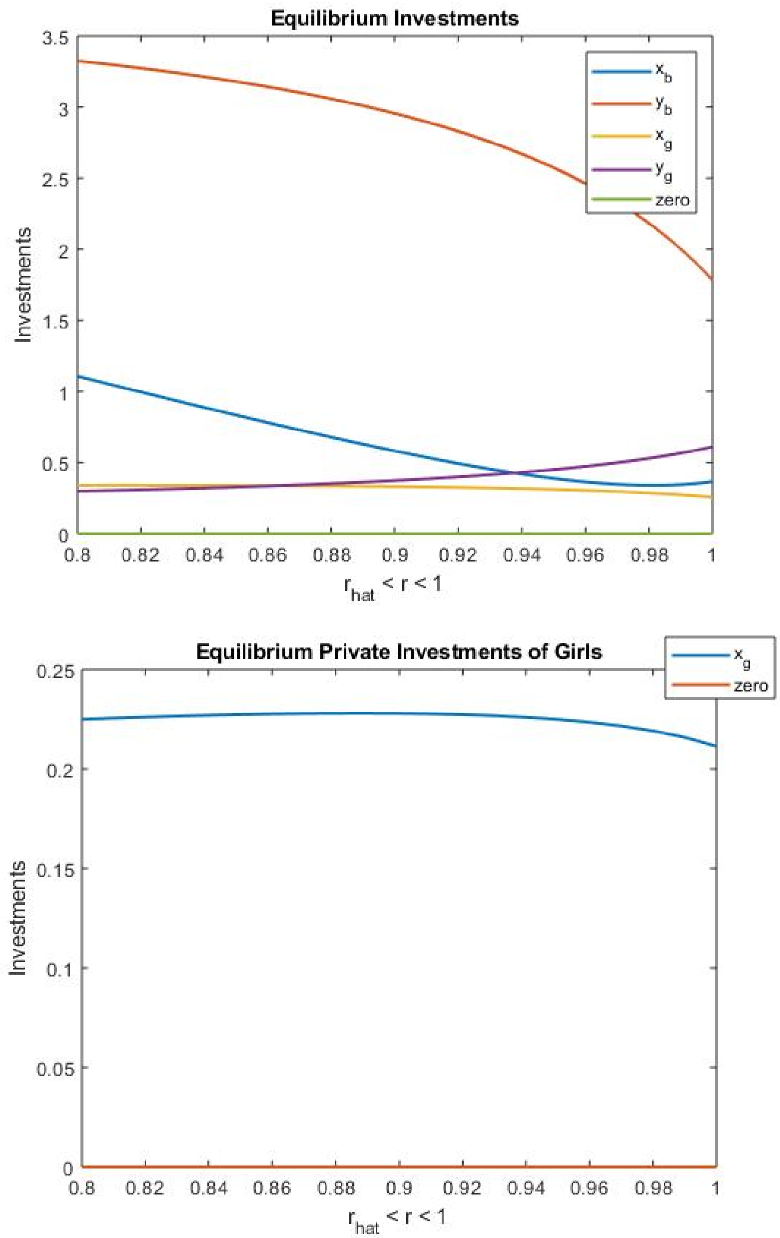


Figure A3 Equilibrium investments with unbalanced sex ratios

Notes: This figure is generated in Online Appendix C. It depicts the level of equilibrium investments when sex ratios are unbalanced.

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 2010 CFPS 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 after regressing the first-son dummy on the remaining variables in the table.

Table A2 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
Method	OLS	OLS	OLS	OLS	OLS
Other controls	YES	YES	YES	YES	YES
County fixed effects	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 is based on a subsample that excludes families with any migrant parents. Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table A3 Robustness results: 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

Notes: Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table A4 Robustness results: Premarital 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

Notes: Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table A5 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 (β_3)					
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 families					
P-value	0.209	0.039	0.226	0.384	0.498
Method	OLS	OLS	OLS	OLS	OLS
Other controls	YES	YES	YES	YES	YES
County fixed effects	YES	YES	YES	YES	YES

Notes: Panel A reports results for rural families and panel B, for urban families. Panel C reports p-values from Hausman's general specification tests that the estimated coefficient on the interaction in each column of panels A and B is equal. Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table A6 Rural and urban families: Premarital 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 families						
P-value	0.286	0.805	0.791	0.782	0.970	0.012
Method	OLS	OLS	OLS	OLS	OLS	OLS
Other controls	YES	YES	YES	YES	YES	YES
County fixed effects	YES	YES	YES	YES	YES	YES

Notes: Panel A reports results for rural families and panel B , for urban families. Panel C reports p-values from Hausman's general specification tests that the estimated coefficient on the interaction in each column of panels A and B is equal. Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

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

Table A7 Results on child human capital outcomes

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: Historical fertility as proxies for son preference						
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 * Proxies	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 between county-level historical fertility—"1981 Completed fertility" and "1981 Birth rate" as defined in notes to Table 1—and 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 * Penalty ratio," where "Penalty ratio" is defined in notes to Table 1. Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

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

Table A8 Regressions for proportion of sons and fertility

Dependent variable	Proportion of sons		Number of children	
	(1)	(2)	(3)	(4)
First son	0.787***	0.787***	-0.206***	-0.201***
	(0.013)	(0.013)	(0.025)	(0.023)
	[59.21]	[59.58]	[-8.239]	[-8.948]
Number of brothers of the father	-0.011***	-0.010***	-0.142***	-0.132***
	(0.003)	(0.003)	(0.012)	(0.011)
	[-4.011]	[-3.686]	[-12.10]	[-12.25]
Observations	3,772	3,772	3,772	3,772
R-squared	0.797	0.800	0.096	0.261
Other controls		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 and the number of brothers of the father. Columns 3 and 4 are the regression results of the number of children of the household on the first-son dummy and the number of brothers of the father. Columns 1 and 3 include only the two key regressors; columns 2 and 4 further add baseline 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 A9 Controlling for fertility in the unrestricted sample: 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	0.313***	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
B: IV results (penalty ratio, first-son dummy, and number of brothers of father as instruments)					
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: Regressions are based on equation (34). Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Table A10 Controlling for fertility in the unrestricted sample: Premarital 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	0.562***	0.278**	17.714	-0.586	-0.366*	-0.295**
	(0.212)	(0.118)	(11.309)	(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
B: IV results (penalty ratio, first-son dummy, and number of brothers of father as instruments)						
Prop. sons * Sex ratio	0.694**	0.440**	16.412	-2.686	-0.981***	-0.429*
	(0.323)	(0.216)	(20.803)	(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: Regressions are based on equation (34). Estimations are weighted by the CFPS survey sampling weights. Standard errors given in parentheses are clustered at the county level.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.