Gender Imbalance, Assortative Matching, and Household Income Inequality in China*

Chen Huang† Serhiy Stepanchuk‡

January 12, 2026

Abstract

This paper studies the effect of gender imbalance on assortative matching and household income inequality. Using data across prefectures in China, we show that a higher sex ratio in the marriage market is negatively associated with both assortative marriage and household income inequality. Motivated by empirical evidence, we develop a heterogeneous agent model to study the mechanism behind the pattern. The quantitative results of the model match the empirical evidence: a higher sex ratio is associated with a lower degree of assortative matching, which leads to a decrease in household income inequality. When we allow men and women to choose their level of education endogenously before entering the marriage market, we find that a higher sex ratio leads to a higher level of education investment among both men and women, with men investing more significantly than women.

Keywords: Sex Ratio, Assortative Marriage, Household Income Inequality, Education Attainment.

JEL Classification: J12; J16; J24; D31

---

*We would like to thank two anonymous referees and an associate editor for their comments. Their feedback has greatly helped improve this paper. We also thank the comments from Alessandro Mennuni, Sekyu Choi, Michael Hatcher, Héctor Calvo Pardo, and other comments at the annual conference of the Chinese Economists Society 2023. Specifically, we thank Dr. Yuankun Li's help.

†Tsinghua University, PBC School of Finance. Email: huangch@pbcsf.tsinghua.edu.cn

‡University of Southampton, Department of Economics. Email: s.stepanchuk@soton.ac.uk

1 Introduction

Does gender imbalance impact household income inequality? The growing surplus of males in China has attracted the interest of many economists,[1] with past research primarily exploring the effects of this imbalance on household saving rates, housing prices, crime rates, and subjective well-being (see Edlund et al., 2013; Li, 2021; Wei and Zhang, 2011; Wei et al., 2017). There are also studies examining the effects of sex imbalances on marriage matching outcomes (Abramitzky et al., 2011; Du et al., 2015; Yao et al., 2023). However, the literature that discusses the relationship between gender imbalance and household income inequality is still limited. In this paper, we investigate how the increasing sex ratio (the sex ratio in this paper is defined as the ratio of males to females in the marriage market) has affected matching patterns in the Chinese marriage market and study its implications for household income inequality.

Income inequality in China experienced a significant rise from 1978 to 2008 but has witnessed a modest reduction in the last few decades.[2] The literature has identified several factors contributing to this recent decrease in income inequality, including the diffusion of technology, urban-rural migration, the returns on human capital, and trade liberalization, as highlighted by Zhang (2021). Our research explores an innovative mechanism driven by gender imbalances in China, suggesting that these imbalances have shifted the patterns of assortative matching, subsequently influencing income inequality.

Marriage market matching patterns have important implications for many economic outcomes. Sorting in marriages affects the next generation's educational attainment, human capital accumulation, income inequality, and labor force participation decisions, as widely documented in the literature by Becker (1974), Fernandez et al. (2005), and Greenwood et al. (2014). An important determinant of marriage market outcomes is the sex ratio, i.e., the ratio of males to females in the population. Studies such as Abramitzky et al. (2011), Du et al. (2015), and Yao et al. (2023) find that sex imbalance influences the relative status (bargaining power, parent's wealth gap, personal income gap, etc.) of men and women in the marriage market, which has changed the degree of assortative mating.

In this study, we begin with an empirical analysis of how the sex ratio in the marriage market af-

[1] In Figure A1, we plot the sex ratio at birth in China from 1973 to 2015 using multiple rounds of Chinese population census data. The figure shows a rising male abundance since the introduction of the One-Child Policy, suggesting that men have faced increasing competition in the marriage market.

[2] We show the evolution of Gini coefficient in China in the appendix Figure A2.

fects assortative matching and household income inequality. Using household surveys, regional data, and population census information, we find that a higher sex ratio is associated with a lower degree of assortative marriage, measured by educational correlation, and with lower household income inequality. To address potential endogeneity concerns, we employ an instrumental variable strategy. The instrument is both strong and valid, and the estimated coefficients on the sex ratio remain negative and statistically significant, consistent with the OLS results. These findings strengthen the credibility of our causal interpretation.

Inspired by empirical observations, we develop a heterogeneous agent general equilibrium model to explain the findings and explore the underlying mechanisms. Our model, inspired by the approach of Fernandez et al. (2005), employs a two-stage matching framework that aims to capture the features of the marriage market characterized by an imbalance in the number of men and women. Our main modeling contribution relative to Fernandez et al. (2005) is the introduction of an exogenous and variable sex ratio in the marriage market, which enables us to analyze how demographic imbalances shape assortative matching and household income inequality—an extension beyond their original framework. Within this model, individuals can be classified as either skilled or unskilled. Before entering the marriage market, each person decides on their level of education, taking into consideration the anticipated returns from matching outcomes in the marriage market. Upon completion of their education, individuals enter the marriage market where they undergo one or two rounds of matching. In this market, an individual randomly encounters a member of the opposite sex. They observe the type of skill of each other, which perfectly predicts the future income of an individual. In addition, a matched couple draws a match-specific quality ('love') shock, and jointly decide whether to form a married couple or to remain single.

The quantitative results reveal that a higher sex ratio in the marriage market lowers assortative matching and reduces household income inequality. The mechanism operates through men's declining reservation match quality: as the surplus of men rises, they become less selective, leading to a greater share of random matches formed in the first stage. This shift weakens positive assortative mating and, in turn, diminishes household income inequality. The negative relationship between the sex ratio and inequality is robust across different measures of income dispersion. Lastly, we explore gender-specific responses in educational attainment decisions when they expect marriage competition among men to be more intensive. The model simulations indicate that an increasing sex ratio prompts both men and women to enhance their educational investments, with men investing more significantly than women.

Related Literature This study is closely related to two strands of the literature. Firstly, it relates to the research that investigates the interaction between household assortative matching and income inequality. Becker (1973, 1974) highlighted that males often marry females with similar socioeconomic backgrounds and that assortative matching is positively associated with household income inequality. Greenwood et al. (2014, 2016, 2017) use several waves of US census data to analyze the trend of positive assortative mating. They found that there is a rise in the positive assortative mating in the US between 1960 and 2005, and that positive assortative mating contributed to the increase of income inequality across households. Fernandez et al. (2005) reached a similar conclusion, using data from 33 countries to show a positive correlation between assortative matching and inequality. Their developed structural model further supports this finding. Fernández and Rogerson (2001) created a dynamic model of intergenerational education acquisition and marital sorting, illustrating that increased sorting significantly elevates income inequality. Eika et al. (2019) analyzed data from Denmark, Germany, Norway, the UK, and the US, documenting the extent of educational assortative matching and its contribution to household income inequality in each country. However, they found that variations in the degree of assortative matching have a weak influence on the time trends of household income inequality.

Secondly, our study relates to the research on the impact of sex ratio imbalances on economic outcomes, such as education investment, household savings rates, housing prices, and crime rates. Bhaskar and Hopkins (2016) theoretically suggested that an excess of males in the marriage market leads to parents over-investing in sons while under-investing in daughters. Abramitzky et al. (2011) and Angrist (2002) have shown that declining sex ratios lead to males marrying up in the social hierarchy in the United States and France, respectively. Lafortune (2013) empirically showed that a sex imbalance enhances male pre-marital investment in education. Wei and Zhang (2011) and Wei et al. (2017) found that the increasing sex ratio in China prompted families with sons to save more and invest in more expensive housing, where the savings and family housing value are used to indicate male's attractiveness in the marriage market. Edlund et al. (2013) discovered a positive correlation between rising sex ratios and increasing crime rates in China. Du et al. (2015) observed that the gender imbalance in China has resulted in females marrying up in terms of income disparity, and parental wealth gaps. Yao et al. (2023) found that when females are scarce in the marriage market, males tend to marry women from lower social classes, i.e., less educated or originating from rural regions.

We contribute to the academic literature in three distinct ways. First and foremost, we enrich the existing literature on the impact of sex imbalances on economic outcomes by concentrating on whether and how gender imbalances affect household income inequality. We focus on how competition within the marriage market leads to a decrease in assortative mating, subsequently affecting household income inequality. Second, after using both aggregate and household-level data to reveal the negative association between sex imbalance and household income inequality, we offer a structural interpretation of how the sex ratio in the marriage market affects household income inequality. Our findings reveal that an elevated sex ratio in the marriage market is inversely related to the level of assortative matching, thereby reducing household income inequality. Moreover, our analysis shows that as the sex ratio increases, men become less selective compared to women. Lastly, we incorporate the endogenization of educational choices in the context of gender imbalance before entering the marriage market. This approach enables us to examine the effects of marriage market competition on educational attainment in China, revealing that an increased sex ratio prompts both genders to enhance their educational investments.

The rest of the paper is organized as follows. Section 2 provides empirical evidence. Section 3 constructs a model to rationalize the empirical findings and explore the underlying mechanisms at play. Section 4 presents the quantitative results. Section 5 concludes.

2 Empirical Evidence

2.1 Data and sample

The data for this study are sourced primarily from the Chinese Household Income Projects (hereafter CHIPs), the Chinese Population Census, and the China City Statistical Yearbook. We utilize data from four CHIPs waves (2002, 2007, 2013, 2018) to gather household-level information. For each wave, the CHIPs datasets include both urban and rural samples, and we utilize data from both to compile comprehensive household-level information. We pool these 4 waves of CHIPs to form the data sample that will be used in reduced-form analysis.

For cleaning household-level data, we limit the age range for husbands and wives to 20-35 years,[3]

[3] CHIPs datasets do not include information about the timing of an individual's marriage, preventing us from determining the exact sex ratio at the time when the cohort enters the marriage market. Therefore, we use the sex ratio of the 20 to 35-year-old group as a proxy to estimate the sex ratio in the marriage market.

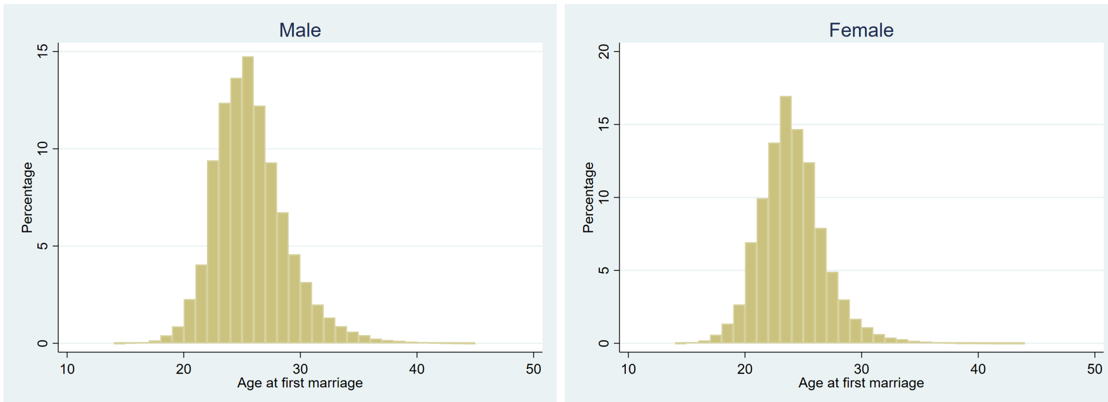


Figure 1: Age at first marriage in China

Note: We use 1% survey data (individual level) of 2005 China's population census to plot the age distribution for men and women.

focusing on first marriages. The rationale for concentrating on this specific age bracket and first marriages is twofold: Firstly, the majority of Chinese couples enter into their first marriage within the age range of 20-35, as shown in Figure 1. This figure presents the age distribution of first marriages for men and women in China, highlighting that most occur between the ages of 20 and 35. Also, the legal marriage age in China is set at 22 for men and 20 for women. Secondly, examining 'young marriages' provides a clearer picture of how the sex ratio in the marriage market influences matching outcomes. This is because marriage decisions for older individuals may be affected by various factors such as children, health conditions, and housing.

We measure educational correlation, household income inequality, and the sex ratio in the marriage market at the prefecture level, a regional division in China akin to a city. Educational correlation is calculated from the education levels of husbands and wives in the CHIPs survey. The dataset distinguishes nine education categories,[4] and for each prefecture, we use all couples in the 20–35 age cohort to compute the correlation coefficient between spouses' education levels. This coefficient serves as our measure of assortative marriage.

Household income is calculated as the combined job incomes of both the husband and wife. We then assess income inequality at the prefecture-level using these aggregated household incomes. The sex ratio (SR_{jt}) is defined for each prefecture j and wave t. Analyzing data at the prefectural level allows us to capture a more diverse range of sex ratio variations. This heterogeneity is depicted

[4] (1) Never schooled, (2) Elementary school, (3) Junior middle school, (4) Senior middle school, (5) Vocational senior secondary school, (6) Specialized secondary school, (7) Polytechnic college, (8) Undergraduate (bachelor's degree), (9) Graduate (master's degree or above).

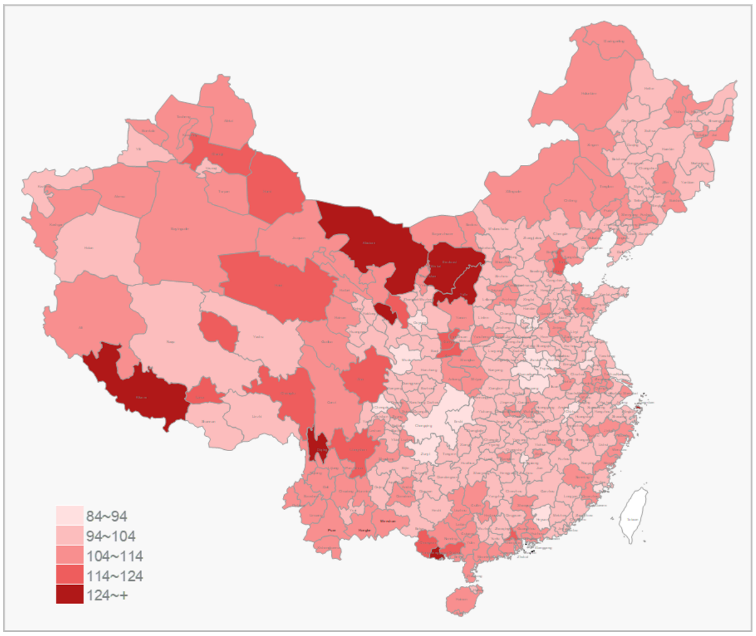


Figure 2: Sex ratio for the 20 to 35 age cohort across prefectures
Note: The darker the color, the higher the sex ratio. We employ data from the 2010 census to map out the sex ratio for the 20 to 35 age cohort at the prefecture level. The 2010 census provides data on the number of male and female residents in each county, categorized by various age groups (e.g., ages 5-9, 10-14, etc.). We aggregate these figures to the prefecture-level to calculate the sex ratio.

in Figure 2, where we illustrate the sex ratio for the 20 to 35 age cohort at the prefectural level, highlighting regional differences across China. As detailed in Wei et al. (2017), due to policy constraints on internal migration and barriers related to language and culture, marriages in China are predominantly local. This suggests that the sex ratio of male to female residents, as reported in the population census, sufficiently captures the intensity of local marriage market competition. For different waves of CHIPs data, we deduce the corresponding sex ratio from various waves of the population census. Table 1 provides detailed insights into how we determine the sex ratio in the marriage market.

