Gender-biased Dynastic Savings and the Marriage Market: Evidence from China

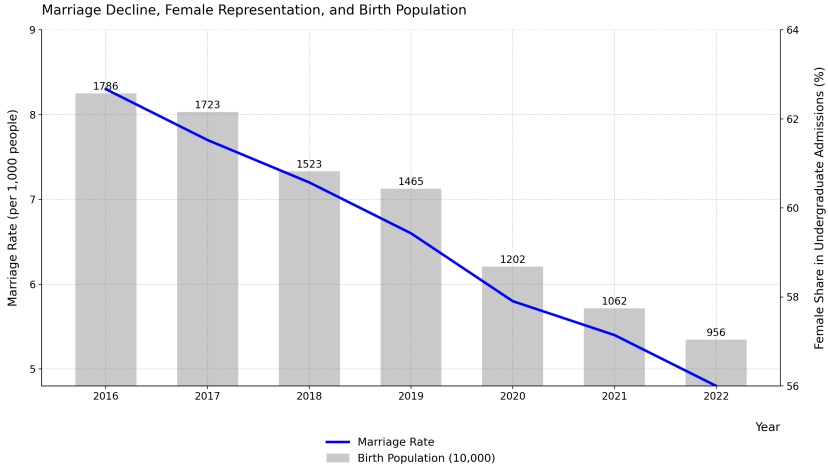
Qilin Zhang

Department of Economics
University of Florida

November 3, 2026

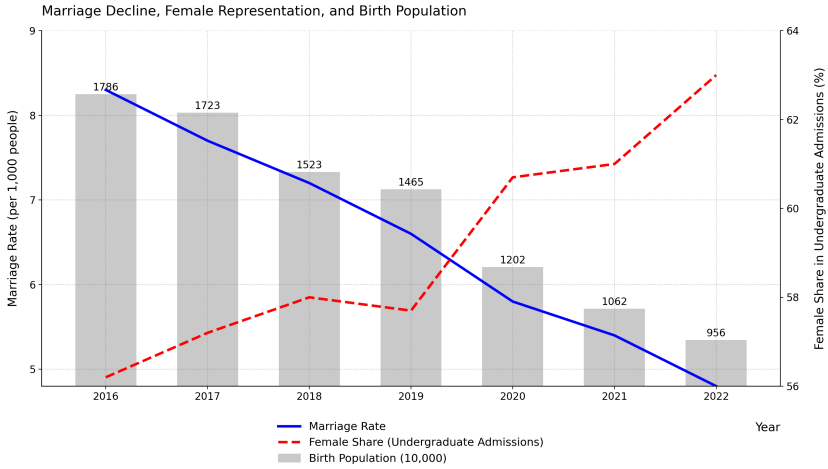
Motivation: From fertility to marriage

- Most children are born within marriage (93-98% Asian countries, 95% China).



Motivation: From fertility to marriage

- Falling marriage and the reversal of the education gender gap



Background: Marriage in China

- Marriage remains assortative: "like marries like".

Background: Marriage in China

- Marriage remains assortative: "like marries like".
- Financial capital matters for marriage, especially for men.

Background: Marriage in China

- Marriage remains assortative: "like marries like".
- Financial capital matters for marriage, especially for men.
- Human capital reflects individual quality, and single women are more educated. Figure

Background: Marriage in China

- Marriage remains assortative: "like marries like".
- Financial capital matters for marriage, especially for men.
- Human capital reflects individual quality, and single women are more educated. Figure
- Parents invest in children's financial capital (savings) and human capital (education), increasing their value of being single and helping them attract better matches.

This paper

Parents invest in children's financial capital (savings) and human capital (education)

This paper

Parents invest in children's financial capital (savings) and human capital (education)

Q1. How does the investment mix differ by child gender?

Q2. Does this gender-biased investment matter for the marriage market?

This paper

Parents invest in children's financial capital (savings) and human capital (education)

Q1. How does the investment mix differ by child gender?

Q2. Does this gender-biased investment matter for the marriage market?

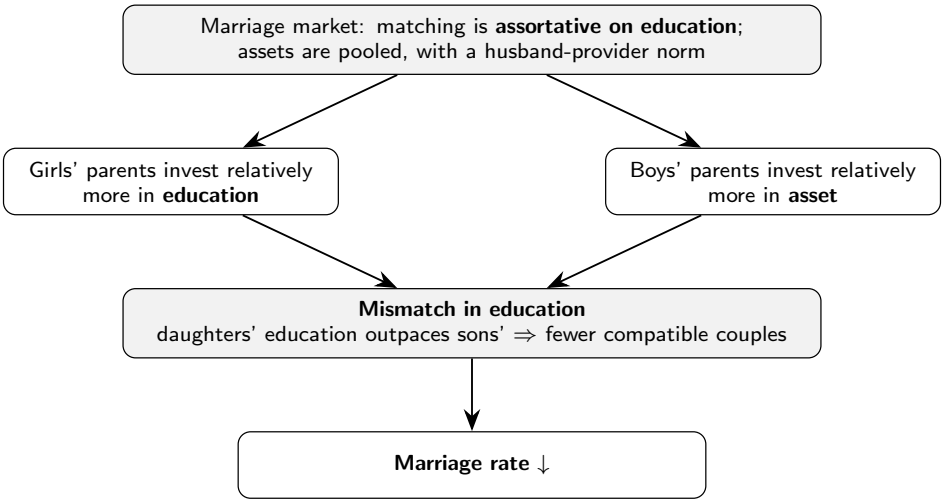
What I do

- Document new facts on marriage matching and parental investment.
- Develop a dynastic model of parental investment in competitive marriage markets.
- Conduct counterfactuals on economic growth, parental subsidies, and marriage subsidies.

Related literature and contribution

- **Gender and parental investment in children** (Qian, 2008; Goldin et al., 2006; Geng and Qin, 2019)
  $\Rightarrow$ Gender bias operates across margins: education for daughters, financial assets for sons.
- **Marriage matching and premarital investment** (Choo and Siow, 2006; Chiappori et al., 2009; Bhaskar et al., 2023)
  $\Rightarrow$ Parents' investment shapes who marries whom.
- **Saving for children's marriage** (Wei and Zhang, 2011; Du and Wei, 2013; Boar, 2021)
  $\Rightarrow$ Marriage incentives shape both the level and the composition of dynastic investment.
- **The decline of marriage** (Greenwood et al., 2016; Fang et al., 2026; Chambers et al., 2026)
  $\Rightarrow$ Quantify how much of China's marriage decline comes from rising parental resources.

My Story


Marriage market: matching is **assortative on education**;
assets are pooled, with a husband-provider norm
Girls' parents invest relatively more in **education**
Boys' parents invest relatively more in **asset**
Mismatch in education
daughters' education outpaces sons' ⇒ fewer compatible couples
Marriage rate ↓

Preview of findings

Q1. How does the investment mix differ by child gender?

Q2. Does this gender-biased investment matter for the marriage market?

Preview of findings

Q1. How does the investment mix differ by child gender?

- The education–income elasticity is consistently higher for daughters.
- Parents invest relatively more in education for daughters and assets for sons.
- The investment gap reflects assortative matching and gender-specific education technology.

Q2. Does this gender-biased investment matter for the marriage market?

Preview of findings

Q1. How does the investment mix differ by child gender?

- The education–income elasticity is consistently higher for daughters.
- Parents invest relatively more in education for daughters and assets for sons.
- The investment gap reflects assortative matching and gender-specific education technology.

Q2. Does this gender-biased investment matter for the marriage market?

- Yes, and it lowers the marriage rate.
- Closing the gender difference in education technology raises the marriage rate.
- The model explains 35% of the drop in marriages from 2010 to 2024. By 2035, marriage falls 33% below its 2010 level, as the model predicts.
- Transfers to parents lower the marriage rate, while equal-sized subsidies to newlyweds raise it.

Overview

1. Data and Stylized Facts
2. Model
3. Counterfactuals
4. Conclusion

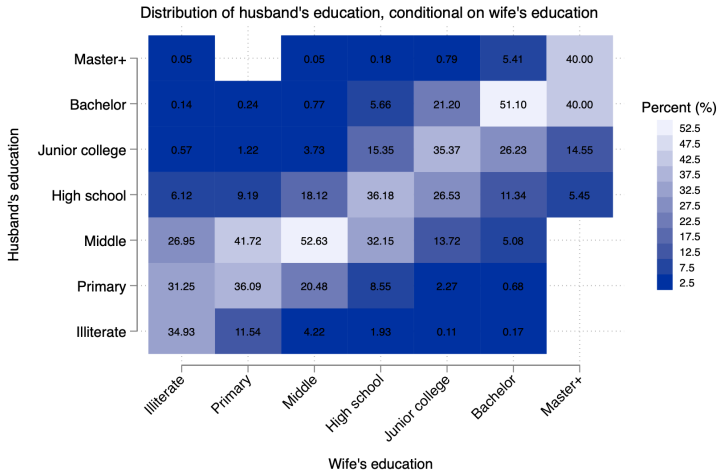
Data: CFPS 2010–2022

- **CFPS**: nationally representative panel of individuals and households in China.
- **Family links**: link children to parents to measure parental investment and family resources.
- **Marriage links**: link spouses and their families to construct the marriage-market sample.
- **Model**: linked data provide moments for estimation and calibration.

data construction summary statistics

Fact 1: Assortative matching in education

- Spouses have similar education: 86% of couples share the same education level (high/low).



matching by gender

Fact 2: parents invest more in daughters' education at the margin

$$\ln e_{ift} = \beta_0 + \beta_1 Male_i + \beta_2 Sch_{it} + \beta_3 \left(Male_i \times Sch_{it}\right) + \beta_4 \ln y_{ft} + \beta_5 \left(Male_i \times \ln y_{ft}\right) + \gamma' X_{ift} + \lambda_t + \varepsilon_{ift}$$

- e_{ift}: parental educational expenditure on child i in family f; $Male_i$: son; Sch_{it}: years of schooling; y_{ft}: household income.
- X_{ift}: age, age^2, mother's years of schooling, urban residence, household equivalence scale; only-child indicator in the full samples.
- $\beta_5 < 0$: spending responds less to income for sons than for daughters.
- No family FE: the question is how parents with different resources invest in sons and daughters, so the relevant variation is between families (in the only-child samples, gender also varies only across families).
- Standard errors clustered at the family level.

with family FE regression samples

Fact 2: parents invest more in daughters' education at the margin

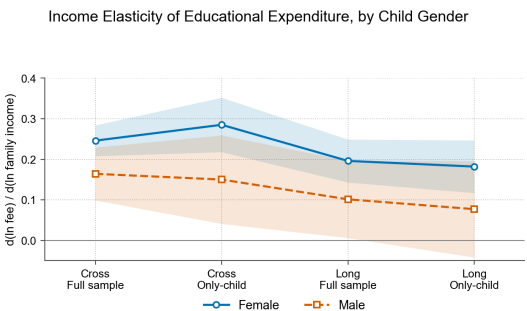
Male $\times \ln y < 0$ in all four samples; full cross-sectional sample: elasticity 0.25 for daughters, 0.16 for sons.

	Cross-sectional sample		Longitudinal sample	
	Full sample (1)	Only child (2)	Full sample (3)	Only child (4)
Male	0.450	0.915*	0.807*	0.939*
	(0.291)	(0.476)	(0.420)	(0.532)
Sch	0.182***	0.135***	0.215***	0.193***
	(0.012)	(0.020)	(0.018)	(0.021)
Male $\times$ *Sch*	0.036***	0.041***	0.017*	0.010
	(0.006)	(0.010)	(0.010)	(0.012)
$\ln y$	0.246***	0.285***	0.196***	0.182***
	(0.019)	(0.034)	(0.027)	(0.033)
Male $\times \ln y$	−0.082***	−0.135***	−0.095**	−0.105**
	(0.027)	(0.044)	(0.040)	(0.051)
Observations	28,299	9,407	8,032	5,004
R^2	0.277	0.231	0.360	0.338
Family FE	No	No	No	No
Year FE	Yes	Yes	Yes	Yes

Notes: The dependent variable is the log of parental educational expenditure on the child, $\ln e$. *Male* indicates a son, *Sch* is the child's years of schooling, and $\ln y$ is log family income. Controls include age, age squared, mother's years of schooling, urban residence, the household equivalence scale $1 + 0.7 \times (\text{adults} - 1) + 0.5 \times$ children, and, in the full samples, an only-child indicator. Columns (1)–(2) use the cross-sectional child sample and columns (3)–(4) the longitudinal child sample; columns (2) and (4) restrict the sample to only children. Standard errors clustered at the family level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

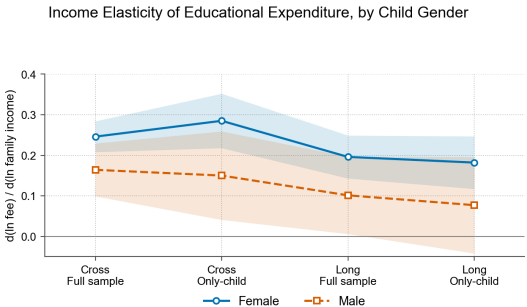
Fact 2: parents invest more in daughters' education at the margin

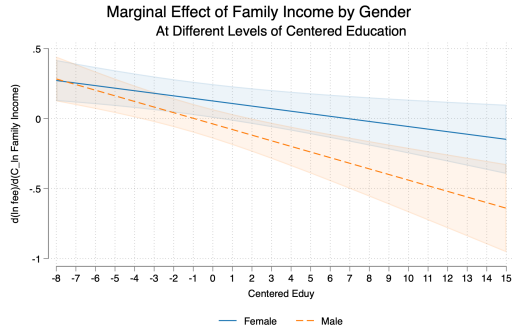
- As family income rises, educational expenditure increases more for daughters.



Fact 2: parents invest more in daughters' education at the margin

- As family income rises, educational expenditure increases more for daughters.





income × education regression

saving regression

Fact 3: Parents save more for sons' marriage

- Households with sons save more than households with daughters, in both rural and urban areas.

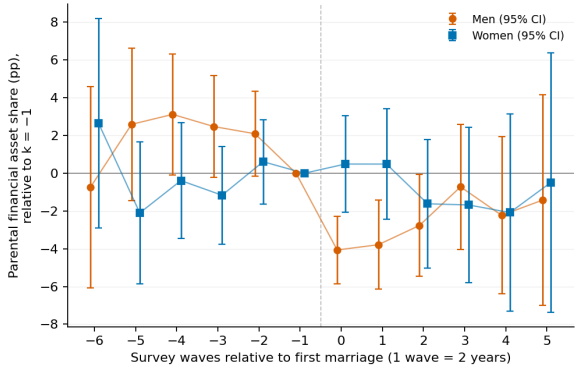
Household Type	Mean	Median	Max	Min	SD	Observations
Rural:						
Only son	0.392	0.460	4.315	-7.581	1.081	1,444
Only daughter	0.316	0.347	3.337	-4.531	0.970	519
Families with boys	0.413	0.457	5.181	-9.748	1.094	3,641
Families with girls	0.255	0.321	3.938	-9.368	1.078	2,408
All families	0.350	0.397	5.181	-9.748	1.091	6,049
Urban:						
Only son	0.428	0.480	3.688	-6.583	0.912	2,023
Only daughter	0.293	0.339	3.972	-5.663	0.843	1,251
Families with boys	0.414	0.480	5.181	-9.002	0.948	3,413
Families with girls	0.271	0.323	3.972	-5.663	0.884	2,745
All families	0.351	0.407	5.181	-9.002	0.923	6,158

Note: This table reports summary statistics for household savings by children's gender composition and urban/rural status among households with unmarried children aged 20–40. The savings rate is defined as log(income/consumption). "Only son" and "Only daughter" refer to households with only one unmarried child of the corresponding gender. "Families with boys" and "Families with girls" refer to households with at least one unmarried son or daughter, respectively.

event-study specification

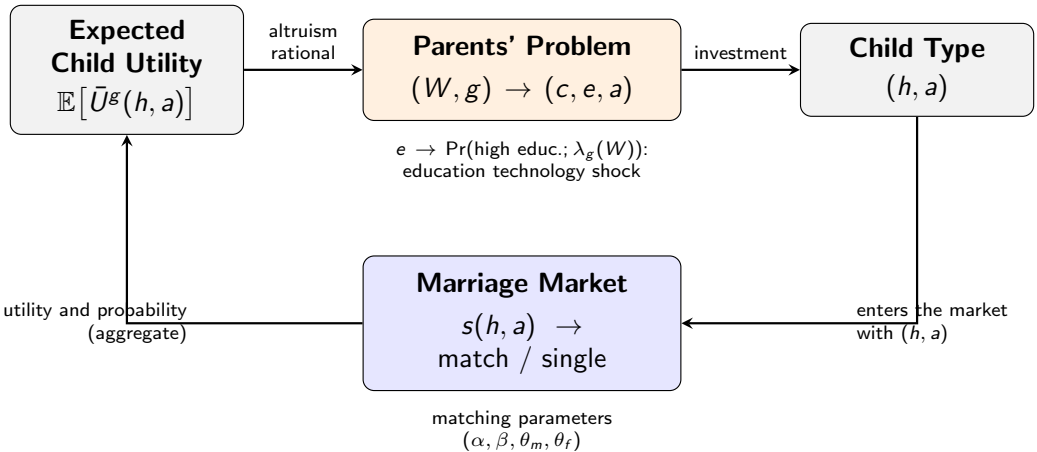
Fact 3: Parents save more for sons' marriage