Table 1: Cohort Age Selection

CHIPs wave	Cohort age	Inferred from	Corresponding age group in the census
2002	20 to 35	Census 2000	18 to 33
2007	20 to 35	Census 2010	23 to 38
2013	20 to 35	Census 2010	17 to 32
2018	20 to 35	Census 2015	17 to 32

Note: The 2000, 2010, and 2015 censuses provide population data on males and females across various age groups (such as 5–9, 10–14, etc.).

The summary statistics for the key variables are reported in Table 2. The mean educational correlation is 0.659, indicating a generally positive pattern of assortative marriage among couples in China. The Gini coefficient exhibits significant regional variation, with an average value of 0.294 and a standard deviation of 0.076. Its range extends from a minimum of 0.104 to a maximum of 0.605, indicating substantial heterogeneity in income inequality between regions. Similar patterns are observed in the analysis of two other income inequality measures: the logarithm of the income ratio between the 80th and 20th percentiles, and the logarithm of the income ratio between the 90th and 10th percentiles of households, both of which underscore the high regional variability of income inequality. Examining the sex ratio at the prefecture level, we find a standard deviation of 0.082, with the minimum and maximum values being 0.689 and 1.277, respectively, further highlighting its heterogeneity.

Table 2: Summary Statistics

Variables	Obs	Mean	S.D.	Min	Max
Educational Correlation	331	0.659	0.197	0.000	0.976
Gini Coefficient	253	0.294	0.076	0.104	0.605
Log of the income ratio between the 90th and 10th percentiles of households	236	1.662	0.786	0.032	5.726
Log of the income ratio between the 80th and 20th percentiles of households	236	0.964	0.436	0.000	3.797
Sex Ratio	331	1.022	0.082	0.689	1.277
Log(GDP)	331	3.341	0.946	1.048	5.257
Log(Population Size)	331	8.411	0.529	6.898	9.600
Log(Local Government General Budgetary Revenue)	331	2.078	1.358	-1.016	5.869
Urban-Rural Income Ratio	331	2.372	0.464	1.314	4.336
Log(Number of Hospitals and Health Clinics)	331	5.069	0.668	3.045	7.444
Log(Number of Hospital Beds)	331	9.660	0.662	8.090	11.819
Log(Number of Middle Schools)	331	5.463	0.505	3.871	6.568
Log(Number of Primary Schools)	331	6.459	0.847	4.111	8.428

Note: All variables are measured at the prefecture level. The Gini coefficient and the two other income inequality measures are computed by the authors using household income information from four waves of CHIPs data. The sex ratio is derived from the China Population Censuses of 2000, 2010, and 2015. All other control variables are obtained from the *China City Statistical Yearbook*.

2.2 Sex ratio, assortative marriage, and household income inequality

In this section, we empirically investigate how the sex ratio in the marriage market affects the degree of household assortative marriage and income inequality. Our hypothesis is that the link between the sex ratio in the marriage market and income inequality consists of two parts. First, the gender imbalance in the marriage market influences marital matching outcomes. Du et al. (2015) and Yao et al. (2023) find that the scarcity of females in China lead to men marry women from a lower social class (i.e., less educated or lower income), suggesting a decrease of assortative mating. Second, the degree of assortative matching in marriage affects household income inequality, as extensively documented in the works of Eika et al. (2019), Fernandez et al. (2005), and Greenwood et al. (2014). These studies have found a positive correlation between educational assortative marriage and income inequality.

Table 3: Sex Ratio, Degree of Assortative Matching, and Household Income Inequality

	(1)	(2)	(3)	(4)
Dependent Variable:	*Educational Correlation*	*Gini*	*Wage9010*	*Wage8020*
Sex Ratio	-0.297*	-0.175**	-1.435**	-0.286**
	(0.178)	(0.082)	(0.698)	(0.139)
Log(GDP)	0.180*	-0.009	-0.198	-0.017
	(0.098)	(0.040)	(0.580)	(0.343)
Log(Population Size)	0.010	0.233***	0.473	0.132
	(0.185)	(0.083)	(0.988)	(0.546)
Log(Local Government General Budgetary Revenue)	-0.061	0.017	-0.384	-0.006
	(0.071)	(0.033)	(0.457)	(0.237)
Urban-Rural Income Ratio	0.024	0.040	-0.614*	-0.114
	(0.068)	(0.029)	(0.337)	(0.207)
Log(Number of Hospitals and Health Clinics)	-0.031	-0.016	-0.048	-0.161
	(0.043)	(0.015)	(0.172)	(0.133)
Log(Number of Hospital Beds)	0.098	0.061	0.905*	0.579**
	(0.097)	(0.047)	(0.538)	(0.271)
Log(Number of Middle Schools)	-0.134	-0.084	-1.917*	-0.530
	(0.151)	(0.066)	(1.013)	(0.678)
Log(Number of Primary Schools)	0.072	-0.018	0.465	0.203
	(0.051)	(0.023)	(0.358)	(0.237)
Constant	-0.176	-1.530**	1.082	-2.783
	(1.479)	(0.688)	(5.723)	(3.469)
Observations	331	253	236	236
R-squared	0.478	0.606	0.598	0.560
Prefecture FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. We have consolidated data from four waves of CHIPs, encompassing both rural and urban samples, to gather household income information. In our analysis, we focus on couples where both the husband and wife are between 20 and 35 years old and are in their first marriage. The sex ratio in the marriage market is derived from population census data.

Table 3 reports the regression results, where both region and year fixed effects are controlled. To study the relationship between gender imbalances in the marriage market and the degree of assortative matching between the spouses, column (1) reports the regression results where education correlation between the spouses is used as a dependent variable. In columns (2)-(4), we study the relationship between the sex ratio and household income inequality. Household income inequality measures are represented by three metrics: the Gini coefficient (*Gini*), the logarithm of the income ratio between the 90th and 10th percentiles (*Wage9010*), and the logarithm of the income ratio between the 80th and 20th percentiles (*Wage8020*). In column (1), we find that the regression coefficient for the sex ratio is -0.297, indicating that a higher sex ratio in the marriage market is associated with a lower degree of assortative matching. This finding is consistent with the existing literature. Turning to income inequality, the results in columns (2)-(4) also show negative and significant effects across all three measures, suggesting that an increase in the sex ratio reduces household income inequality. However, these OLS estimates are descriptive correlations rather than causal effects; to address potential endogeneity concerns, we turn to an instrumental-variable approach in the next section.

2.3 Instrumental variable approach

To resolve the potential endogeneity concern, we use a leave-one-out spatial instrument: the average sex ratio across other cities within the same province (excluding the target city). This variable serves as a proxy for shared regional demographic pressures, policy enforcement intensity (e.g., local implementation of the one-child policy), and cultural norms that jointly influence sex ratio imbalances across cities within a province.

This instrument satisfies the two key requirements for a valid instrumental variable: (i) Relevance: cities within the same province typically share similar demographic policies, migration patterns, birth planning norms, and economic conditions. As a result, the sex ratio in other cities is highly predictive of the target city's own sex ratio. Our first-stage regression confirms this strong relationship, with F-statistics well above conventional thresholds. This ensures that the instrument is informative for explaining variation in the endogenous regressor (local sex ratio). (ii) Exogeneity: the leave-one-out construction ensures that the instrument does not mechanically contain information about the target city's own outcome (i.e., educational correlation, household income inequality). Furthermore, by controlling for prefecture fixed effects and year fixed effects, we purge the influence of time-invariant local characteristics and national trends. Under these conditions, any direct effect of other cities' sex ratios on local income inequality is highly unlikely, leaving the local sex ratio as the only plausible channel through which the instrument affects the outcome.

This approach aligns with the empirical logic of spatial and contextual instruments employed in previous studies, such as Chetty and Hendren (2018) and Kleven et al. (2019), where leave-one-out neighborhood averages are used to isolate local causal effects without contamination from the target unit's own data.

The 2SLS estimates are reported in Table 4. Across all specifications, the coefficients on the sex ratio remain negative, consistent with the baseline regressions.[5] The effects on household income inequality are negative and statistically significant across all three measures. Moreover, the magnitudes are larger in absolute terms compared with the OLS estimates, implying that failing to account for endogeneity biases the baseline estimates toward zero. For instance, the IV coefficient for the 90/10 income ratio is -5.627, substantially larger than its OLS counterpart. This suggests that unobserved factors—such as selective migration or local shocks correlated with both sex ratios and income distribution—may have attenuated the baseline results.

Taken together, these findings reinforce the causal interpretation of our main result: demographic imbalances in the marriage market systematically reduce household income inequality.

[5]For educational correlation, the coefficient continues to be negative but becomes statistically insignificant. A plausible explanation is that assortative matching is jointly shaped by a wider range of cultural and social preferences that are not fully captured by demographic variation alone, making the IV estimates less precise.

Table 4: The Effect of Sex Ratio on Assortative Marriage and Household Income Inequality: 2SLS Estimates

Panel A: 2SLS Results				
	(1)	(2)	(3)	(4)
Dependent Variable:	*Educational Correlation*	*Gini*	*Wage9010*	*Wage8020*
Sex Ratio	-0.138	-0.359**	-5.627**	-2.641*
	(0.202)	(0.156)	(2.253)	(1.335)
Log(GDP)	-0.005	0.003	0.379	0.333
	(0.045)	(0.043)	(0.609)	(0.388)
Log(Population Size)	-0.036	0.252**	1.538	1.326
	(0.052)	(0.101)	(1.336)	(0.867)
Log(Local Government General Budgetary Revenue)	0.037	0.013	-0.234	0.273
	(0.033)	(0.031)	(0.462)	(0.284)
Urban-Rural Income Ratio	0.053**	0.037	-0.698*	0.081
	(0.026)	(0.027)	(0.404)	(0.271)
Log(Number of Hospitals and Health Clinics)	-0.041*	-0.021	-0.045	-0.170
	(0.021)	(0.017)	(0.255)	(0.151)
Log(Number of Hospital Beds)	0.085**	0.067	-0.087	0.226
	(0.043)	(0.046)	(0.655)	(0.423)
Log(Number of Middle Schools)	-0.618	-0.092	-1.279	-0.154
	(0.057)	(0.072)	(0.945)	(0.660)
Log(Number of Primary Schools)	0.010	-0.024	0.159	0.046
	(0.020)	(0.023)	(0.314)	(0.208)
Observations	331	253	170	213
R-squared	0.161	0.069	0.099	0.135
Prefecture FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Panel B: First Stage				
	(1)	(2)	(3)	(4)
Dependent Variable:	*Sex Ratio*	*Sex Ratio*	*Sex Ratio*	*Sex Ratio*
Instrument: Average sex ratio in other cities of the same province	0.968***	0.970***	0.927***	0.976***
	(0.660)	(0.141)	(0.200)	(0.156)
Anderson canonical corr. LM statistic	132.833	66.479	34.926	56.247
Cragg-Donald Wald F statistic	214.833	25.127	10.989	20.274
Sargan statistic	0.000	1.109	1.697	0.664

Note: All regressions control for prefecture and year fixed effects. Robust standard errors clustered at the prefecture level are in parentheses. Panel B reports first-stage results and instrument diagnostics. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The number of observations is smaller in the IV regressions because the construction of the instrument (prefecture–year average sex ratio) and the absorption of fixed effects drop prefecture–year cells with missing instruments or singleton groups. The reduction in sample size is therefore mechanical and does not reflect selective exclusion of data.

3 The Model

In this section, we develop a heterogeneous-agent general equilibrium model that builds on Fernandez et al. (2005), and use it to study the impact of the increasing sex ratio ("male abundance") on household assortative matching and its implications for income inequality. Different from their framework, we explicitly introduce gender imbalance in the marriage market by allowing the sex ratio to be exogenously determined and varied across scenarios. This modeling innovation enables us to move beyond the balanced setting of Fernandez et al. (2005) and directly analyze how demographic imbalances shape marriage market outcomes.

3.1 Setup

The economy is populated by individuals of two genders, men and women. Both men and women can be either skilled, 's', or unskilled, 'u'. There are four types of agents: skilled men, skilled women, unskilled men, and unskilled women. Following Fernandez et al. (2005), we employ a two-stage matching model. In the first stage, matches are formed randomly across genders, while in the second stage, unmatched individuals are paired within their education group (i.e., skilled with skilled and unskilled with unskilled).

Before entering the marriage market, each individual decides what education level to acquire: they can become skilled or unskilled. Once they have finished their education, all agents enter the marriage market, where they go through one or two rounds of matching. They observe each other's skill type, which perfectly predicts the future income of an individual. In addition, a matched couple draws a match-specific quality ('love') shock, and jointly decide whether to form a married couple or to remain single.

Men and women in a married couple share a common joint utility function. Each agent is assumed to derive utility from consumption and from the quality of his (her) match. The indirect utility function for a couple with match quality q and household income I is:

$$V(I,q) = U(I) + q = U[e(w_m^z + w_f^z)] + q \tag{1}$$

where U is a continuous and strictly increasing function of I; w_m^z denotes the wage of males with skill type z (z is either skilled 's' or unskilled 'u') and w_f^z denotes the wage of females with skill type z. e is used to measure the consumption equivalence scale, which is a measure of the cost of living for a household given family size and demographic composition. We follow the literature and set $e = 0.7$ in this study. Household income is:

$$I = \begin{cases} w_m^s + w_f^s, \\ w_m^s + w_f^u, \\ w_m^u + w_f^s, \\ w_m^u + w_f^u, \end{cases} \tag{2}$$

The total number of men is normalized to 1 ($N^m = 1$). As a result of the 'male abundance', the number of women is smaller than the number of men ($N^f < 1$). In the model, $\frac{N^m}{N^f} = \frac{1}{N^f}$ denotes the sex ratio in the marriage market. ψ_m is the share of skilled men among all men, ψ_f is the share of skilled women among all women.