- Households with sons save more than households with daughters, in both rural and urban areas.
- Parents of sons hold more financial assets before the son's marriage and draw them down at marriage.



event-study specification

Model: Overview



Parents' problem

Parents heterogeneous by wealth W and child gender $g \in \{m, f\}$. Altruistic parents allocate W between own consumption c, education e, and asset transfer a:

$$\begin{aligned} \max_{c,\, e,\, a} \quad & \ln c + \underbrace{\delta\, \mathbb{E}\left[\bar{U}_g(h(e), a)\right]}_{\text{altruism}} \qquad \delta \geq 0 \\ \text{s.t.} \quad & c + e + a = W, \qquad W \sim \text{Lognormal}(\mu_W, \sigma_W^2), \qquad c, e, a \geq 0. \end{aligned} \tag{1}$$

Since h is stochastic, the altruism term is an expectation over the education outcome:

$$\mathbb{E}[\bar{U}_g(h(e), a)] = p_g(e)\, \bar{U}_g(\bar{h}, a) + [1 - p_g(e)]\, \bar{U}_g(1, a) \tag{2}$$

$p_g(e)$ is realized through a **gender-biased** stochastic technology governed by $\lambda_g(W)$.
$h \in \{1, \bar{h}\}$: discrete education outcome; δ: altruism weight;
$\bar{U}_g(h, a)$: child's expected utility in marriage market, specified below.

education stochastic shock

Marriage market: child's individual utility

Parents choose (e, a) before the marriage-market shocks are realized.

$$\bar{U}_g(h, a) = \mathbb{E}_\varepsilon \left[U_g^k(h, a) \right].$$

Entering the marriage market, child k of gender g with (h, a) observes idiosyncratic preference shocks and picks the best available option among all potential partners (h', a'):

$$U_g^k(h, a) = \max \left\{ \underbrace{U_g^k(h, a; \emptyset)}_{\text{single}}, \underbrace{\max_{(h', a')} U_g^k(h, a; h', a')}_{\text{married with } k'} \right\}. \tag{3}$$

Derive closed form of $\bar{U}_g(h, a)$

Marriage market: utility by marital status

Given type (h, a), individual k's utility Single:

$$U_g^k(h, a; \emptyset) = \underbrace{\ln a}_{\text{asset value}} + \underbrace{\kappa_g(h)}_{\text{education value}} + \varepsilon^k, \qquad \varepsilon^k \sim \text{T1EV}(\sigma_\varepsilon). \tag{4}$$

Married:

$$\begin{aligned} U_g^k(h, a; h', a') = & \underbrace{\ln\left(\frac{a + a'}{1 + \chi}\right)}_{\text{asset pooling}} + \underbrace{\mu_g^k(h, h')}_{\text{education matching}} \\ & - \underbrace{\phi_m \mathbf{1}\{g = m\} \max\{0, \ln a' - \ln a\}}_{\text{male relative-asset norm}} - \tau_g + \varepsilon^k. \end{aligned} \tag{5}$$

- κ_g: gender-specific return to education if single, calibrated based on income.
- ε^k: idiosyncratic match-preference shock, $\varepsilon^k \sim \text{T1EV}(\sigma_\varepsilon)$ (Choo and Siow, 2006).
- $\phi_m \geq 0$: penalty on a man's asset disadvantage relative to his spouse.
- $\chi > 0$: economies of scale.
- τ_g: equilibrium price of a match, a technical requirement.

Solving the model: joint surplus

Differencing married and single choice probabilities and using market clearing condition (11).

$$\begin{aligned}\omega(h,a;\,h',a') &= \underbrace{\ln\frac{(a+a')^2}{(1+\chi)^2aa'}-\phi_m[\ln a'-\ln a]_+}_{\text{asset surplus: pooling}-\text{norm}}+\underbrace{\mu^m(h,h')-\kappa_m(h)}_{\text{net edu. surplus, male}}+\underbrace{\mu^f(h',h)-\kappa_f(h')}_{\text{net edu. surplus, female}}\\ &= \underbrace{\ln\frac{(a+a')^2}{(1+\chi)^2aa'}-\phi_m\big[\ln a'-\ln a\big]_+}_{\text{asset surplus: pooling}-\text{norm}}+\underbrace{\alpha hh'-\beta\left(h-h'\right)^2+\theta_m h+\theta_f h'}_{\text{education-matching surplus}}.\end{aligned} \tag{6}$$

Solving the model: joint surplus

Differencing married and single choice probabilities and using market clearing condition (11).

$$\begin{aligned}\omega(h,a;\,h',a') &= \underbrace{\ln\frac{(a+a')^2}{(1+\chi)^2aa'} - \phi_m[\ln a' - \ln a]_+}_{\text{asset surplus: pooling − norm}} + \underbrace{\mu^m(h,h') - \kappa_m(h)}_{\text{net edu. surplus, male}} + \underbrace{\mu^f(h',h) - \kappa_f(h')}_{\text{net edu. surplus, female}} \\ &= \underbrace{\ln\frac{(a+a')^2}{(1+\chi)^2aa'} - \phi_m\big[\ln a' - \ln a\big]_+}_{\text{asset surplus: pooling − norm}} + \underbrace{\alpha hh' - \beta\left(h-h'\right)^2 + \theta_m h + \theta_f h'}_{\text{education-matching surplus}}.\end{aligned} \tag{6}$$

The Choo–Siow (Choo and Siow, 2006) transferable-utility structure delivers a closed-form joint surplus:

$$\omega(h,a;\,h',a') = \sigma_\varepsilon \ln\frac{M(h,a;\,h',a')^2}{M_{m,\emptyset}(h,a)\,M_{f,\emptyset}(h',a')}. \tag{7}$$

full derivation & surplus functions

Solving the model: joint surplus

Differencing married and single choice probabilities and using market clearing condition (11).

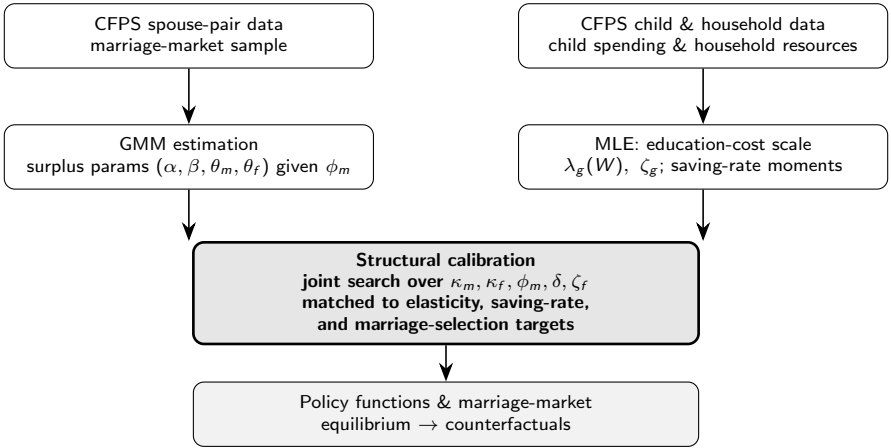
$$\begin{aligned}\omega(h,a;\,h',a') &= \underbrace{\ln\frac{(a+a')^2}{(1+\chi)^2aa'} - \phi_m[\ln a' - \ln a]_+}_{\text{asset surplus: pooling − norm}} + \underbrace{\mu^m(h,h') - \kappa_m(h)}_{\text{net edu. surplus, male}} + \underbrace{\mu^f(h',h) - \kappa_f(h')}_{\text{net edu. surplus, female}} \\ &= \underbrace{\ln\frac{(a+a')^2}{(1+\chi)^2aa'} - \phi_m\left[\ln a' - \ln a\right]_+}_{\text{asset surplus: pooling − norm}} + \underbrace{\alpha hh' - \beta\left(h-h'\right)^2 + \theta_m h + \theta_f h'}_{\text{education-matching surplus}}.\end{aligned} \tag{6}$$

The Choo–Siow (Choo and Siow, 2006) transferable-utility structure delivers a closed-form joint surplus:

$$\omega(h,a;\,h',a') = \sigma_\varepsilon \ln\frac{M(h,a;\,h',a')^2}{M_{m,\emptyset}(h,a)\,M_{f,\emptyset}(h',a')}. \tag{7}$$