3.2 Matching

There are two rounds of matching in the marriage market. In the first round, agents meet randomly and draw a random match-specific quality q. We assume that q is non-negative and follows a continuous distribution, taking the possible values in the $[0, q_{\max}]$ interval.[6] The cumulative distribution function (CDF) of q is Q, with the expected value equal to μ. The match can be accepted by both agents (resulting in a marriage), or rejected by at least one of the agents. In the latter case, both agents go to the second round of matching.

In the second round, all remaining unmatched women are guaranteed a partner within their own education group (skilled unmarried women are matched with skilled unmarried men, while unskilled unmarried women are matched with unskilled unmarried men). A new random match quality is drawn from the same distribution Q. Since the number of men exceeds the number of women ($N^m > N^f$), all women will eventually be matched, but some men may remain single after the second stage. This feature ensures the consistency of the model and reflects the demographic imbalance that motivates our analysis.

Let $\bar{q}_g^{zz}$ be the cutoff match quality for an agent with gender g and in a match type zz, where $g \in \{m, f\}$, and $zz \in \{ss, su, us, uu\}$. Below this cutoff, the agent will not accept the match. Thus,

[6]The assumption of $q \geq 0$ ensures that all women accept a marriage in the second stage, whereas some men will remain single due to there being a surplus of men relative to women.

the probability of agents getting married is $1 - Q\left[\max(\bar{q}_m^{zz}, \bar{q}_f^{zz})\right]$.

In the first round, the total number of meetings is N^f. Since $N^m > N^f$, some men will not meet a potential partner. For example, the number of meetings between skilled male and unskilled female is: $N^f \psi_m (1 - \psi_f)$.

In the second round, some part of the remaining unmarried skilled males match (and get married) with skilled females, and some part of the remaining unmarried unskilled males match (and get married) with unskilled females. Let θ_g^z stand for the number of unmarried agents after the first round of the match with gender g and skill type z. For instance, the number of unmarried unskilled males after round 1 is:

$$\theta_m^u = (1-N^f)(1-\psi_m)+N^f(1-\psi_m)\psi_f Q\left[\max(\bar{q}_m^{us}, \bar{q}_f^{us})\right]+N^f(1-\psi_m)(1-\psi_f)Q\left[\max(\bar{q}_m^{uu}, \bar{q}_f^{uu})\right] \quad (3)$$

There is a trade-off between accepting the first round of matching and rejecting it, in which case the agent goes to the second round of matching. In particular, a skilled male who meets an unskilled female in the first round and draws a high match quality will encounter a trade-off between forming a lower-income household with high match quality, or rejecting this match and forming a higher-income household with a skilled female in the second round but with an expected match quality that is μ.

3.3 Equilibrium

Given $V(I, q) = U(I) + q$, a skilled male will be indifferent between accepting a first-round match with an unskilled female and rejecting that match and entering into second round if:

$$V(I, \bar{q}_m^{su}) = E[V(I, \mu)]$$

where $\bar{q}_m^{su}$ denotes the reservation match quality for the skilled male when he meets with an unskilled female. The first character in the superscript 's' is used to denote the male's skill type, whereas the second character 'u' is used to denote the female's skill type (we follow the same notations in the rest of the paper).

In the following part, we use this notation to represent the reservation match quality for each type of agent. Thus, there are eight reservation match qualities in total. The equilibrium is determined by eight match quality cutoffs in the first round. Solving for the value of q at the equilibrium yields the reservation quality of $\bar{q}_g^{zz}$.

In round 1, a skilled male matches with a skilled female will be indifferent if:

$$U[e(w_m^s + w_f^s)] + \bar{q}_m^{ss} = \min(\frac{\theta_f^s}{\theta_m^s}, 1) \left[U[e(w_m^s + w_f^s)] + \mu\right] + \left[1 - \min(\frac{\theta_f^s}{\theta_m^s}, 1)\right] U(w_m^s)$$

And if

$$U[e(w_m^s + w_f^s)] + q_{\min} > \min(\frac{\theta_f^s}{\theta_m^s}, 1) \left[U[e(w_m^s + w_f^s)] + \mu\right] + \left[1 - \min(\frac{\theta_f^s}{\theta_m^s}, 1)\right] U(w_m^s)$$

then all the skilled male that are matched with a skilled female will want to get married in round 1.

If

$$U[e(w_m^s + w_f^s)] + q_{\max} < \min(\frac{\theta_f^s}{\theta_m^s}, 1) \left[U[e(w_m^s + w_f^s)] + \mu\right] + \left[1 - \min(\frac{\theta_f^s}{\theta_m^s}, 1)\right] U(w_m^s)$$

then no skilled male matched with a skilled female will want to get married in round 1. Where $q_{\max}$, $q_{\min}$ represent the upper bound and lower bound of match quality.

In summary, the equilibrium is characterized by eight indifference conditions. As an illustration, consider the indifference condition for an unskilled male when paired with a skilled female; the remaining conditions are presented in Appendix B.

$$U[e(w_m^u + w_f^s)] + \bar{q}_m^{us} = \min(\frac{\theta_f^u}{\theta_m^u}, 1) \left[U[e(w_m^u + w_f^u)] + \mu\right] + \left[1 - \min(\frac{\theta_f^u}{\theta_m^u}, 1)\right] U(w_m^u) \tag{4}$$

3.4 Equilibrium distribution of households

Since we have solved the reservation match quality for different types of match, we can generate the equilibrium distribution of household types. Let π_{zz} denotes the number of different types of households in the population, thus we have:

The number of skilled husband with skilled wife in the population:

$$\pi_{ss} = \underbrace{N^f \psi_m \psi_f \left(1 - Q[\max(\bar{q}_m^{ss}, \bar{q}_f^{ss})]\right)}_{\text{marriage at the first round}} \quad + \quad \underbrace{\min(\theta_m^s, \theta_f^s)}_{\text{marriage at the second round}}$$

The number of skilled husband with unskilled wife in the population:

$$\pi_{su} = N^f \psi_m (1 - \psi_f) \left(1 - Q[\max(\bar{q}_m^{su}, \bar{q}_f^{su})]\right)$$

For other types of households in the population, we present them in Appendix B.

Note that we have:

$$\pi_{ss} + \pi_{su} + \pi_{us} + \pi_{uu} = N^f$$

Also have:

$$\pi_{ss} + \pi_{su} + \pi_{us} + \pi_{uu} + \pi_s + \pi_u = 1$$

where the total number of marriages equals the number of females in the population.

3.5 Education correlation between husband and wife

Definition: we define the education correlation between husband and wife ρ as:

$$\rho = \frac{\frac{\pi_{ss}}{N^f} - \frac{\pi_{ss}+\pi_{su}}{N^f} \frac{\pi_{ss}+\pi_{us}}{N^f}}{\sqrt{\left(\frac{\pi_{ss}+\pi_{su}}{N^f}\right) - \left(\frac{\pi_{ss}+\pi_{su}}{N^f}\right)^2} \sqrt{\left(\frac{\pi_{ss}+\pi_{us}}{N^f}\right) - \left(\frac{\pi_{ss}+\pi_{us}}{N^f}\right)^2}} \tag{5}$$

We show the detailed derivations and descriptions in Appendix B.

4 Quantitative Analysis

4.1 Calibration

The model is calibrated to match the empirical evidence and reproduce quantitative properties. The utility function takes the form of:

$$U(I) = log(I)$$

We assume the match quality q follows a continuous uniform distribution. The support of q is $[0, 1]$, where the expected value is $\mu = 0.5$ in this setting. A skilled agent is defined as someone who possesses educational qualifications beyond junior middle school, including those from senior middle school, vocational senior secondary school, polytechnic colleges, undergraduate programs, and graduate degrees.[7]

The wages for different types of agents are exogenously calibrated. We employ CHIPs 2018 data to project lifetime wages for four types of agents, as it covers the most number of provinces in China compared with other waves of CHIPs data.[8] The life cycle of an individual is divided into 5-year intervals, the age range of men is from 20 to 60, and the age range of women is from 20 to 55, the different selection is due to the mandatory retirement age in China, which is different for men and women (the restricted age is 60 for men and 55 for women). For the projections of lifetime wage of males, the interval is 20-24, 25-29, 30-34, ..., 56-60. For females, it's up to 55. The wage we used is the before-tax net labor wage, as tax may play a role in redistributing incomes. We initially compute the average wage of unskilled women in the age group of 20-24, and do so for other age groups until the group of 51-55, then use the annual discount factor of 0.96 to get the present value of unskilled women's lifetime wages.[9] We follow the same procedures to compute the lifetime wages of skilled women, unskilled men, and skilled men. The wage of the unskilled female is normalized to 1, and we use the wage premiums to obtain those for skilled females, unskilled males, and skilled males. Where the wage premium of a skilled female is equal to the ratio of the projected lifetime wage for a skilled female over an unskilled female, the same setting applies to other types of agents.

The skilled shares of men (ψ^m) and women (ψ^f) are also calibrated to match empirical evidence.

[7]This study divides individual education level into 9 categories, including ① never schooled, ② elementary school, ③ junior middle school, ④ senior middle school, ⑤ vocational senior secondary school, ⑥ specialized secondary school, ⑦ polytechnic college, ⑧ undergraduate (bachelor's degree), ⑨ graduate (master's degree or above)

[8]We use both urban and rural samples to project the wages. CHIPs 2018 covers 15 provinces in China (there are 31 provinces in China), which gives us the most number of observations compared with other waves.

[9]For each age group, the group discount factor is $0.96^5 = 0.815$

The study uses the Chinese population census 2015 to measure the fraction of skilled men and skilled women.

We vary the number of women in the model, N^f, in the range from 0.81 to 0.98, which implies the sex ratio varies from 1.031 to 1.235. The detailed calibration results can be found in Table 5.

Table 5: Calibrated parameters

Parameters	Meaning	Value	Source or target(s)
ψ_m	fraction of skilled male	0.453	Census 2015
ψ_f	fraction of skilled female	0.412	Census 2015
w_f^u	wage of unskilled female	1	Normalized to 1
w_m^u	wage of unskilled male	1.411	CHIPs
w_f^s	wage of skilled female	1.624	CHIPs
w_m^s	wage of skilled male	2.050	CHIPs

4.2 Education correlation between couples

To study the impact of gender imbalances on the degree of assortative matching, we vary the sex ratio in our model and investigate what happens to the education correlation (computed using the formula in equation 5) in equilibrium. Using education correlation between couples as a proxy for the degree of assortative matching in marriages follows the idea of Eika et al. (2019) and Greenwood et al. (2014).

As illustrated in Figure 3, the education correlation between couples is decreasing with the increase in the sex ratio. The intuition here is that the rising number of males in the marriage market makes men less likely to find a wife, thus males are less selective, which results in a decrease in assortative matching[10]. This finding is consistent with the results of other papers; Du et al. (2015) and Yao et al. (2023) find that when females are scarce, males tend to marry females of a lower social class, i.e., those who are less educated, have lower personal income and come from families with lower wealth.

To check how the model fits with the data, we compare the moments of education correlation between model output and data sample. The results are reported in Table 6, we find the mean of education correlation in the model output is close to the value in the data sample, which is 0.6256

[10]Table C1 shows how the 'reservation match quality' changes for men and women in various types of matched couples as one changes the sex ratio in the model. When the sex ratio increases (men become relatively more 'abundant'), the reservation match quality cut-offs for men decrease, which means that they become less selective

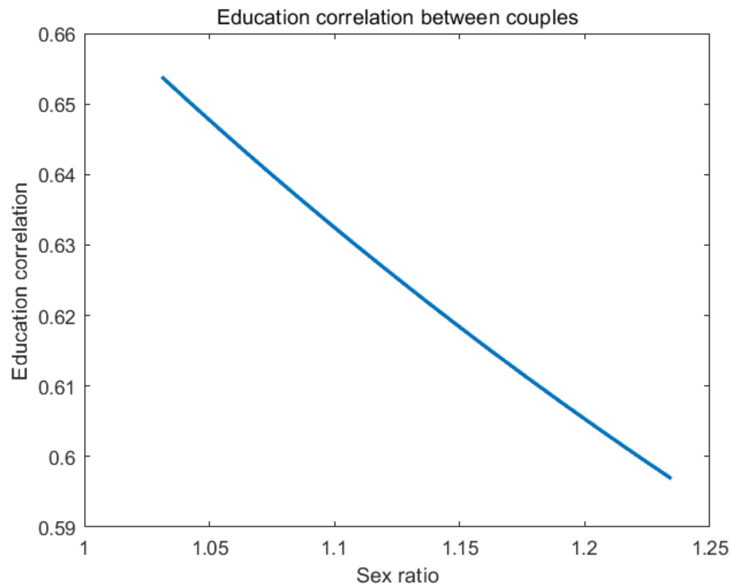


Figure 3: Assortative matching and sex ratio

and 0.6183 respectively.

Table 6: How model fit to data

Moment	Model	Data Sample
Education correlation (mean)	0.6256	0.6183

Note: We compute the mean of the education correlation using the data sample from the empirical analysis.

4.3 Implication for household income inequality

This section reports on how the sex ratio in the marriage market impacts household income inequality. Multiple indices are used to measure income inequality. First, we use the Gini coefficient to reveal the relationship between the sex ratio in the marriage market and household income inequality. By constructing the Lorenz curve (formulating the cumulative household share and cumulative income share), the Gini coefficient can be derived analytically, the derivations are reported in Appendix D.

Figure 4 illustrates that in the marriage market, the sex ratio is inversely correlated with the Gini coefficient, suggesting that a higher sex ratio correlates with lower household income inequality. The underlying rationale for this pattern relates to the positive association between assortative marriage and household income inequality, a theme explored in several studies (e.g., Eika et al.,

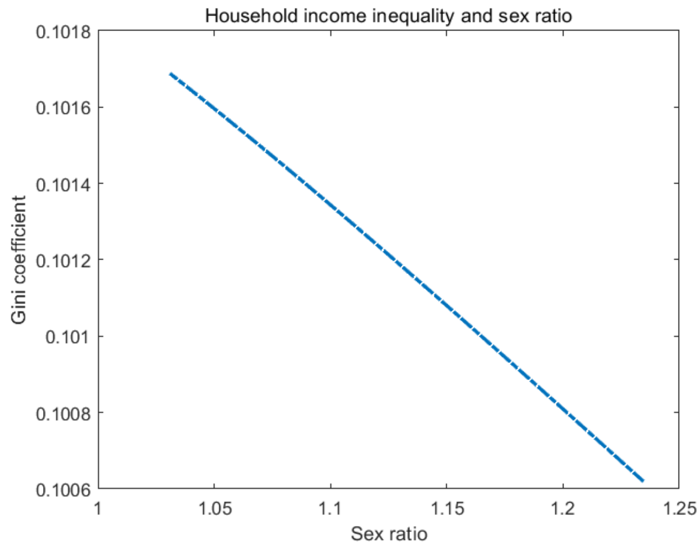


Figure 4: Gini coefficient and sex ratio

2019; Fernandez et al., 2005; Greenwood et al., 2014). Conversely, our prior analysis reveals that an elevated sex ratio corresponds to reduced assortative matching. This inverse relationship indicates that a higher sex ratio may lead to decreased household income inequality. Intuitively, a decline in assortative marriage suggests a reduction in the proportion of couples with similar education levels (positive assortative mating) and an increase in couples with differing education levels, as shown in Table C2. These shifts in household formation contribute to the reduction in household income inequality, reinforcing our conclusion that the sex ratio is negatively associated with household income disparity.