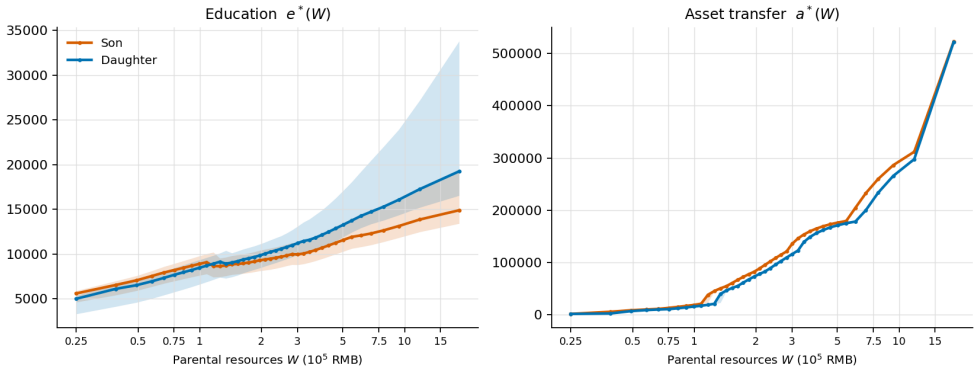
full derivation & surplus functions

From data to model estimation and calibration



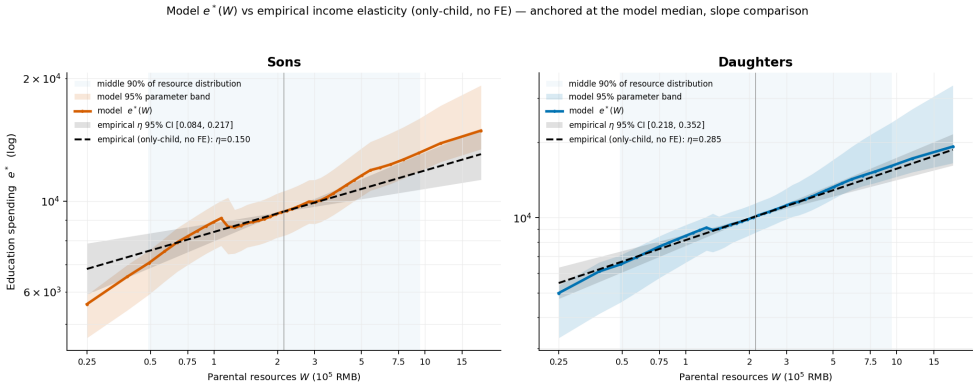
Parents' policy functions

Q1. How does the mix of human capital investment and saving differ by child gender?



95% band: parametric Monte Carlo over 150 draws of (λ_g, ζ_g) (from their estimated sampling s.e.'s) and $(\alpha, \beta, \theta_m, \theta_f)$ (from the GMM covariance); the parents' problem is re-solved at each draw, and the band is the 2.5th--97.5th percentile across draws at each W. (zeta_f=0.1225, delta=0.40, phi_m=0.75, feasible=True; 150/150 draws)

Model fit



Education–income elasticity by gender.

education–income elasticity | saving rate by gender | matching and selection

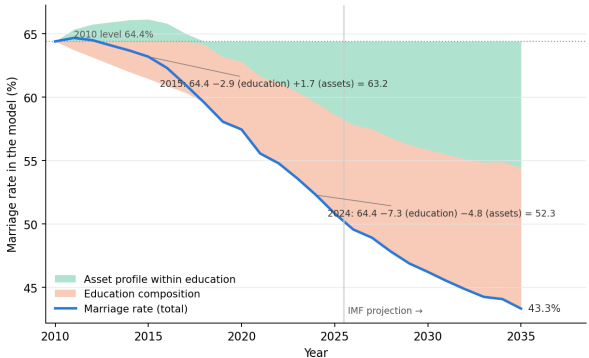
Counterfactual Overview

As parents become richer with China's growth, what happens to marriage? Does gender-biased investment matter for the marriage market, and does where a policy targets matter as much as its size?

- Richer parents: grow parental resources along China's GDP path from 2010. By 2024 the model's marriage rate falls by 18.8%, about 35% of the 53.8% fall in the crude marriage rate, and it keeps falling to 2035.
- Closing the technology difference: give sons daughters' education technology, $\lambda_m(W) = \lambda_f(W)$. The marriage rate rises by $+0.62$pp.
- Same transfer, two targets: a fixed RMB amount (20% of mean parental resources) paid to parents lowers the marriage rate by -8.73pp; paid directly to newlyweds, it raises it by $+14.22$pp.

Policies that route through parents' investment choices erode the marriage-market surplus; policies that reach the market directly do not.

Richer Parents, Fewer Marriages



Relax the limited resources: parental resources grow with real GDP (×2.40 by 2024), education cost fixed. Model −18.8% vs. crude rate −53.8% by 2024 ⇒ **about 35% of the decline**; 33% below 2010 by 2035.

Bands split the total by where it shows up (Shapley): education composition −7.3, asset profile −4.8 of −12.1pp by 2024.

crude marriage rate | table | decomposition method | rising education cost

Counterfactuals: Back to Q2

Q2. Does this gender-biased investment matter for the marriage market?

Same limited resources W; remove the gender difference in education technology: give sons daughters' technology (level and slope in W), $\lambda_m(W) = \lambda_f(W)$.

	Baseline	$\lambda_m = \lambda_f$	Δ (pp)
Men	0.609	0.615	+0.58
Women	0.683	0.689	+0.65
Overall	0.644	0.650	+0.62

Counterfactuals: Back to Q2

Q2. Does this gender-biased investment matter for the marriage market?

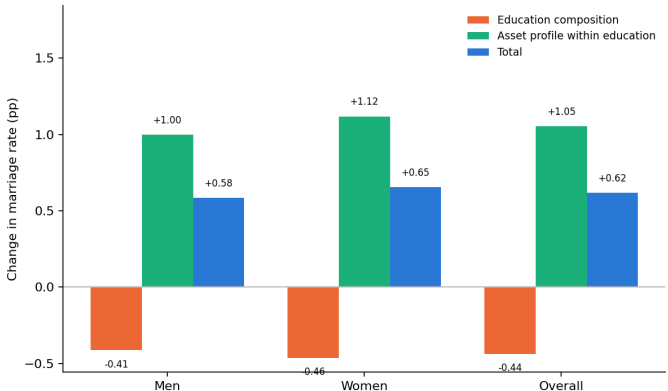
Same limited resources W; remove the gender difference in education technology: give sons daughters' technology (level and slope in W), $\lambda_m(W) = \lambda_f(W)$.

	Baseline	$\lambda_m = \lambda_f$	Δ (pp)
Men	0.609	0.615	+0.58
Women	0.683	0.689	+0.65
Overall	0.644	0.650	+0.62

- **Yes**: with the same resources, equalizing education technology across genders raises the marriage rate by +0.62**pp**.
- The gender difference in education technology lowers the marriage rate.

by education RAS/IPFP construction

Back to Q2: Where Does the Change Show Up?



The total effect is +0.62pp. The Shapley split records where this change shows up in the marriage market: men's high-education share rises from 33.2% to 38.1% (education composition, −0.44pp), and the sons who remain low-educated bring more premarital assets (asset profile, +1.05pp). The two parts add up to the total.

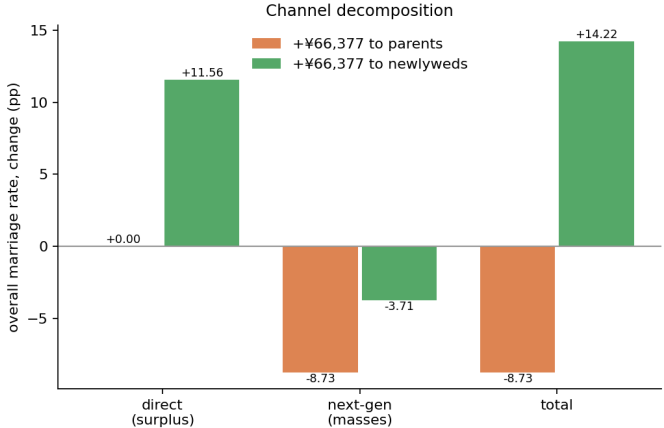
Counterfactuals: Subsidy to Parents / Couples

A fixed RMB amount (20% of mean parental resources), paid either into parents' budget constraint or directly to newlyweds' pooled resources.

	Baseline	+Parents	+Newlyweds	Δ (pp)
Men	0.609	0.527	0.744	−8.26 / +13.46
Women	0.683	0.590	0.833	−9.25 / +15.08
Overall	0.644	0.557	0.786	−8.73 / +14.22

Paid to parents, it funds more education, which this model's negative matching surplus penalizes — the marriage rate *falls*. Paid directly to newlyweds, it raises today's surplus enough to outweigh that same erosion channel — the rate *rises*, by more. *Where* a transfer lands matters as much as its size. sizing the transfer

Counterfactuals: Subsidy — Which Channel Drives It?