We also use other measures to study how the sex ratio impacts the inequality of household income. Figure 5 presents four measures of income inequality, which are the coefficient of variation, Theil L index, Theil T index, and logarithmic variance, respectively. These indices provide a more comprehensive understanding of the income distribution and its skewness. For instance, the coefficient of variation emphasizes the standard deviation of income relative to the mean, offering insights into the overall variability of household income. On the other hand, Theil's indices, particularly the L and T variants, offer a decomposition feature that allows for a detailed analysis of inequality within and between different population subgroups. Similarly, the logarithm of variance addresses scale dependency issues by normalizing the variance, thus providing a measure of dispersion that is not influenced by the income level itself (see Cowell, 2011).

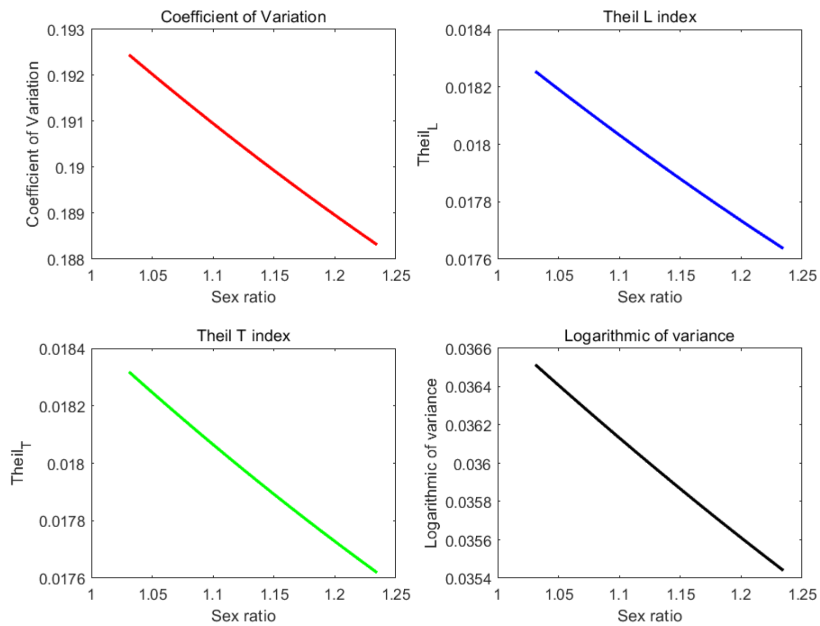


Figure 5: Sex ratio and household income inequality

Our analysis, as shown in Figure 5, demonstrates that the sex ratio in the marriage market exhibits a consistent inverse relationship with these indices. Specifically, the coefficient of variation, which gauges the dispersion of income around the mean, shows a downward trend as the sex ratio increases, indicating a contraction in relative income disparities. Furthermore, the Theil indices, both L and T, show a similar downward trend with an increasing sex ratio. The Theil L index, which places greater emphasis on the lower tail of the income distribution, suggests that a higher sex ratio may contribute to raising incomes at the lower end, thereby reducing inequality. The Theil T index, more sensitive to the upper tail, implies that the top earners' share of total income decreases as the sex ratio rises. This could be interpreted as a decline in the income advantage that typically accrues to households with high educational or occupational homogamy, as the sex ratio increases. Lastly, the logarithmic variance, which mitigates the scale dependency of variance, also shows a reduction with an elevated sex ratio. This indicates that income differences between households are becoming less pronounced, leading to a more even spread of earnings among families.

4.3.1 Quantitative comparison between the model and data

In this section, we present a quantitative comparison between the model and the empirical results. Specifically, we replicate the regressions reported in Table 3 using the simulated data from the calibrated model and contrast the resulting coefficients with those obtained from the empirical sample.

The comparison highlights two key findings. First, the model reproduces the relationship between sex ratios and assortative marriage well. The estimated coefficient of the sex ratio on educational correlation is -0.289, closely aligned with the empirical estimate of -0.297. This correspondence suggests that the model successfully captures the mechanism by which a male-biased marriage market increases the share of random matches, thereby weakening assortative matching.

Second, the model predicts a smaller impact of sex ratios on household income inequality than observed in the data. For the Gini coefficient, the model delivers an estimate of -0.0052, compared with an empirical coefficient of -0.175, While the direction of the effect is consistent, the magnitude is underestimated. This limitation stems from the parsimonious structure of the model, which includes only four discrete household types. Such structure flattens the Lorenz curve and may dampen the sensitivity of inequality measures to changes in household composition. We view this as a promising direction for future research: extending the framework to incorporate richer heterogeneity—such as multiple skill levels or continuous wage distributions—would allow the model to generate a more nuanced income distribution and better match the empirical magnitudes of inequality responses.

Table 7: Model-implied coefficients vs. data estimates

Outcome / Regressor	**Model coeff.**	**Data coeff.**
Educational correlation	-0.289	-0.297
Gini coefficient	-0.0052	-0.175
Estimation	OLS on model output	OLS on data

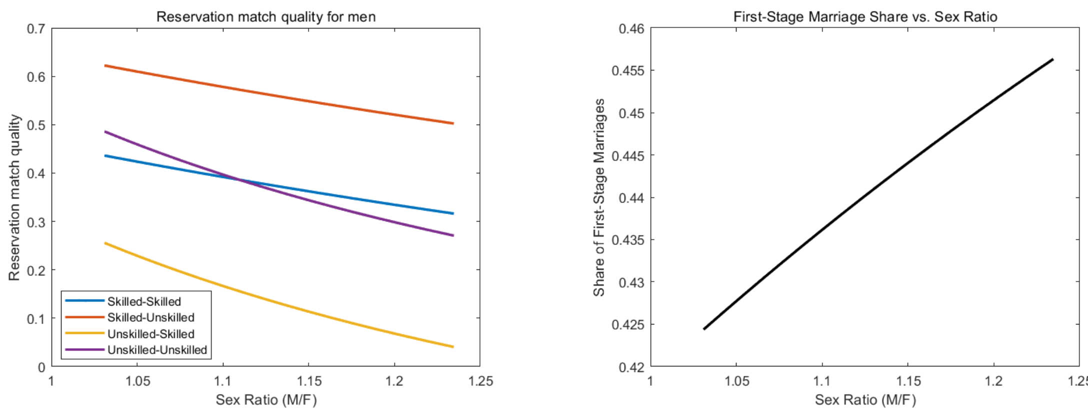
Note: Coefficients are estimated from regressions of each outcome on the sex ratio. "Model coeff." is based on simulated data from the calibrated model; "Data coeff." is based on the empirical sample.

4.4 Inspecting the mechanisms

The negative effect of the sex ratio on assortative matching in our model operates through the interaction of reservation thresholds and the two-stage matching structure. When the number of men increases, competition in the marriage market intensifies. As a result, men become less selective: they lower the reservation match quality that they are willing to accept in the first round. Figure 6a illustrates this pattern across different skill pair types, showing that the reservation cutoff declines as the sex ratio rises.

This decline in reservation thresholds has direct implications for the share of first-stage marriages. Because the first round involves random matching across education groups, a lower reservation cutoff increases the probability that men accept these random matches. As shown in Figure 6b, the first-stage marriage rate rises monotonically with the sex ratio. More marriages are therefore concluded in the random stage, reducing the role of the second-stage within-group (assortative) matching.

The shift toward a greater share of random matches has important distributional consequences. Since random matches weaken the positive correlation between spouses' education levels, the overall degree of assortative matching falls. With less assortative sorting by education, income differences across households narrow, leading to lower household income inequality in equilibrium.



(a) Reservation match quality for men

(b) Marriage rate in the first stage matching

Figure 6: Reservation match quality for men and first stage marriage rate

4.5 Robustness check: alternative matching function

To examine the robustness of our results, we consider a more general specification of the matching function. Following Den Haan et al. (2000), we allow the meeting probabilities on both sides of the market to vary with the sex ratio rather than assuming that each woman is guaranteed to meet a partner in the first stage. The general matching function is defined as:

$$M(X,Y) = \frac{XY}{\left(X^{\frac{1}{\omega}} + Y^{\frac{1}{\omega}}\right)^{\omega}} \tag{6}$$

where M denotes the number of matches between two groups, X and Y represent the number of agents on each side of the market (men and women), and $\omega > 0$ is the elasticity of substitution between them. In our context, the formulation becomes:

$$M(N^m, N^f) = \frac{N^m N^f}{\left((N^m)^{\frac{1}{\omega}} + (N^f)^{\frac{1}{\omega}}\right)^{\omega}} \tag{7}$$

Our baseline assumption corresponds to the limiting case as $\omega \to 0$, in which each woman is guaranteed to meet a man in the first stage. By allowing finite values of ω, this specification relaxes the baseline assumption and makes women's meeting probabilities responsive to demographic imbalances.

Quantitative experiments under different values of ω show that our main results remain robust. Although a higher probability of meeting a partner allows women to be more selective, the overall pattern persists: a higher sex ratio reduces assortative matching and, in turn, lowers household income inequality. This confirms that our key insights are not an artifact of the baseline assumption but hold under a broader class of matching technologies. Detailed derivations and numerical results are reported in Appendix E.

4.6 Education decision

In previous sections, the skilled labor shares for both males and females were exogenously calibrated. This section, however, explores how agents choose their education levels. Specifically, it aims to conduct a counterfactual experiment examining the choices of males and females in education, considering their expectations of intensified competition in the marriage market among men and reduced competition among women due to changing sex ratios. To investigate this, we model agents making educational investment decisions before entering the marriage market. We denote the period when agents decide on their education as t and the period when they enter the marriage market competition as $t+1$.

In period t, agents form expectations regarding the utility gains from matches occurring in the following period, $t+1$. These expectations guide their decision to pursue either a skilled or unskilled path. The process of deciding to become a skilled or unskilled agent involves various scenarios, and we have delineated the expected value function for clarity. Initially, we elucidate the expected payoff for a skilled male using a decision tree diagram, followed by a formal expression of the expected payoff equation. The expected payoff equations for the other three agent types are detailed in Appendix F, along with corresponding decision tree diagrams. These decision trees carefully map out each step's potential outcomes, allowing for the calculation of the associated expected utility.

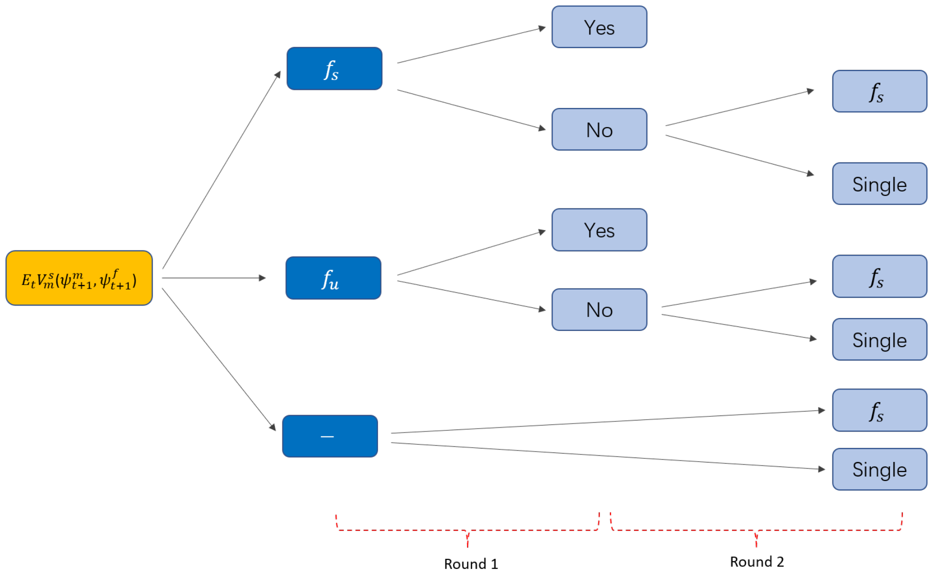


Figure 7: Decision tree for the male who wants to become skilled

Figure 7 presents a decision tree for a man who wants to become skilled. In the first round of matching, he encounters one of three possible outcomes: meeting a skilled female (f_s), an unskilled female (f_u), or no one at all ('-'), a scenario that arises when the number of males ($N^m = 1$) exceeds the number of females (N^f). If he meets a skilled female (f_s), he may either agree to marry ('Yes') or decline the match ('No'), moving to the second round. In the second round, he can either successfully pair with a skilled female and marry or remain single. Similarly, if he meets an unskilled female (f_u), he faces the choice to either form a household with her ('Yes') or refuse ('No') and proceed to the second round. Here, he may marry a skilled female (f_s) or end up single ('Single'). If he meets no one in the first round ('-'), he directly advances to the second round, where he either finds a skilled female to marry or remains single ('Single').