Direct: today's population, only the surplus changes (zero for the parent transfer, since W never enters the surplus). *Next-gen*: baseline surplus, only the next cohort's education/asset composition changes. The newlywed transfer's direct surplus effect more than offsets its own next-gen erosion — an extra channel the parent transfer structurally cannot have.

Conclusion

- I study how gender-biased parental investment affects the marriage market.
- I focus on parents' choices between education and savings for sons and daughters.
- I find that parents invest relatively more in education for daughters and more in assets for sons.
- Growing parental resources along China's GDP path accounts for about 35% of the 2010–2024 decline in the crude marriage rate, and the model predicts a further decline as growth continues.
- Using a dynastic marriage-market model, I find that closing the gender gap in education technology raises the marriage rate by 0.62 pp, and that the same transfer can have very different effects depending on whether it goes to parents or directly to new couples.

Appendix: data construction — longitudinal sample

← back

- **Selection**: age 8 for females and above age 10 for males; $\geq$ 6 waves and unmarried at 2010; at least one parent must have completed the household economic questionnaire.
- **Sample**: 2,803 individuals (18,201 person-years; 2,405 families). Gender balanced (49.4% female). 63.7% only children (within-year observations, e.g. an only child in 2010 may have siblings in later years).
- **Marriage variation**: 1,064 individuals transition from unmarried to married (510 F, 554 M), providing within-person identifying variation.

Appendix: data construction — cross-sectional sample

← back

- **Selection**: age 8 for females and above age 10 for males. Drop females older than 45 (98th percentile among single females) and males older than 50 (93rd percentile among single males) to account for fertility considerations. Parental information is not required.
- **Sample**: 145,859 observations. Gender composition: 46.9% female (68,410) and 53.1% male (77,449). 36.5% only children (within-year observations).
- **Urban–rural distribution**: balanced across gender (45.7% urban for males; 47.0% urban for females).

Appendix: data construction — cross-sectional sample with parental info

← back

- **Selection**: keep observations where at least one parent completed the household economic questionnaire.
- **Sample**: 91,528 observations with parental information (60% of the sample without this restriction). Gender composition: 41.5% female (37,968) and 58.5% male (53,560). 50.1% only children (within-year definition).
- **Only-child subsample**: 45,816 observations; 34.4% female (15,756) and 65.6% male (30,060).
- **Urban–rural distribution**: 44.4% urban for males and 47.8% urban for females — broadly balanced across gender.

Appendix: overview of the empirical samples

← back

	Observations	Individuals	Families	Years
Cross-sectional child-year sample	145,859	42,524	17,738	2010–2022
Longitudinal child-year sample	18,201	2,803	3,487	2010–2022
Marriage-market individual-year sample	89,410	30,795	14,340	2010–2022

Notes: The cross-sectional and longitudinal datasets use child-year observations. The marriage-market dataset uses individual-year observations constructed from spouse links and includes the unmarried comparison group.

Appendix: summary statistics — cross-sectional child sample

← back

	Girls	Boys
Panel A: Summary statistics in levels		
Child characteristics		
Child age	27.51 (10.88)	31.14 (11.86)
Child years of schooling	8.47 (4.65)	9.01 (4.11)
Only child	0.44 (0.50)	0.59 (0.49)
Urban residence	0.47 (0.50)	0.46 (0.50)
Married	0.04 (0.20)	0.05 (0.22)
Parental characteristics		
Mother's years of schooling	6.24 (4.53)	4.80 (4.59)
Father's years of schooling	7.84 (3.96)	7.08 (4.18)
Parental household income (RMB)	86,457.98 (197,581.73)	85,516.21 (196,288.26)
Parental savings (RMB)	52,079.74 (157,050.19)	48,135.54 (174,136.51)
Parental financial assets (RMB)	77,916.58 (305,610.05)	70,578.41 (346,553.27)
Child outcome		
Educational expenditure for the child (RMB)	6,372.15 (9,781.49)	6,927.55 (10,107.29)
Observations	68,410	77,449
Unique children	20,511	22,286

Notes: The unit of observation is the child-year. Statistics are reported in levels as means with standard deviations in parentheses. Girls are defined as gender=0 and boys as gender=1. Amounts are in the units recorded in the CFPS data.

Appendix: summary statistics — longitudinal child sample

← back

	Girls	Boys
Panel A: Summary statistics in levels		
Child characteristics		
Child age	20.23 (6.34)	21.32 (7.04)
Child years of schooling	10.67 (4.45)	10.24 (4.33)
Only child	0.61 (0.49)	0.66 (0.47)
Urban residence	0.49 (0.50)	0.47 (0.50)
Married	0.18 (0.39)	0.18 (0.39)
Parental characteristics		
Mother's years of schooling	6.51 (4.43)	6.16 (4.56)
Father's years of schooling	8.10 (3.75)	7.87 (3.95)
Parental household income (RMB)	79,362.97 (171,768.40)	86,484.38 (182,421.28)
Parental savings (RMB)	51,518.26 (130,611.18)	55,378.58 (157,186.07)
Parental financial assets (RMB)	73,078.06 (197,048.33)	80,897.00 (277,048.33)
Child outcome		
Educational expenditure for the child (RMB)	7,579.36 (10,345.28)	7,382.30 (10,459.35)
Observations	8,398	9,803
Unique children	1,314	1,529
Unique families	1,820	1,949

Notes: The unit of observation is the child-year. Statistics are reported in levels as means with standard deviations in parentheses. The table includes repeated observations for the same child. Amounts are in the units recorded in the CFPS data.

Appendix: summary statistics — regression samples

← back ← regression

	(1) Cross, daughters	(2) Cross, boys	(3) Long, daughters	(4) Long, boys
Panel A: Full regression samples				
Educational expenditure for the child (RMB)	6,373.49 (9,469.43)	6,933.50 (9,991.27)	7,773.11 (10,507.75)	7,570.73 (10,655.13)
Child age	14.67 (4.79)	15.84 (4.45)	16.24 (4.04)	16.01 (4.07)
Child years of schooling	7.57 (4.64)	8.45 (4.16)	9.13 (4.39)	8.74 (4.43)
Only child	0.28 (0.45)	0.39 (0.49)	0.62 (0.49)	0.63 (0.48)
Mother's years of schooling	6.84 (4.49)	6.55 (4.55)	6.67 (4.38)	6.63 (4.57)
Parental household income (RMB)	75,779.34 (175,274.65)	75,504.67 (160,828.44)	64,311.03 (162,628.74)	67,646.14 (150,367.83)
Urban residence	0.44 (0.50)	0.44 (0.50)	0.43 (0.50)	0.44 (0.50)
Observations	15,102	13,197	4,032	4,000
Unique children	5,846	5,314	1,088	1,125
Panel B: Only-child regression samples				
Educational expenditure for the child (RMB)	8,957.07 (11,826.45)	8,721.18 (11,797.22)	8,712.88 (11,610.93)	8,564.91 (11,500.24)
Child age	15.17 (5.06)	16.51 (4.85)	16.05 (4.04)	16.16 (4.08)
Child years of schooling	8.29 (4.78)	9.21 (4.18)	9.16 (4.50)	9.08 (4.41)
Mother's years of schooling	9.03 (4.29)	8.17 (4.58)	7.60 (4.30)	7.66 (4.48)
Parental household income (RMB)	90,528.95 (203,067.71)	85,984.29 (195,116.51)	70,381.92 (193,844.57)	75,372.69 (180,670.56)
Urban residence	0.66 (0.47)	0.59 (0.49)	0.49 (0.50)	0.51 (0.50)
Observations	4,208	5,199	2,483	2,521
Unique children	2,083	2,479	692	747

Notes: The unit of observation is the child-year. Each panel uses the complete-case observations for the corresponding educational-expenditure regression sample. Columns (1) and (3) report daughters; columns (2) and (4) report boys. Variables are reported in levels as means with standard deviations in parentheses. The regressions use the natural logarithms of parental household income and educational expenditure.