The expected utility of being a skilled male given the fraction of ψ^f_{t+1} skilled female is:

$$
\begin{aligned}
\mathrm{E}_t V^s_m(\psi^m_{t+1}, \psi^f_{t+1}) =& \psi^f_{t+1} N^f_{t+1} \Bigg\{ \int_{\max(\bar{q}^{ss}_m, \bar{q}^{ss}_f)}^{q_{\max}} \left[U\left(e[w^s_m + w^s_f] \right) + q^{ss}_m \right] dq \\
& + Q[\max(\bar{q}^{ss}_m, \bar{q}^{ss}_f)] \min(\frac{\theta^s_f(\psi^m_{t+1}, \psi^f_{t+1})}{\theta^s_m(\psi^m_{t+1}, \psi^f_{t+1})}, 1) \int_0^{q_{\max}} \left[U\left(e[w^s_m + w^s_f] \right) + q^{ss}_m \right] dq \\
& + Q[\max(\bar{q}^{ss}_m, \bar{q}^{ss}_f)] \left[1 - \min(\frac{\theta^s_f(\psi^m_{t+1}, \psi^f_{t+1})}{\theta^s_m(\psi^m_{t+1}, \psi^f_{t+1})}, 1) \right] U[w^s_m] \Bigg\} \\
& + (1 - \psi^f_{t+1}) N^f_{t+1} \Bigg\{ \int_{\max(\bar{q}^{su}_m, \bar{q}^{su}_f)}^{q_{\max}} \left[U\left(e[w^s_m + w^u_f] \right) + q^{su}_m \right] dq \\
& + Q[\max(\bar{q}^{su}_m, \bar{q}^{su}_f)] \min(\frac{\theta^s_f(\psi^m_{t+1}, \psi^f_{t+1})}{\theta^s_m(\psi^m_{t+1}, \psi^f_{t+1})}, 1) \int_0^{q_{\max}} \left[U\left(e[w^s_m + w^s_f] \right) + q^{ss}_m \right] dq \\
& + Q[\max(\bar{q}^{su}_m, \bar{q}^{su}_f)] \left[1 - \min(\frac{\theta^s_f(\psi^m_{t+1}, \psi^f_{t+1})}{\theta^s_m(\psi^m_{t+1}, \psi^f_{t+1})}, 1) \right] U\left[w^s_m \right] \Bigg\} \\
& + (1 - N^f_{t+1}) \Bigg\{ \min(\frac{\theta^s_f(\psi^m_{t+1}, \psi^f_{t+1})}{\theta^s_m(\psi^m_{t+1}, \psi^f_{t+1})}, 1) \int_0^{q_{\max}} \left[U\left(e[w^s_m + w^s_f] \right) + q^{ss}_m \right] dq \\
& + \left[1 - \min(\frac{\theta^s_f(\psi^m_{t+1}, \psi^f_{t+1})}{\theta^s_m(\psi^m_{t+1}, \psi^f_{t+1})}, 1) \right] U\left[w^s_m \right] \Bigg\}
\end{aligned}
\tag{8}
$$

4.6.1 Education decision and sex ratio

Becoming a skilled agent requires some effort ϕ_g ($g \in \{male, female\}$), which is $\in [0, +\infty)$, the effort is i.i.d. for all men and women, with the CDF of Φ_g. An agent i wants to become skilled if:

$$\mathrm{E}V_i^s - \mathrm{E}V_i^u \geq \phi_i \tag{9}$$

Where $\mathrm{E}V_i^s$ is the expected payoff of becoming skilled for agent i. The skilled-unskilled payoff difference is generated when a fraction of ψ is skilled:

$$\begin{aligned} \bar{\phi}(\psi_{t+1}^m) &\equiv \mathrm{E}_t V_m^s(\psi_{t+1}^m) - \mathrm{E}_t V_m^u(\psi_{t+1}^m) \\ \bar{\phi}(\psi_{t+1}^f) &\equiv \mathrm{E}_t V_f^s(\psi_{t+1}^f) - \mathrm{E}_t V_f^u(\psi_{t+1}^f) \end{aligned} \tag{10}$$

For any individual i: when $\phi_i \leq \bar{\phi}$ (the effort that has to be devoted to becoming a skilled individual is smaller or equal to the reservation effort $\bar{\phi}$, an agent want to become skilled).

Without considering the education costs, agents can borrow freely, and contingent on their efforts, the young agent will make the same decision irrespective of their family wealth type (family provides the funding for education). This implies that in equilibrium: a fraction $\Phi(\bar{\phi}_g)$ of individuals would become skilled, then the equilibrium distribution of education fraction for two genders are:

$$\begin{aligned} \psi_{t+1}^m &= \Phi(\bar{\phi}_m) \\ \psi_{t+1}^f &= \Phi(\bar{\phi}_f) \end{aligned} \tag{11}$$

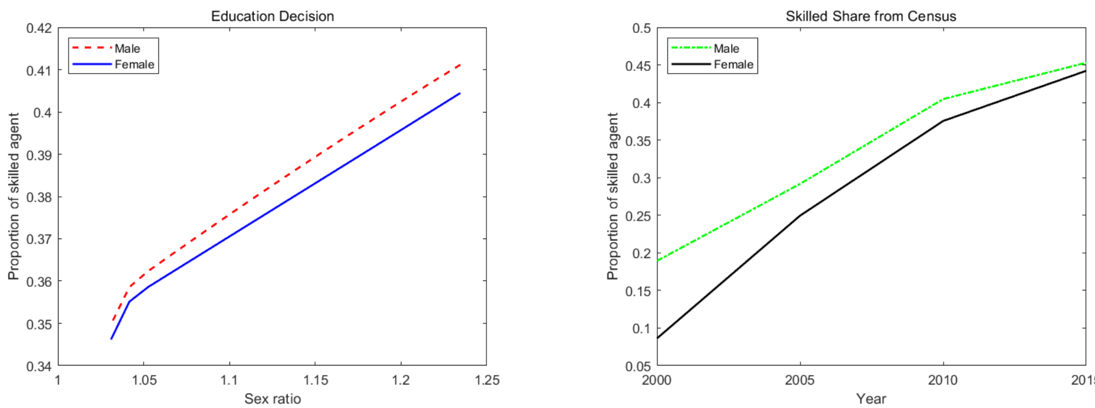
Thus:

$$\begin{aligned} \bar{\phi}(\Phi(\bar{\phi}_m)) &= \mathrm{E}_t V_m^s(\Phi(\bar{\phi}_m)) - \mathrm{E}_t V_m^u(\Phi(\bar{\phi}_m)) \\ \bar{\phi}(\Phi(\bar{\phi}_f)) &= \mathrm{E}_t V_f^s(\Phi(\bar{\phi}_f)) - \mathrm{E}_t V_f^u(\Phi(\bar{\phi}_f)) \end{aligned} \tag{12}$$

We firstly assume that ϕ_g is a uniform distribution for both male and female, which is $\in [0, 1]$. Under this assumption, we have:

$$\psi_{t+1}^{m} = \mathrm{E}_t V_m^s(\psi_{t+1}^{m}) - \mathrm{E}_t V_m^u(\psi_{t+1}^{m})$$

$$\psi_{t+1}^{f} = \mathrm{E}_t V_f^s(\psi_{t+1}^{f}) - \mathrm{E}_t V_f^u(\psi_{t+1}^{f}) \tag{13}$$



(a) Share of skilled agent from model

(b) Share of skilled agent from census data

Figure 8: Education decision and sex ratio

Note: The results presented in the left-hand figure are derived from model outputs, while those on the right-hand side are based on four waves of population census data (census 2000, 2005, 2010, 2015). We measure the fraction of skilled men and women respectively, by restricting cohorts' age between 20 to 35.

After integrating both sides of equation 10 and evaluating it on the whole range of ϕ (that is $\in [0, 1]$), we can get equation 13. After solving it, we generate the results plotted in the Figure 8a. In our matching model, agents of each gender make a binary decision regarding their education level, opting to be skilled or unskilled. This choice leads to an equilibrium distribution in the proportion of skilled agents within the population.

As illustrated in Figure 8a, we find that both men and women increasingly become skilled as the sex ratio in the marriage market increases, where the trend of men is above the trend of women, suggesting that men spend more effort in becoming skilled compared with women. Additionally, when we examine the results of the model alongside the data trends shown in Figure 8b, we observe an upward trajectory in the proportion of skilled men and women over time. This pattern suggests that our model offers insight into the increasing levels of human capital in China, indicating that the increase in male abundance induces higher human capital.

The reasons behind the trend of both men and women becoming more skilled as the sex ratio increases are as follows. For men, firstly, being skilled typically means higher earnings, which can boost their appeal in the increasingly competitive marriage market. Second, when examining the data in Table C1, we see that the probabilities of an unskilled man (π_u) and a skilled man (π_s) of remaining single rise with an increasing sex ratio. However, the likelihood of being single is higher for unskilled men ($\pi_u < \pi_s$), suggesting that it's riskier for them to find a partner. Additionally, should an unskilled man marry, he is more likely to marry an unskilled woman, as indicated by the higher values of (π_{uu}) in comparison to other household types, leading to a lower utility gain in forming a household. These factors motivate men to become skilled.

For women, firstly, choosing to remain unskilled means they are more likely to marry an unskilled husband (as $\pi_{uu} < \pi_{su}$), which generally offers less benefit. Secondly, if a woman becomes skilled, she is more likely to marry a skilled husband than an unskilled one (as $\pi_{ss} > \pi_{us}$). Third, as the sex ratio in the marriage market increases, even though an unskilled woman's chances of marrying a skilled husband increase, the likelihood of forming a household with skilled men is still lower than marrying an unskilled husband (as $\pi_{uu} > \pi_{su}$). Therefore, women are more inclined to become skilled.

This finding offers a potential explanation for the observed increase in education levels in China. Analysis of census data suggests that a growing surplus of men is associated with higher educational attainment for both men and women. In addition, we enrich the literature that has studied the interconnection between premarital investment in education and marriage market competition. Lafortune (2013) finds that males increased their human capital investment when faced with an increase in the sex ratio. Bhaskar and Hopkins (2016) indicate that if there is an unbalanced sex ratio, the abundant sex invests more, while the scarcer one invests less. Additionally, Bhaskar et al. (2023) illustrate that parents of boys invest less in education in regions with a more male-biased sex ratio but they invest more in housing. Our findings enrich these studies, showing that in response to an increasing surplus of males, both men and women enhance their educational investments. Notably, men's investments in education are more pronounced compared to those of women.

5 Conclusion

The increasing gender imbalance in China has gained significant attention from economists. In this study, we investigate whether and how gender imbalance affects household income inequality. Our research suggests that growing gender imbalance may lead to a reduction in assortative mating among Chinese couples, thereby contributing to a decrease in household income inequality within the country. We first present empirical evidence to assess the influence of the sex ratio in the marriage market on household income inequality. Then, we construct a structural model to elucidate the empirical observations and investigate the underlying mechanisms.

Drawing from empirical evidence, we observe a negative correlation between the sex ratio in the marriage market and household income inequality. Specifically, an increase in the sex ratio is linked to a reduction in household income inequality. To deepen our analysis, Section 3 introduces a structural interpretation, in which we use a model to inspect the dynamics of the marriage market and matching outcomes in China.

Theoretically, the results from the model indicate that men become less selective in the marriage market when the male abundance in the marriage market increases. This results in a decrease of assortative matching. This finding is consistent with the literature's conclusion that men tend to marry women from a lower social class in the presence of male abundance. Furthermore, our results consistently show that a rise in the sex ratio is related to a decline in household income inequality, a relationship that is robust across different measures of income inequality. Lastly, we incorporate the endogenization of individuals' educational choices in the context of gender imbalance before they enter into the marriage market. This approach allows us to examine the impact of marriage market competition on educational attainment in China. Our results show that an elevated sex ratio leads to increased educational investments by both men and women, with men investing more significantly than women.

While our analysis focuses on China due to its unique demographic history—including the one-child policy and strong son preference—we believe that the mechanisms uncovered in our study may be applicable to other societies with significant gender imbalances. For example, countries such as India and parts of the Middle East have also experienced persistently high male-to-female sex ratios, which could lead to similar patterns of marriage market competition and its implications for household-level outcomes. Nevertheless, cultural norms and institutional settings may shape how these dynamics manifest, and the magnitude of the effects may differ across contexts. Future research could explore the extent to which our findings generalize beyond China by applying similar structural approaches to other settings.

Our analysis yields several policy relevant insights. First, the evidence suggests that male-biased sex ratios increase marriage market competition among men, which in turn encourages higher educational investment, particularly among men in affected regions. Second, higher sex ratios tend to reduce assortative matching, thereby contributing to lower household income inequality. At the same time, a persistent imbalance may leave a larger share of men unmarried, raising potential social and economic challenges. Policymakers may therefore need to complement the observed inequality-reducing effects with targeted interventions aimed at supporting unmarried and disadvantaged individuals in regions facing sustained gender imbalances.

References

Abramitzky, R., Delavande, A., & Vasconcelos, L. (2011). Marrying up: The role of sex ratio in assortative matching. *American Economic Journal: Applied Economics*, *3*(3), 124–57.

Angrist, J. (2002). How do sex ratios affect marriage and labor markets? evidence from america's second generation. *The Quarterly Journal of Economics*, *117*(3), 997–1038.

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

Becker, G. S. (1974). A theory of marriage: Part ii. *Journal of political Economy*, *82*(2, Part 2), S11–S26.

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

Bhaskar, V., Li, W., & Yi, J. (2023). Multidimensional premarital investments with imperfect commitment. *Journal of Political Economy*, *131*(10), 2893–2919.

Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, *133*(3), 1163–1228.

Cowell, F. (2011). *Measuring inequality*. Oxford University Press.

Den Haan, W. J., Ramey, G., & Watson, J. (2000). Job destruction and propagation of shocks. *American economic review*, *90*(3), 482–498.

Du, J., Wang, Y., & Zhang, Y. (2015). Sex imbalance, marital matching and intra-household bargaining: Evidence from china. *China Economic Review*, *35*, 197–218.

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

Eika, L., Mogstad, M., & Zafar, B. (2019). Educational assortative mating and household income inequality. *Journal of Political Economy*, *127*(6), 2795–2835.

Fernandez, R., Guner, N., & Knowles, J. (2005). Love and money: A theoretical and empirical analysis of household sorting and inequality. *The Quarterly Journal of Economics*, *120*(1), 273–344.

Fernández, R., & Rogerson, R. (2001). Sorting and long-run inequality. *The Quarterly Journal of Economics*, *116*(4), 1305–1341.

Greenwood, J., Guner, N., Kocharkov, G., & Santos, C. (2014). Marry your like: Assortative mating and income inequality. *American Economic Review*, *104*(5), 348–53.

Greenwood, J., Guner, N., Kocharkov, G., & Santos, C. (2016). Technology and the changing family: A unified model of marriage, divorce, educational attainment, and married female labor-force participation. *American Economic Journal: Macroeconomics*, *8*(1), 1–41.

Greenwood, J., Guner, N., & Vandenbroucke, G. (2017). Family economics writ large. *Journal of Economic Literature*, *55*(4), 1346–1434.

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, *11*(4), 181–209.

Lafortune, J. (2013). Making yourself attractive: Pre-marital investments and the returns to education in the marriage market. *American Economic Journal: Applied Economics*, *5*(2), 151–78.

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

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

Wei, S.-J., Zhang, X., & Liu, Y. (2017). Home ownership as status competition: Some theory and evidence. *Journal of Development Economics*, *127*, 169–186.

Yao, Y., Zhang, M., & Li, Y. (2023). Sex ratios and marital matching outcomes in the chinese marriage market. *Economics Letters*, *222*, 110927.

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

Zhang, J. (2021). A survey on income inequality in china. *Journal of Economic Literature*, *59*(4), 1191–1239.

Appendix

Appendix A. Sex ratio at birth and Gini coefficient in China

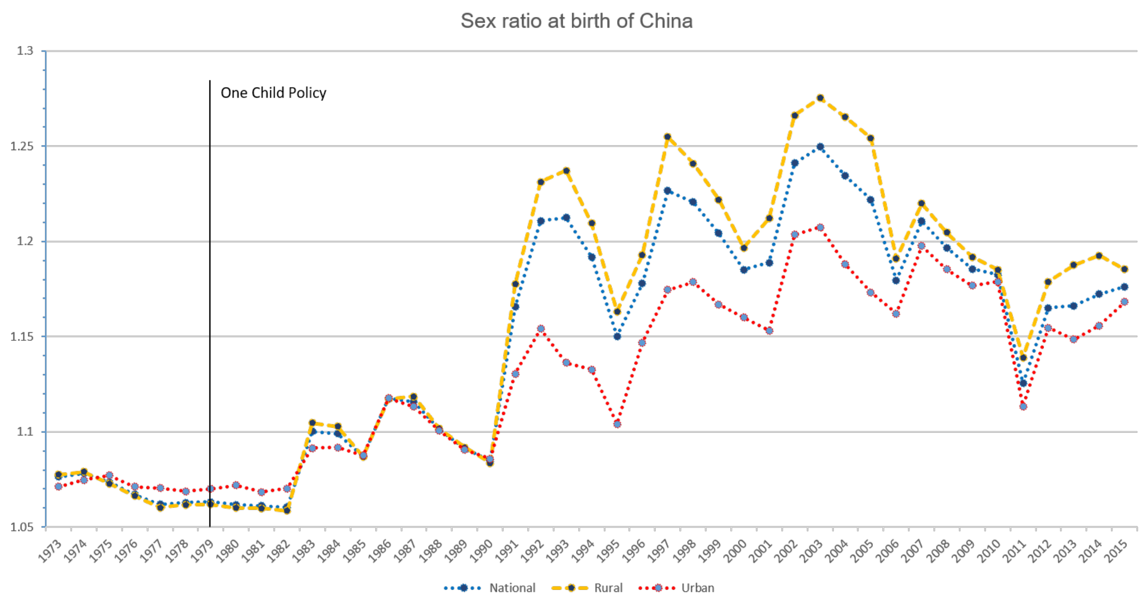


Figure A1: Sex ratio at birth in China

Note: We utilize data from multiple waves of the Chinese population census (1982, 1985,...2015) to calculate the sex ratio at birth in China. Each population census provides data on the number of residents in both rural and urban areas across various age groups. We utilize this information to deduce the sex ratio at birth for different years.

Figure A1 displays the trend of the sex ratio at birth in China, revealing a significant increase following the introduction of the One Child Policy (OCP) in 1979. The connection between the One Child Policy and the sex ratio has been thoroughly examined by researchers such as Ebenstein (2010, 2011) and Zhang (2017), who emphasize that a preference for sons led to the selective abortion of female fetuses, facilitated by advancements in biotechnology, as a consequence of the OCP. This policy skewed the sex ratio, with a more pronounced distortion in rural areas due to a stronger preference for sons. Notably, from 1990 onwards, the sex ratio at birth in China has risen markedly, indicating that two decades later, the marriage market experienced heightened competition among men.

Figure A2 shows the evolution of the Gini coefficient in China, revealing an upward trend in income inequality beginning in 1978, followed by a modest decline in recent years. For a detailed exploration of the changes in income inequality in China, Zhang (2021) provides an extensive survey.

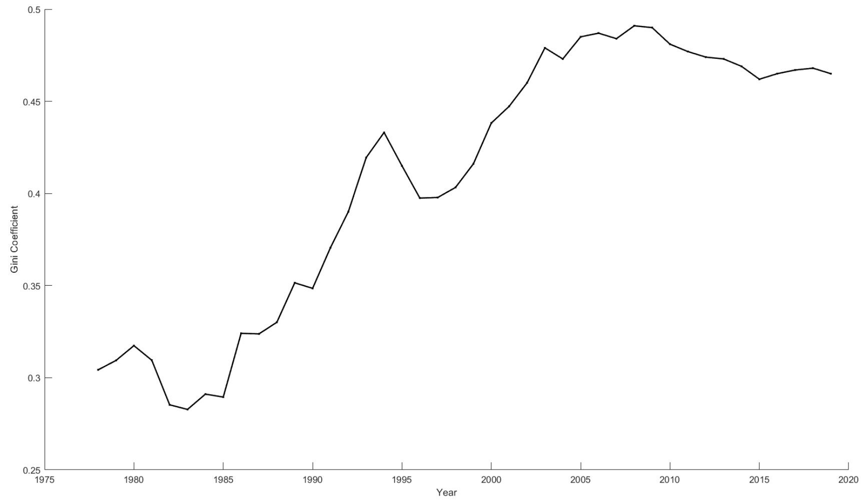


Figure A2: Gini coefficient in China

Note: The Gini coefficient data are sourced from the National Bureau of Statistics.

Appendix B. The remaining part of the model

We illustrate the remaining part of the model in this section.

The number of meetings between individuals of different skill types that take place in the first round is:

$$\begin{array}{ll} N^f\psi_m\psi_f : & \text{the number of meetings between skilled male and skilled female} \\ N^f\psi_m(1-\psi_f) : & \text{the number of meetings between skilled male and unskilled female} \\ N^f(1-\psi_m)\psi_f : & \text{the number of meetings between unskilled male and skilled female} \\ N^f(1-\psi_m)(1-\psi_f) : & \text{the number of meetings between unskilled male and unskilled female} \end{array} \tag{14}$$

The number of unmarried individuals after the first round of matching:

$$\theta_m^s = (1-N^f)\psi_m + N^f\psi_m\psi_f Q\left[\max(\bar{q}_m^{ss},\bar{q}_f^{ss})\right] + N^f\psi_m(1-\psi_f)Q\left[\max(\bar{q}_m^{su},\bar{q}_f^{su})\right]$$

the number of unmarried skilled males after round 1

$$\theta_m^u = (1-N^f)(1-\psi_m) + N^f(1-\psi_m)\psi_f Q\left[\max(\bar{q}_m^{us},\bar{q}_f^{us})\right] + N^f(1-\psi_m)(1-\psi_f)Q\left[\max(\bar{q}_m^{uu},\bar{q}_f^{uu})\right]$$

the number of unmarried unskilled males after round 1

$$\theta_f^s = N^f\psi_m\psi_f Q\left[\max(\bar{q}_m^{ss},\bar{q}_f^{ss})\right] + N^f(1-\psi_m)\psi_f Q\left[\max(\bar{q}_m^{us},\bar{q}_f^{us})\right]$$

the number of unmarried skilled females after round 1

$$\theta_f^u = N^f\psi_m(1-\psi_f)Q\left[\max(\bar{q}_m^{su},\bar{q}_f^{su})\right] + N^f(1-\psi_m)(1-\psi_f)Q\left[\max(\bar{q}_m^{uu},\bar{q}_f^{uu})\right]$$

the number of unmarried unskilled females after round 1

(15)

Appendix B1. The full equilibrium conditions in the model

The indifference condition for a skilled male matched with a skilled female:

$$U[e(w_m^s + w_f^s)] + \bar{q}_m^{ss} = \min(\frac{\theta_f^s}{\theta_m^s}, 1)\left[U[e(w_m^s + w_f^s)] + \mu\right] + \left[1 - \min(\frac{\theta_f^s}{\theta_m^s}, 1)\right] U(w_m^s)$$

The indifference condition for a skilled female matched with a skilled male:

$$U[e(w_m^s + w_f^s)] + \bar{q}_f^{ss} = \min(\frac{\theta_m^s}{\theta_f^s}, 1)\left[U[e(w_m^s + w_f^s)] + \mu\right] + \left[1 - \min(\frac{\theta_m^s}{\theta_f^s}, 1)\right] U(w_f^s)$$

The indifference condition for an unskilled male matched with a skilled female:

$$U[e(w_m^u + w_f^s)] + \bar{q}_m^{us} = \min(\frac{\theta_f^u}{\theta_m^u}, 1)\left[U[e(w_m^u + w_f^u)] + \mu\right] + \left[1 - \min(\frac{\theta_f^u}{\theta_m^u}, 1)\right] U(w_m^u)$$

The indifference condition for a skilled female matched with an unskilled male:

$$U[e(w_m^u + w_f^s)] + \bar{q}_f^{us} = \min(\frac{\theta_m^s}{\theta_f^s}, 1)\left[U[e(w_m^s + w_f^s)] + \mu\right] + \left[1 - \min(\frac{\theta_m^s}{\theta_f^s}, 1)\right] U(w_f^s)$$

The indifference condition for a skilled male matched with an unskilled female:

$$U[e(w_m^s + w_f^u)] + \bar{q}_m^{su} = \min(\frac{\theta_f^s}{\theta_m^s}, 1) \left[U[e(w_m^s + w_f^s)] + \mu\right] + \left[1 - \min(\frac{\theta_f^s}{\theta_m^s}, 1)\right] U(w_m^s)$$

The indifference condition for an unskilled female matched with a skilled male:

$$U[e(w_m^s + w_f^u)] + \bar{q}_f^{su} = \min(\frac{\theta_m^u}{\theta_f^u}, 1) \left[U[e(w_m^u + w_f^u)] + \mu\right] + \left[1 - \min(\frac{\theta_m^u}{\theta_f^u}, 1)\right] U(w_f^u)$$

The indifference condition for an unskilled male matched with an unskilled female:

$$U[e(w_m^u + w_f^u)] + \bar{q}_m^{uu} = \min(\frac{\theta_f^u}{\theta_m^u}, 1) \left[U[e(w_m^u + w_f^u)] + \mu\right] + \left[1 - \min(\frac{\theta_f^u}{\theta_m^u}, 1)\right] U(w_m^u)$$

The indifference condition for an unskilled female matched with an unskilled male:

$$U[e(w_m^u + w_f^u)] + \bar{q}_f^{uu} = \min(\frac{\theta_m^u}{\theta_f^u}, 1) \left[U[e(w_m^u + w_f^u)] + \mu\right] + \left[1 - \min(\frac{\theta_m^u}{\theta_f^u}, 1)\right] U(w_f^u)$$

Appendix B2. The number of different types of household in the model

The number of unskilled husband with skilled wife in the population:

$$\pi_{us} = N^f (1 - \psi_m)\psi_f \left(1 - Q[\max(\bar{q}_m^{us}, \bar{q}_f^{us})]\right)$$

The number of unskilled husband with unskilled wife in the population:

$$\pi_{uu} = N^f (1 - \psi_m)(1 - \psi_f) \left(1 - Q[\max(\bar{q}_m^{uu}, \bar{q}_f^{uu})]\right) + \min(\theta_m^u, \theta_f^u)$$

The number of unmarried skilled male after two rounds of matching in the population:

$$\pi_s = \max(\theta_m^s - \theta_f^s, 0)$$

The number of unmarried unskilled male after two rounds of matching in the population:

$$\pi_u = \max(\theta_m^u - \theta_f^u, 0)$$

Appendix B3. Education correlation between husband and wife

Education correlation between husband and wife ρ is been defined as:

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E(X)^2}\sqrt{E(Y^2) - E(Y)^2}}$$

Where X and Y stands for events $\{husband\}$ and $\{wife\}$ respectively, let:

$$X = \begin{cases} 1 & \text{husband is skilled} \\ 0 & \text{husband is unskilled} \end{cases} \qquad Y = \begin{cases} 1 & \text{wife is skilled} \\ 0 & \text{wife is unskilled} \end{cases} \tag{16}$$

Hence, the expectations of X and Y are:

$$E(X) = \tfrac{\pi_{ss}+\pi_{su}}{N^f} \qquad E(Y) = \tfrac{\pi_{ss}+\pi_{us}}{N^f} \tag{17}$$

And we have:

$$E(X^2) = \tfrac{\pi_{ss}+\pi_{su}}{N^f} \qquad E(Y^2) = \tfrac{\pi_{ss}+\pi_{us}}{N^f} \tag{18}$$

$$E(XY) = \frac{\pi_{ss}}{N^f}$$

Thus:

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E(X)^2}\sqrt{E(Y^2) - E(Y)^2}} = \frac{\frac{\pi_{ss}}{N^f} - \frac{\pi_{ss}+\pi_{su}}{N^f}\frac{\pi_{ss}+\pi_{us}}{N^f}}{\sqrt{(\frac{\pi_{ss}+\pi_{su}}{N^f}) - (\frac{\pi_{ss}+\pi_{su}}{N^f})^2}\sqrt{(\frac{\pi_{ss}+\pi_{us}}{N^f}) - (\frac{\pi_{ss}+\pi_{us}}{N^f})^2}} \tag{19}$$

Appendix C. Additional results in the quantitative analysis

Appendix C1. Reservation match quality

The solved reservation match qualities are reported in Table C1, which illustrates several dynamics of marriage matching in the presence of an increasing sex ratio. First, we find that with the increasing of the sex ratio, all reservation match qualities for males in different types of matches are decreasing, this implies that men are decreasingly selective in all types of marriage due to the male abundance. Second, comparing the reservation match qualities between men and women in four types of matches, we find that men always have lower reservation match quality than women (except for the match of a skilled man meeting an unskilled woman), which suggests that males are less selective than females in those matches.[11] Third, looking at the reservation match quality for skilled woman and unskilled woman respectively, whenever they meet an unskilled man, they always have higher cutoff match quality compared to meeting with a skilled man, i.e., $\bar{q}_f^{ss} < \bar{q}_f^{us}$ and $\bar{q}_f^{su} < \bar{q}_f^{uu}$. Lastly, when looking at the reservation match quality for all women, it indicates that the skilled woman who meets with an unskilled man has the highest reservation match quality compared with other women. The reservation match quality is 0.691, which implies the probability that she rejects the marriage with an unskilled man is 69.1%, intuitively, the match quality for a skilled woman who gets married to an unskilled man has to be higher enough to compensate for the loss in utility, which is the trade-off between love and money.