Appendix: summary statistics — marriage-market sample

← back

	Men	Women
Panel A: Married		
Age	39.01 (7.39)	37.51 (7.70)
Years of schooling	8.99 (4.00)	8.00 (4.61)
High education (junior college or above)	0.14 (0.34)	0.13 (0.33)
Education difference (spouse minus own, years)	-0.99 (3.71)	0.99 (3.71)
Marries up (spouse more educated)	0.20 (0.40)	0.35 (0.48)
Marries down (spouse less educated)	0.38 (0.48)	0.22 (0.41)
Same education level as spouse	0.42 (0.49)	0.42 (0.49)
Observations	35,903	35,903
Panel B: Unmarried		
Age	27.80 (6.33)	23.80 (4.06)
Years of schooling	11.16 (4.51)	12.84 (3.56)
High education (junior college or above)	0.41 (0.49)	0.62 (0.49)
Observations	9,820	7,784

Notes: The unit of observation is the individual-year in `marriage_market.dta`. Entries are means with standard deviations in parentheses. Education is the level recorded in the survey year. Panel A uses married individuals with nonmissing spouse identifiers and spouse education; Panel B uses never-married individuals. High education is junior college (associate degree) or above, as in the model. Marries up (down) indicates that the spouse has a higher (lower) education level than the individual, computed separately for men and women.

Appendix: summary statistics — model marriage-market sample

← back

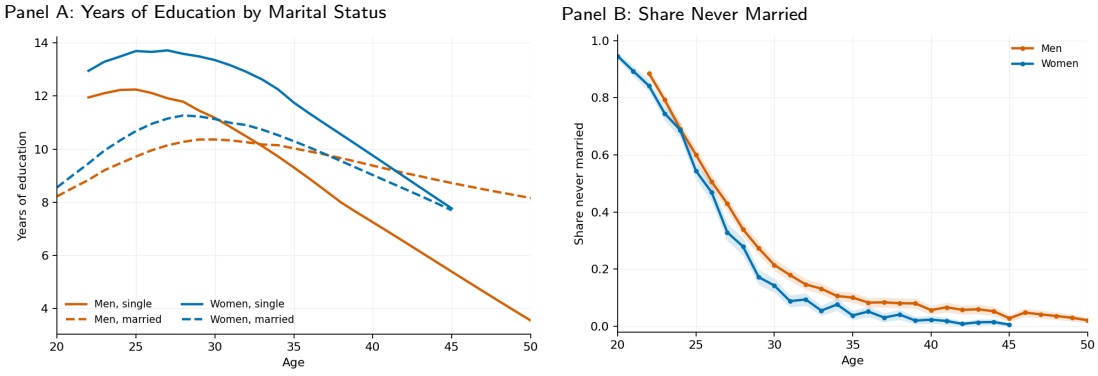
	Men	Women
Panel A: Married		
Age	33.76 (4.90)	32.08 (4.71)
Years of schooling	10.36 (3.86)	10.05 (4.14)
High education (junior college or above)	0.24 (0.43)	0.23 (0.42)
Spouse age	32.06 (4.72)	33.73 (4.90)
Spouse years of schooling	10.03 (4.16)	10.36 (3.86)
Education difference (spouse minus own, years)	-0.32 (3.10)	0.31 (3.08)
Marries up (spouse more educated)	0.23 (0.42)	0.29 (0.45)
Marries down (spouse less educated)	0.32 (0.46)	0.26 (0.44)
Same education level as spouse	0.45 (0.50)	0.45 (0.50)
Observations	16,644	16,569
Panel B: Unmarried		
Age	26.90 (4.54)	23.90 (3.66)
Years of schooling	12.21 (4.01)	13.68 (3.21)
High education (junior college or above)	0.49 (0.50)	0.73 (0.45)
Observations	8,271	6,776

Notes: The sample is the pooled 2012–2022 marriage-market sample used in the model: women aged 20–40 and men aged 22–42, and for married individuals the spouse's age must also lie in the spouse's own window. Entries are weighted means with weighted standard deviations in parentheses, using the model's market weights (CFPS sampling weights normalized by year); observations are unweighted counts. Education is the level recorded in the survey year. High education is junior college (associate degree) or above. Marries up (down) indicates that the spouse has a higher (lower) education level than the individual, computed separately for men and women.

Appendix: Education and marriage by age and gender

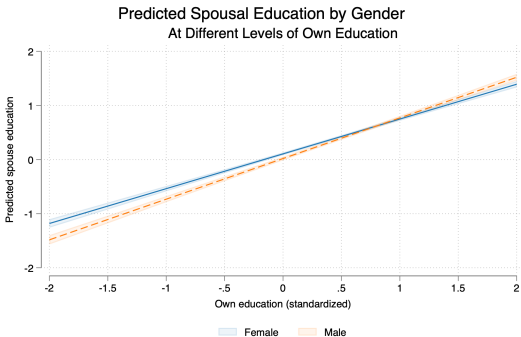
← back

Figure: Education and Marriage by Age and Gender

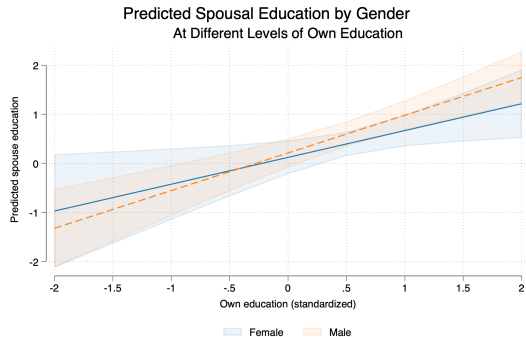


Appendix: weak gender differences in educational matching

← back



Panel A: Regression 2



Panel B: Regression 4

Appendix: income regression — robustness with family FE

← back

	Cross-sectional sample		Longitudinal sample	
	Full sample (1)	Only child (2)	Full sample (3)	Only child (4)
Male	0.306	1.283	0.456	1.073
	(0.304)	(0.834)	(0.501)	(0.889)
Sch	0.112***	−0.008	0.143***	0.119***
	(0.012)	(0.022)	(0.020)	(0.025)
Male × *Sch*	0.035***	0.068***	0.019*	0.009
	(0.007)	(0.014)	(0.011)	(0.014)
ln *y*	0.096***	0.160***	0.081***	0.067*
	(0.021)	(0.058)	(0.030)	(0.038)
Male × ln *y*	−0.059**	−0.169**	−0.049	−0.081
	(0.029)	(0.076)	(0.049)	(0.069)
Observations	26,116	7,177	7,758	4,691
R^2	0.562	0.606	0.555	0.564
Family FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes: The dependent variable is the log of parental educational expenditure on the child, ln *e*. *Male* indicates a son, *Sch* is the child's years of schooling, and ln *y* is log family income. Controls include age, age squared, mother's years of schooling, urban residence, the household equivalence scale $1 + 0.7 \times (\text{adults} - 1) + 0.5 \times$ children, and, in the full samples, an only-child indicator. Columns (1)–(2) use the cross-sectional child sample and columns (3)–(4) the longitudinal child sample; columns (2) and (4) restrict the sample to only children. Standard errors clustered at the family level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix: saving regression — specification

← back

$$\ln(1+s_{ft}) = \beta_0+\beta_1 Male_i+\beta_2 Sch_{it}+\beta_3 (Male_i \times Sch_{it})+\beta_4 \ln y_{ft}+\beta_5 (Male_i \times \ln y_{ft})+\gamma' X_{ift}+\lambda_t+\varepsilon_{ift}$$

- Same specification as the education regression, with log household saving as the outcome; s_{ft} = family income − family expenditure.
- Sample: unmarried children with a bachelor's degree or above, whose parents are approaching the marriage stage; households with negative saving drop out.
- X_{ift}, no family FE, and clustering as in the education regression.
- $\beta_5 > 0$: saving responds more to income for sons than for daughters — the mirror image of the education result.

estimates

Appendix: saving regression — estimates (no family FE)

← back

Male $\times \ln y > 0$ in all four samples, between 0.12 and 0.25, significant in three.