Table C1: Reservation match quality $\bar{q}_g^{zz}$

N^f	Sex ratio	$\bar{q}_m^{ss}$	$\bar{q}_f^{ss}$	$\bar{q}_m^{us}$	$\bar{q}_f^{us}$	$\bar{q}_m^{su}$	$\bar{q}_f^{su}$	$\bar{q}_m^{uu}$	$\bar{q}_f^{uu}$
0.810	1.235	0.316	0.500	0.040	0.691	0.502	0.265	0.271	0.500
0.820	1.220	0.324	0.500	0.052	0.691	0.510	0.265	0.282	0.500
0.830	1.205	0.332	0.500	0.064	0.691	0.518	0.265	0.294	0.500
0.840	1.190	0.340	0.500	0.076	0.691	0.526	0.265	0.306	0.500
0.850	1.176	0.347	0.500	0.089	0.691	0.533	0.265	0.319	0.500
0.860	1.163	0.355	0.500	0.101	0.691	0.541	0.265	0.331	0.500
0.870	1.149	0.363	0.500	0.114	0.691	0.549	0.265	0.344	0.500
0.880	1.136	0.370	0.500	0.127	0.691	0.556	0.265	0.357	0.500
0.890	1.124	0.378	0.500	0.140	0.691	0.564	0.265	0.371	0.500
0.900	1.111	0.385	0.500	0.154	0.691	0.571	0.265	0.384	0.500
0.910	1.099	0.393	0.500	0.168	0.691	0.579	0.265	0.398	0.500
0.920	1.087	0.400	0.500	0.182	0.691	0.586	0.265	0.412	0.500
0.930	1.075	0.407	0.500	0.196	0.691	0.593	0.265	0.426	0.500
0.940	1.064	0.415	0.500	0.211	0.691	0.601	0.265	0.441	0.500
0.950	1.053	0.422	0.500	0.225	0.691	0.608	0.265	0.456	0.500
0.960	1.042	0.429	0.500	0.240	0.691	0.615	0.265	0.471	0.500
0.970	1.031	0.436	0.500	0.256	0.691	0.622	0.265	0.486	0.500

[11] For the match of a skilled man meets with an unskilled woman, the reservation match quality of the man is higher than the woman, but decreases with the rising of male abundance

Appendix C2. Equilibrium distribution of households

This section presents the results of the equilibrium distribution of households. The results in previous sections suggest that men are less selective and the degree of assortative matching decreases with the rise of sex ratio in the marriage market, these forces impact the formation of households in China, which has implications for household income inequality. Table C2 illustrates the equilibrium distribution of different types of households given different values of the sex ratio. When looking at the singles, it indicates that the number of unskilled men π_u is larger than the number of skilled men π_s; this is because, in most types of matches, men are relatively less selective compared with women.[12] When a skilled woman matches with a skilled man and an unskilled man respectively, she has higher reservation match quality when the man is unskilled (as $\bar{q}_f^{ss} < \bar{q}_f^{us}$, based on the results in Table C1), which means she is more likely to reject the marriage with an unskilled man.

The analysis of couples' distribution indicates that only the number of "*su*" households (where a skilled husband is paired with an unskilled wife) increases with the rising sex ratio in the marriage market. In contrast, for other types of households, their numbers decrease as the sex ratio rises. This finding also implies that males are less selective, while females are more inclined to "marry up". Furthermore, the variation in household formation has significant implications for aggregate economic outcomes, such as income inequality.

Table C2: Equilibrium distribution of households

N^f	Sex ratio	π_{ss}	π_{su}	π_{us}	π_{uu}	π_s	π_u
0.810	1.235	0.2773	0.1074	0.0564	0.3689	0.0683	0.1217
0.820	1.220	0.2808	0.1070	0.0571	0.3752	0.0653	0.1147
0.830	1.205	0.2842	0.1066	0.0578	0.3815	0.0623	0.1077
0.840	1.190	0.2876	0.1061	0.0585	0.3878	0.0593	0.1007
0.850	1.176	0.2910	0.1056	0.0592	0.3942	0.0563	0.0937
0.860	1.163	0.2944	0.1051	0.0599	0.4005	0.0534	0.0866
0.870	1.149	0.2979	0.1046	0.0606	0.4070	0.0505	0.0795
0.880	1.136	0.3013	0.1040	0.0613	0.4134	0.0477	0.0723
0.890	1.124	0.3047	0.1034	0.0620	0.4199	0.0449	0.0651
0.900	1.111	0.3081	0.1028	0.0627	0.4264	0.0421	0.0579
0.910	1.099	0.3116	0.1021	0.0634	0.4330	0.0393	0.0507
0.920	1.087	0.3150	0.1014	0.0641	0.4395	0.0366	0.0434
0.930	1.075	0.3184	0.1007	0.0647	0.4461	0.0339	0.0361
0.940	1.064	0.3218	0.1000	0.0654	0.4527	0.0312	0.0288
0.950	1.053	0.3253	0.0992	0.0661	0.4594	0.0285	0.0215
0.960	1.042	0.3287	0.0984	0.0668	0.4661	0.0259	0.0141
0.970	1.031	0.3321	0.0976	0.0675	0.4728	0.0233	0.0067

[12] The results in Table C1 show that except for the match of a skilled man meeting an unskilled woman, the reservation match qualities of men are lower than women in other matches

Appendix D. Constructing the Gini coefficient and Lorenz curve

This section offers an in-depth explanation of how the Gini coefficient and Lorenz curve are constructed. We begin by establishing the cumulative share of household income. Let I_i represent the total income across various household types, we have:

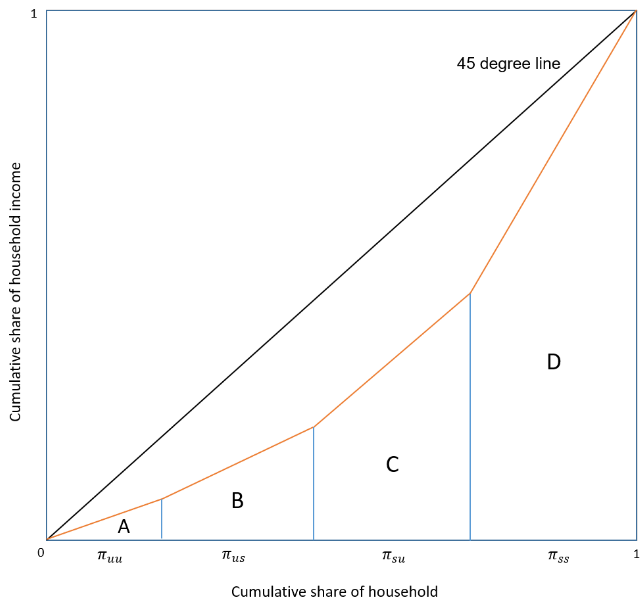


Figure D1: Gini coefficient and Lorenz curve

$$
\begin{aligned}
I_1 &= I_{ss} = \pi_{ss}(w_m^s + w_f^s) \\
I_2 &= I_{su} = \pi_{su}(w_m^s + w_f^u) \\
I_3 &= I_{us} = \pi_{us}(w_m^u + w_f^s) \\
I_4 &= I_{uu} = \pi_{uu}(w_m^u + w_f^u) \\
I &= \sum I_i
\end{aligned}
\tag{20}
$$

Let χ_{zz} denotes the share of income for household type zz over total population, then we have:

$$
\begin{aligned}
\chi_{ss} &= \tfrac{I_1}{I} \\
\chi_{su} &= \tfrac{I_2}{I} \\
\chi_{us} &= \tfrac{I_3}{I} \\
\chi_{uu} &= \tfrac{I_4}{I}
\end{aligned}
\tag{21}
$$

For the share of different types of households, we have:

$$\pi_{ss} + \pi_{su} + \pi_{us} + \pi_{uu} = N^f$$

Let λ_{zz} denotes the share of different types of households over total population, we set:

$$\begin{aligned} \lambda_{ss} &= \tfrac{\pi_{ss}}{N^f} \\ \lambda_{su} &= \tfrac{\pi_{su}}{N^f} \\ \lambda_{us} &= \tfrac{\pi_{us}}{N^f} \\ \lambda_{uu} &= \tfrac{\pi_{uu}}{N^f} \end{aligned} \tag{22}$$

Following Greenwood et al. (2014), we construct the Lorenz curve and the Gini coefficient as follow: the Gini coefficient associated with the Lorenz curve equals twice the area between the Lorenz curve and the 45-degree line. The Gini coefficient can be computed as: $1 - 2\Delta$, where Δ is the area below the Lorenz curve. In this research, we divide Δ into four areas, which are A,B,C,D respectively. Thus, the Gini coefficient g is:

$$g = 1 - 2\left(A + B + C + D\right) \tag{23}$$

Plugging in the coordinates we have:

$$\begin{aligned} area_A &= \tfrac{\lambda_{uu}\chi_{uu}}{2} \\ area_B &= \tfrac{\lambda_{us}[\chi_{uu} + (\chi_{uu} + \chi_{us})]}{2} \\ area_C &= \tfrac{\lambda_{su}[(\chi_{uu} + \chi_{us}) + (\chi_{uu} + \chi_{us} + \chi_{su})]}{2} \\ area_D &= \tfrac{\lambda_{ss}[(\chi_{uu} + \chi_{us} + \chi_{su}) + (\chi_{uu} + \chi_{us} + \chi_{su} + \chi_{ss})]}{2} \end{aligned} \tag{24}$$

Thus, the Gini coefficient is:

$$g = 1 - 2\sum_{j=A}^{D} area_j \tag{25}$$

Appendix E. Robustness check of the model: alternative matching function

In the baseline model, we assumed that in the first stage of the marriage market, each woman is guaranteed to meet a man, while some men remain unmatched due to a male-biased sex ratio. This assumption implies that the probability of women meeting a partner is fixed at one and does not vary with the sex ratio. This may shut down an important countervailing force: when the number of men relative to women increases, women's probability of meeting a partner should also increase, which in turn makes them more selective in the first stage. To address this concern, we extend our analysis by adopting a more general matching function in which the meeting probabilities of both men and women vary with the sex ratio.

Following Den Haan et al. (2000), we consider the matching function:

$$M(X,Y) = \frac{XY}{\left(X^{\frac{1}{\omega}} + Y^{\frac{1}{\omega}}\right)^{\omega}} \tag{26}$$

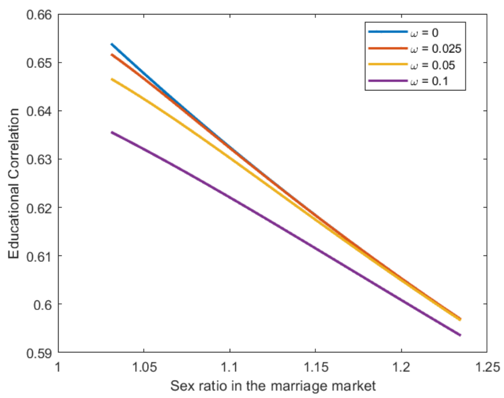
where M is the number of matches, X and Y are the number of agents on each side of the market (men and women), and $\omega > 0$ denotes the elasticity of substitution between them. In our context, we apply:

$$M(N^m, N^f) = \frac{N^m N^f}{\left((N^m)^{\frac{1}{\omega}} + (N^f)^{\frac{1}{\omega}}\right)^{\omega}} \tag{27}$$

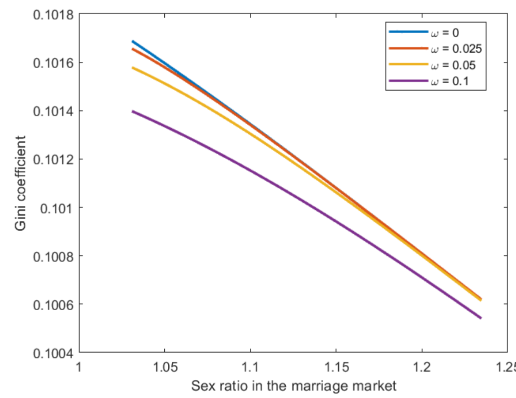
This formulation encompasses our baseline case as a special example. Specifically, as $\omega \to 0$, the matching probability for women converges to 1, recovering our original assumption.

To evaluate the quantitative implications, we simulate the model under different values of ω. The results show that our main findings remain robust. The overall mechanism persists: as the sex ratio increases, assortative matching declines, and household income inequality is reduced. Figures E1a-E1c report the outcomes for educational correlation, the Gini coefficient, and other inequality measures across different values of ω. The qualitative patterns are consistent with the baseline results, though the magnitude of the effects varies with the parameter choice.

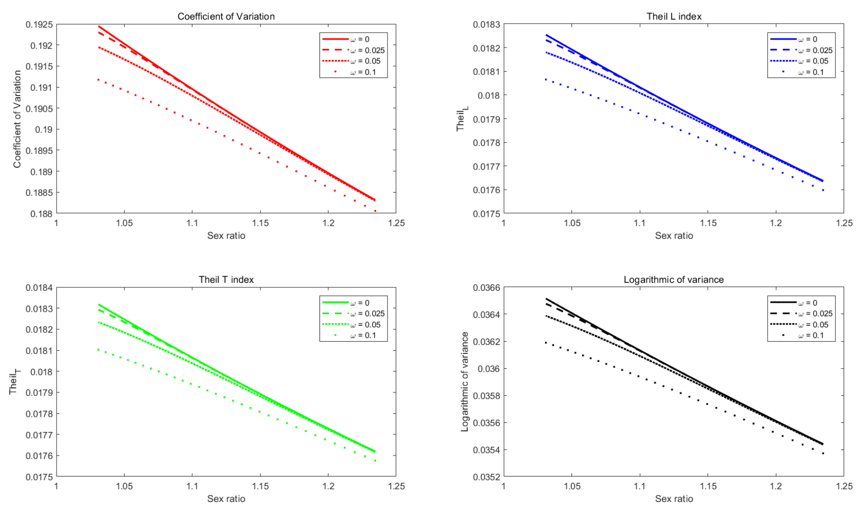
Overall, this robustness exercise demonstrates that our central conclusion, that a male-biased sex ratio reduces assortative matching and lowers household income inequality, is not tied to the special case assumed in the baseline model. Instead, the mechanism holds more generally under alternative and more flexible specifications of the matching function.



(a) Assortative matching vs. sex ratio (ω varies)



(b) Gini vs. sex ratio (ω varies)



(c) Other inequality measures vs. sex ratio (ω varies)

Figure E1: Robustness with a general matching function: assortative matching, and inequality as ω varies.