	Cross-sectional sample		Longitudinal sample	
	Full sample (1)	Only child (2)	Full sample (3)	Only child (4)
Male	−1.596	−0.501	−2.580	−4.176*
	(1.222)	(1.561)	(1.822)	(2.167)
Sch	−0.040	−0.053	−0.057	−0.093
	(0.028)	(0.042)	(0.050)	(0.065)
Male × *Sch*	−0.049	−0.051	−0.001	0.079
	(0.055)	(0.074)	(0.071)	(0.088)
$\ln y$	1.338***	1.374***	1.354***	1.320***
	(0.047)	(0.062)	(0.071)	(0.076)
Male × $\ln y$	0.207***	0.118	0.221**	0.245**
	(0.073)	(0.089)	(0.107)	(0.122)
Observations	1,747	1,004	691	497
R^2	0.492	0.515	0.520	0.610
Family FE	No	No	No	No
Year FE	Yes	Yes	Yes	Yes

Notes: The dependent variable is log household saving, $\ln(1+s)$, where saving s is family income minus family expenditure; households with negative saving drop out of the sample. The sample is restricted to unmarried children with a bachelor's degree or above. *Male* indicates a son, *Sch* is the child's years of schooling, and $\ln y$ is log family income. Controls include age, age squared, mother's years of schooling, urban residence, the household equivalence scale $1 + 0.7 \times (\text{adults} - 1) + 0.5 \times \text{children}$, and, in the full samples, an only-child indicator. Columns (1)–(2) use the cross-sectional child sample and columns (3)–(4) the longitudinal child sample; columns (2) and (4) restrict the sample to only children. Standard errors clustered at the family level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix: income × education regression — specification

← back

Educational expenditure varies systematically across children's schooling stages. To capture this heterogeneity, I allow schooling attainment to enter the specification flexibly by including years of schooling ($eduy_{ift}$) and its squared term.
Age is not included because it is highly correlated with schooling attainment in the sample (correlation coefficient close to 0.9). In addition, the dependent variable is defined only for children currently enrolled in school. Given China's nine-year compulsory education system and the near absence of homeschooling, years of schooling provides a more direct measure of children's educational stage than age.
For ease of interpretation, both years of schooling and household income are mean-centered, denoted as c_eduy_{ift} and $c_lnp_fincome1_{ift}$.

$$
\begin{aligned}
\ln fee_{ift} = {} & \beta_0 + \beta_1 gender_{ift} + \beta_2 c_eduy_{ift} + \beta_3 c_lnp_fincome1_{ift} \\
& + \beta_4 \left(gender_{ift} \times c_eduy_{ift}\right) + \beta_5 \left(gender_{ift} \times c_lnp_fincome1_{ift}\right) + \beta_6 \left(c_eduy_{ift} \times c_lnp_fincome1_{ift}\right) \\
& + \beta_7 \left(gender_{ift} \times c_eduy_{ift} \times c_lnp_fincome1_{ift}\right) \\
& + \gamma X_{ift} + \mu_f + \lambda_t + \varepsilon_{ift}
\end{aligned}
$$

X_{ift} includes control variables such as only-child status, parental education, urban residence, and an indicator for eastern provinces. μ_f denotes family fixed effects and λ_t denotes year fixed effects. Standard errors are clustered at the family level. main table

without family FE regression samples

Appendix: income × education regression — main specification (family FE)

← back

	Cross-sectional sample		Longitudinal sample	
	Full sample (1)	Only child (2)	Full sample (3)	Only child (4)
Male	−0.007	0.097	0.072	0.402
	(0.041)	(0.285)	(0.083)	(0.494)
$\widetilde{Sch}$	0.169***	0.054***	0.169***	0.114***
	(0.006)	(0.020)	(0.015)	(0.027)
$\ln \tilde{y}$	0.071***	0.126**	0.050	0.038
	(0.022)	(0.061)	(0.031)	(0.038)
Male × $\widetilde{Sch}$	0.027***	0.058***	0.022**	0.015
	(0.007)	(0.014)	(0.011)	(0.014)
Male × $\ln \tilde{y}$	−0.051*	−0.164**	−0.066	−0.100
	(0.029)	(0.080)	(0.051)	(0.072)
$\widetilde{Sch} \times \ln \tilde{y}$	−0.017***	−0.018***	−0.018***	−0.012*
	(0.004)	(0.007)	(0.006)	(0.006)
Male × $\widetilde{Sch} \times \ln \tilde{y}$	−0.015**	−0.022*	−0.015	−0.020
	(0.006)	(0.011)	(0.010)	(0.013)
Observations	26,116	7,177	7,758	4,691
R^2	0.562	0.604	0.556	0.566
Mean of *Sch*	8.09	9.13	8.89	9.02
Mean of $\ln y$	10.69	10.88	10.60	10.70
Family FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes: The dependent variable is the log of parental educational expenditure on the child, $\ln e$. $\widetilde{Sch}$ and $\ln \tilde{y}$ are years of schooling and log family income, centered at the sample means reported above. Controls include mother's years of schooling, urban residence, the household equivalence scale $1 + 0.7 \times (\text{adults} - 1) + 0.5 \times \text{children}$, and, in the full samples, an only-child indicator. Columns (1)–(2) use the cross-sectional child sample and columns (3)–(4) the longitudinal child sample; columns (2) and (4) restrict the sample to only children. Standard errors clustered at the family level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix: income × education regression — robustness without family FE

← back

	Cross-sectional sample		Longitudinal sample	
	Full sample (1)	Only child (2)	Full sample (3)	Only child (4)
Male	−0.110***	−0.170***	−0.022	−0.062
	(0.026)	(0.041)	(0.041)	(0.052)
$\widetilde{Sch}$	0.197***	0.154***	0.211***	0.194***
	(0.004)	(0.006)	(0.009)	(0.009)
$\ln \tilde{y}$	0.208***	0.245***	0.177***	0.169***
	(0.020)	(0.034)	(0.026)	(0.032)
Male × $\widetilde{Sch}$	0.025***	0.027***	0.021**	0.012
	(0.006)	(0.010)	(0.010)	(0.011)
Male × $\ln \tilde{y}$	−0.046	−0.101**	−0.106***	−0.116**
	(0.028)	(0.045)	(0.041)	(0.050)
$\widetilde{Sch} \times \ln \tilde{y}$	−0.029***	−0.033***	−0.016***	−0.010
	(0.004)	(0.006)	(0.006)	(0.007)
Male × $\widetilde{Sch} \times \ln \tilde{y}$	−0.006	0.003	−0.015	−0.022*
	(0.007)	(0.010)	(0.010)	(0.012)
Observations	28,299	9,407	8,032	5,004
R^2	0.278	0.231	0.361	0.339
Mean of *Sch*	7.98	8.80	8.94	9.12
Mean of $\ln y$	10.71	10.84	10.59	10.68
Family FE	No	No	No	No
Year FE	Yes	Yes	Yes	Yes

Notes: The dependent variable is the log of parental educational expenditure on the child, $\ln e$. $\widetilde{Sch}$ and $\ln \tilde{y}$ are years of schooling and log family income, centered at the sample means reported above. Controls include mother's years of schooling, urban residence, the household equivalence scale $1 + 0.7 \times (\text{adults} - 1) + 0.5 \times \text{children}$, and, in the full samples, an only-child indicator. Columns (1)–(2) use the cross-sectional child sample and columns (3)–(4) the longitudinal child sample; columns (2) and (4) restrict the sample to only children. Standard errors clustered at the family level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix: marriage event study — specification

← back The outcome variable is the parental financial asset ratio:

$$Y_{ft} = \text{FinanceAssetRatio}_{ft} = 100 \times \frac{\text{FinanceAsset}_{ft}}{\text{TotalAsset}_{ft}}.$$

The specification, estimated separately for sons and daughters, is:

$$Y_{ft} = \sum_{k \neq -1} \beta_k D_{it}^k + X_{it}'\gamma + \mu_f + \lambda_t + \varepsilon_{ift}.$$

Event time k counts CFPS survey waves (two years each) relative to the child's first marriage; $k = -1$, the last wave observed unmarried, is the omitted period, and the marriage occurs between $k = -1$ and $k = 0$. The figure plots β_k from the son and daughter regressions.
Control variables include age, age squared, log parental family income, urban residence, and co-residence with parents. All specifications include family and year fixed effects. Standard errors are clustered at the individual level (*pid*). Bars are 95% confidence intervals.

Appendix: choice probabilities

← back Let V denote the deterministic payoff from U. Given the married and single utilities above and the T1EV assumption, the single and matching choice probabilities are

$$P^{g}_{\emptyset}(h,a) = \frac{\exp(V_{\emptyset,g}/\sigma_{\varepsilon})}{\exp(V_{\emptyset,g}/\sigma_{\varepsilon}) + \sum_{(h',a')} \exp(V_{g}(h,a;h',a')/\sigma_{\varepsilon})} \tag{8}$$

$$P^{g}(h,a;h',a') = \frac{\exp(V_{g}(h,a;h',a')/\sigma_{\varepsilon})}{\exp(V_{\emptyset,g}/\sigma_{\varepsilon}) + \sum_{(\tilde{h},\tilde{a})} \exp(V_{g}(h,a;\tilde{h},\tilde{a})/\sigma_{\varepsilon})} \tag{9}$$

Aggregating (8)–(9) via T1EV properties yields the log-sum closed form, which parents maximise in (1).