Appendix F. Education decision

This section outlines the decision tree for being an unskilled male, a skilled female, and an unskilled female, subsequently providing the explicit expressions for the expected payoff equations.

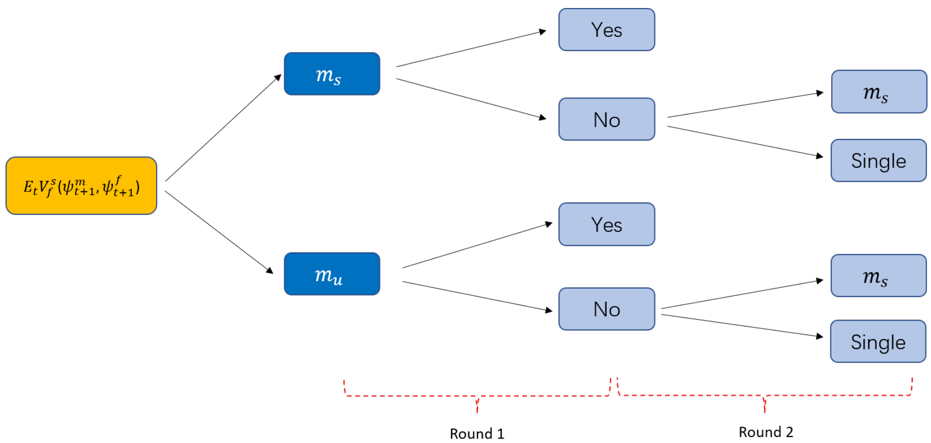


Figure F1: Decision tree for the female who wants to become skilled

Figure F1 presents a decision tree for a female who wants to become skilled. In the first round of matching, she encounters one of two possible outcomes: meeting a skilled man (m_s) or an unskilled man (m_u). If she meets a skilled man (m_s), she may either agree to marry ('Yes') or decline the match ('No'), moving to the second round. In the second round, she can either successfully pair with a skilled man and marry or remain single. Similarly, if she meets an unskilled man (m_u) in the first round, she faces the choice to either form a household with him ('Yes') or refuse ('No') and proceed to the second round. Here, she may marry a skilled man (m_s) or end up single ('Single').

The expected utility of being a skilled female given the fraction of ψ_{t+1}^m skilled male is:

$$\begin{aligned}
\mathrm{E}_t V_f^s(\psi_{t+1}^f, \psi_{t+1}^m) = \min\left(\frac{\psi_{t+1}^m}{N_{t+1}^f}, 1\right) \Bigg\{ & \int_{\max(\bar{q}_m^{ss}, \bar{q}_f^{ss})}^{q_{\max}} \left[U\left(e[w_m^s + w_f^s]\right) + q_f^{ss}\right] dq \\
& + Q[\max(\bar{q}_m^{ss}, \bar{q}_f^{ss})] \min(\frac{\theta_m^s(\psi_{t+1}^m, \psi_{t+1}^f)}{\theta_f^s(\psi_{t+1}^m, \psi_{t+1}^f)}, 1) \int_0^{q_{\max}} \left[U\left(e[w_m^s + w_f^s]\right) + q_f^{ss}\right] dq \\
& + Q[\max(\bar{q}_m^{ss}, \bar{q}_f^{ss})] \left[1 - \min(\frac{\theta_m^s(\psi_{t+1}^m, \psi_{t+1}^f)}{\theta_f^s(\psi_{t+1}^m, \psi_{t+1}^f)}, 1)\right] U\left[w_f^s\right] \Bigg\} \\
+ \min\left(\frac{1 - \psi_{t+1}^m}{N_{t+1}^f}, 1\right) \Bigg\{ & \int_{\max(\bar{q}_m^{us}, \bar{q}_f^{us})}^{q_{\max}} \left[U\left(e[w_m^u + w_f^s]\right) + q_f^{us}\right] dq \\
& + Q[\max(\bar{q}_m^{us}, \bar{q}_f^{us})] \min(\frac{\theta_m^s(\psi_{t+1}^m, \psi_{t+1}^f)}{\theta_f^s(\psi_{t+1}^m, \psi_{t+1}^f)}, 1) \int_0^{q_{\max}} \left[U\left(e[w_m^s + w_f^s]\right) + q_f^{ss}\right] dq \\
& + Q[\max(\bar{q}_m^{us}, \bar{q}_f^{us})] \left[1 - \min(\frac{\theta_m^s(\psi_{t+1}^m, \psi_{t+1}^f)}{\theta_f^s(\psi_{t+1}^m, \psi_{t+1}^f)}, 1)\right] U\left[w_f^s\right] \Bigg\}
\end{aligned} \tag{28}$$

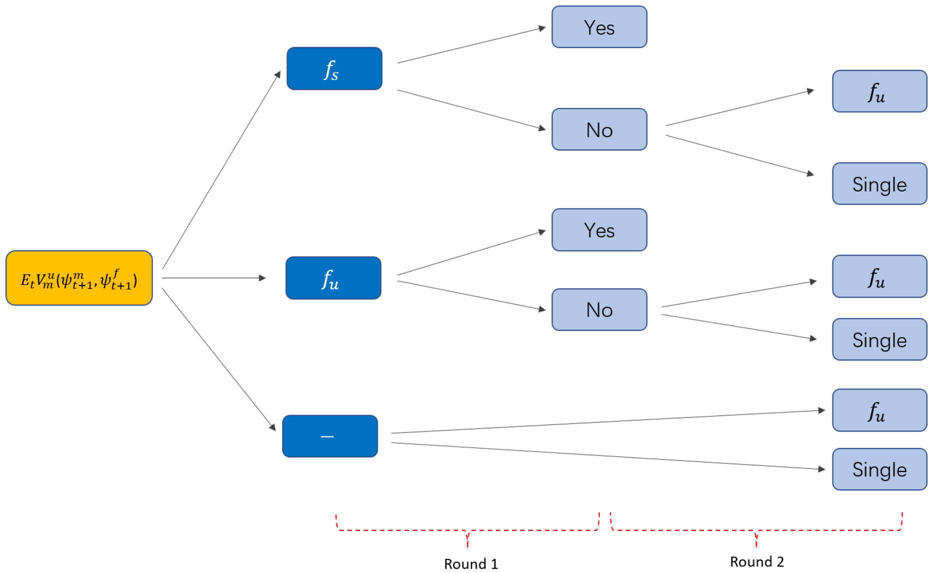


Figure F2: Decision tree for a man who wants to become unskilled

Figure F2 presents a decision tree for a man who wants to become unskilled. In the first round of matching, he encounters one of three possible outcomes: meeting a skilled female (f_s), an unskilled female (f_u), or no one at all ('-'), a scenario that arises when the number of males ($N^m = 1$) exceeds the number of females (N^f). If he meets a skilled female (f_s), he may either agree to marry ('Yes') or decline the match ('No'), moving to the second round. In the second round, he can either successfully pair with an unskilled female and marry or remain single. Similarly, if he meets an unskilled female (f_u), he faces the choice to either form a household with her ('Yes') or refuse ('No') and proceed to the second round. Here, he may marry an unskilled female (f_u) or end up single ('Single'). If he meets no one in the first round ('-'), he directly advances to the second round, where he either finds an unskilled female to marry or remains single ('Single').

The expected utility of being an unskilled male given the fraction of ψ_{t+1}^{f} skilled female is:

$$
\begin{aligned}
\mathrm{E}_t V_m^u\left(\psi_{t+1}^m,\psi_{t+1}^f\right) =& \psi_{t+1}^f N_{t+1}^f \left\{ \int_{\max(\bar{q}_m^{us},\bar{q}_f^{us})}^{q_{\max}} \left[U\left(e[w_m^u + w_f^s]\right) + q_m^{us}\right] dq \right. \\
& + Q[\max(\bar{q}_m^{us},\bar{q}_f^{us})] \min(\frac{\theta_f^u(\psi_{t+1}^m,\psi_{t+1}^f)}{\theta_m^u(\psi_{t+1}^m,\psi_{t+1}^f)},1) \int_0^{q_{\max}} \left[U\left(e[w_m^u + w_f^u]\right) + q_m^{uu}\right] dq \\
& \left. + Q[\max(\bar{q}_m^{us},\bar{q}_f^{us})] \left[1 - \min(\frac{\theta_f^u(\psi_{t+1}^m,\psi_{t+1}^f)}{\theta_m^u(\psi_{t+1}^m,\psi_{t+1}^f)},1)\right] U\left[w_m^u\right] \right\} \\
& + (1-\psi_{t+1}^f) N_{t+1}^f \left\{ \int_{\max(\bar{q}_m^{uu},\bar{q}_f^{uu})}^{q_{\max}} \left[U\left(e[w_m^u + w_f^u]\right) + q_m^{uu}\right] dq \right. \\
& + Q[\max(\bar{q}_m^{uu},\bar{q}_f^{uu})] \min(\frac{\theta_f^u(\psi_{t+1}^m,\psi_{t+1}^f)}{\theta_m^u(\psi_{t+1}^m,\psi_{t+1}^f)},1) \int_0^{q_{\max}} \left[U\left(e[w_m^u + w_f^u]\right) + q_m^{uu}\right] dq \\
& \left. + Q[\max(\bar{q}_m^{uu},\bar{q}_f^{uu})] \left[1 - \min(\frac{\theta_f^u(\psi_{t+1}^m,\psi_{t+1}^f)}{\theta_m^u(\psi_{t+1}^m,\psi_{t+1}^f)},1)\right] U\left[w_m^u\right] \right\} \\
& + (1-N_{t+1}^f) \left\{ \min(\frac{\theta_f^u(\psi_{t+1}^m,\psi_{t+1}^f)}{\theta_m^u(\psi_{t+1}^m,\psi_{t+1}^f)},1) \int_0^{q_{\max}} \left[U\left(e[w_m^u + w_f^u]\right) + q_m^{uu}\right] dq \right. \\
& \left. + \left[1 - \min(\frac{\theta_f^u(\psi_{t+1}^m,\psi_{t+1}^f)}{\theta_m^u(\psi_{t+1}^m,\psi_{t+1}^f)},1)\right] U\left[w_m^u\right] \right\}
\end{aligned} \tag{29}
$$

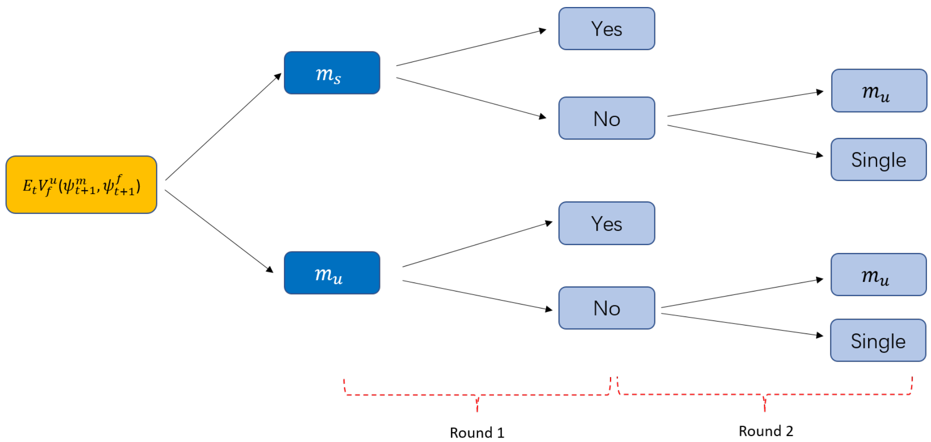


Figure F3: Decision tree for a woman who wants to become unskilled

Figure F3 presents a decision tree for a female who wants to become unskilled. In the first round of matching, she encounters one of two possible outcomes: meeting a skilled man (m_s) or an unskilled man (m_u). If she meets a skilled man (m_s), she may either agree to marry ('Yes') or decline the match ('No'), moving to the second round. In the second round, she can either successfully pair with an unskilled man and marry or remain single. Similarly, if she meets an unskilled man (m_u) in the first round, she faces the choice to either form a household with him ('Yes') or refuse ('No') and proceed to the second round. Here, she may marry an unskilled man (m_u) or end up single ('Single').

The expected utility of being an unskilled female given the fraction of ψ_{t+1}^m skilled male is:

$$
\begin{aligned}
\mathrm{E}_t V_f^u\left(\psi_{t+1}^f, \psi_{t+1}^m\right) = & \min\left(\frac{\psi_{t+1}^m}{N_{t+1}^f}, 1\right)\Bigg\{\int_{\max(\bar{q}_m^{su}, \bar{q}_f^{su})}^{q_{\max}} \left[U\left(e[w_m^s + w_f^u]\right) + q_f^{su}\right] dq \\
& + Q[\max(\bar{q}_m^{su}, \bar{q}_f^{su})] \min(\frac{\theta_m^u(\psi_{t+1}^m, \psi_{t+1}^f)}{\theta_f^u(\psi_{t+1}^m, \psi_{t+1}^f)}, 1) \int_0^{q_{\max}} \left[U\left(e[w_m^u + w_f^u]\right) + q_f^{uu}\right] dq \\
& + Q[\max(\bar{q}_m^{su}, \bar{q}_f^{su})]\left[1 - \min(\frac{\theta_m^u(\psi_{t+1}^m, \psi_{t+1}^f)}{\theta_f^u(\psi_{t+1}^m, \psi_{t+1}^f)}, 1)\right] U\left[w_f^u\right]\Bigg\} \\
& + \min\left(\frac{1-\psi_{t+1}^m}{N_{t+1}^f}, 1\right)\Bigg\{\int_{\max(\bar{q}_m^{uu}, \bar{q}_f^{uu})}^{q_{\max}} \left[U\left(e[w_m^u + w_f^u]\right) + q_f^{uu}\right] dq \\
& + Q[\max(\bar{q}_m^{uu}, \bar{q}_f^{uu})] \min(\frac{\theta_m^u(\psi_{t+1}^m, \psi_{t+1}^f)}{\theta_f^u(\psi_{t+1}^m, \psi_{t+1}^f)}, 1) \int_0^{q_{\max}} \left[U\left(e[w_m^u + w_f^u]\right) + q_f^{uu}\right] dq \\
& + Q[\max(\bar{q}_m^{uu}, \bar{q}_f^{uu})]\left[1 - \min(\frac{\theta_m^u(\psi_{t+1}^m, \psi_{t+1}^f)}{\theta_f^u(\psi_{t+1}^m, \psi_{t+1}^f)}, 1)\right] U\left[w_f^u\right]\Bigg\}
\end{aligned} \tag{30}
$$