Appendix: from log-sum to inclusive value

← back Aggregating both options via T1EV gives the closed-form inclusive value used in (2)–(1):

$$\bar{U}_g(h,a) = V_{\emptyset,g}(h,a) - \sigma_\varepsilon \ln P^g_\emptyset(h,a) + \sigma_\varepsilon \gamma \tag{10}$$

Derivation. Let $\mathcal{D}_g(h,a) \equiv 1 + \sum_{(h',a')} \exp\big(\frac{V_g - V_{\emptyset,g}}{\sigma_\varepsilon}\big)$, so $P^g_\emptyset = 1/\mathcal{D}_g$.

$$\begin{aligned}
\bar{U}_g &= \sigma_\varepsilon \log\Big[\exp\tfrac{V_{\emptyset,g}}{\sigma_\varepsilon} + \sum \exp\tfrac{V_g}{\sigma_\varepsilon}\Big] + \sigma_\varepsilon\gamma \\
&= \sigma_\varepsilon \log\Big[\exp\tfrac{V_{\emptyset,g}}{\sigma_\varepsilon} \cdot \mathcal{D}_g\Big] + \sigma_\varepsilon\gamma = V_{\emptyset,g} + \sigma_\varepsilon \log \mathcal{D}_g + \sigma_\varepsilon\gamma \\
&= V_{\emptyset,g} - \sigma_\varepsilon \log P^g_\emptyset + \sigma_\varepsilon\gamma.
\end{aligned}$$

τ_g is absorbed into $P^g_\emptyset$, so $\bar{U}_g$ is transfer-free and ready for (1).

Appendix: Market Clearing & Transfers

← back

Let $M(h, a;\ h', a')$ denote the equilibrium mass of males (h, a) matched with females (h', a'). Bilateral consistency requires:

$$M(h, a;\ h', a') \ \equiv\ n^m(h, a) \cdot P_m(h, a;\ h', a') \ =\ n^f(h', a') \cdot P_f(h', a';\ h, a), \quad \forall\,(h, a, h', a') \tag{11}$$

The equilibrium transfer satisfies a zero-sum condition:

$$\tau_m(h, a;\ h', a') \ +\ \tau_f(h', a';\ h, a) \ =\ 0 \tag{12}$$

Appendix: Equilibrium

An equilibrium is $\left\{e_g^*,\ a_g^*,\ P_g(h,a;\cdot),\ \tau_g(h,a;\cdot),\ \mu(h,a;\cdot)\right\}$ such that:

1. **Rational expectations.** Parents correctly anticipate $\bar{U}_g$ when solving (1).
2. **Parental optimality.** (e_g^*, a_g^*) solves (1) for each g.
3. **Market clearing.** $\{\tau_g\}$ satisfies (11) and (12) for all (h, a, h', a').

Appendix: joint surplus — full derivation & surplus functions

← back **Derivation.** The marriage joint surplus is

$$\begin{aligned}&\ln P_m(h,a;h',a') - \ln P_m(h,a;\emptyset) + \ln P_f(h',a';h,a) - \ln P_f(h',a';\emptyset)\\ &= \frac{1}{\sigma_\varepsilon}\Big[\underbrace{\ln\tfrac{(a+a')^2}{(1+\chi)^2 aa'} - \phi_m[\ln a' - \ln a]_+}_{\text{asset surplus: pooling} - \text{norm}} + \underbrace{\mu^m(h,h') - \kappa_m(h)}_{\Omega_m(h,h')} + \underbrace{\mu^f(h',h) - \kappa_f(h')}_{\Omega_f(h',h)}\Big] \equiv \frac{\omega(h,a;h',a')}{\sigma_\varepsilon}. \end{aligned} \tag{13}$$

Surplus functions. For side $g \in \{m, f\}$ matched with g':

$$\Omega_g(h,h') = \alpha_g hh' - \beta_g(h-h')^2 + \underbrace{\gamma_g h - \kappa_g h}_{\theta_g h}$$

- α_g: education complementarity; only $\alpha_m + \alpha_f$ is identified. Education assortativeness is governed by $\alpha + 2\beta$ ($= 6.46 > 0$ at the estimates).
- $\beta_g > 0$: mismatch penalty; only $\beta_m + \beta_f$ is identified.
- γ_g: own-education gain in marriage (expected < 0, $|\gamma_f| \gg |\gamma_m|$).
- κ_g: single-state return to education (from Single Utility); nets out against γ_g into θ_g.

Identification: Recovering Side-Specific Surpluses

Using the market-clearing condition:

$$\begin{aligned}\omega(h,a;\,h',a') &= \sigma_\varepsilon \ln \frac{M(h,a;\,h',a')^2}{M_{m,\emptyset}(h,a)\,M_{f,\emptyset}(h',a')} \text{(Not feasible)} \\ &= \ln \frac{(a+a')^2}{(1+\chi)^2 a a'} - \phi_m[\ln a' - \ln a]_+ + \underbrace{\Omega_g + \Omega_{g'}}_{\text{Feasible: } \sigma_\varepsilon \ln \frac{M(h;\,h')^2}{M_{m,\emptyset}(h)\,M_{f,\emptyset}(h')}}\end{aligned} \tag{14}$$

Education-matching surplus.

$$\begin{aligned}\sigma_\varepsilon \ln \frac{M(h,h')^2}{M_{m,\emptyset}(h)\,M_{f,\emptyset}(h')} = \omega(h,h') &\equiv \Omega_m(h,h') + \Omega_f(h',h) \\ &= \underbrace{(\alpha_m+\alpha_f)}_{\alpha} hh' - \underbrace{(\beta_m+\beta_f)}_{\beta}(h-h')^2 + \theta_m h + \theta_f h'\end{aligned} \tag{15}$$

The marriage-market dataset identifies the LHS, $(\alpha, \beta, \theta_m, \theta_f)$ could be estimated by GMM, and by the second equality in (14) we recover the child's expected utility $\bar{U}_g(h,a)$ in (10).

Appendix: Mass Split — Setup

Type notation. Let $x = (h, a)$ denote a **male** type and $y = (h', a')$ a **female** type throughout.

Starting point: equilibrium condition from IPFP.

$$M[x,y] \;=\; \underbrace{\sqrt{M_{m,\emptyset}[x]}}_{s_x} \cdot \underbrace{\exp\!\Big(\frac{\omega[x,y]}{2\sigma_\varepsilon}\Big)}_{K[x,y]} \cdot \underbrace{\sqrt{M_{f,\emptyset}[y]}}_{s_y} \tag{A1}$$

Goal. Given the masses $\{M[x,y],\, M_{m,\emptyset}[x],\, M_{f,\emptyset}[y]\}$ from IPFP and the joint surplus ω, recover the **side-specific** surpluses $u_m[x,y]$ and $u_f[x,y]$ separately — without observing transfers τ directly.

Key identity from the logit. The male log-odds of matching with y vs. staying single equals u_m/σ_ε:

$$\ln \frac{M[x,y]}{M_{m,\emptyset}[x]} \;=\; \frac{u_m[x,y]}{\sigma_\varepsilon}$$

Appendix: IPFP — what it does

← back Male type $x = (h, a)$, female $y = (h', a')$; $N_T = N_h \times N_a = 18$ per side. Given n^m, n^f and surplus ω, IPFP finds the unique equilibrium masses $\{M_{m,\emptyset}, M_{f,\emptyset}, M\}$ with market clearing on both sides:

$$\sum_y M[x,y] + M_{m,\emptyset}[x] = n^m[x], \qquad \sum_x M[x,y] + M_{f,\emptyset}[y] = n^f[y].$$

With kernel $K[x,y] = \exp(\omega[x,y]/2\sigma_\varepsilon)$ and $s_x = \sqrt{M_{m,\emptyset}[x]}$, $s_y = \sqrt{M_{f,\emptyset}[y]}$: $M[x,y] = s_x\, K[x,y]\, s_y$. Unique equilibrium under Choo–Siow (2006).

Appendix: Mass Split — Derivation

Direction: masses $\longrightarrow u_m$
Divide (A1) by $M_{m,\emptyset}[x] = s_x^2$ and take logs:

$$\ln \frac{M[x,y]}{M_{m,\emptyset}[x]} \;=\; \frac{\omega}{2\sigma_\varepsilon} \;+\; \frac{1}{2} \ln \frac{M_{f,\emptyset}[y]}{M_{m,\emptyset}[x]}$$

Multiply through by σ_ε and use the logit identity above:

$$\boxed{u_m(h,a;\,h',a') \;=\; \frac{\omega}{2} \;+\; \frac{\sigma_\varepsilon}{2} \ln \frac{M_{f,\emptyset}(h',a')}{M_{m,\emptyset}(h,a)}, \qquad u_f \;=\; \omega - u_m}$$

Identification interpretation. The gender asymmetry in single rates $M_{f,\emptyset}[y]/M_{m,\emptyset}[x]$ is the identifying variation that splits ω into u_m and u_f separately, recovering μ^m and μ^f individually (Fact 2 identification).

Appendix: GMM estimation of marriage-surplus parameters

← back The marriage-market data provide eight moments: four single rates by gender and education, and four education-match shares. The estimator chooses $\vartheta = (\alpha, \beta, \theta_m, \theta_f)$ to minimize

$$\widehat{\vartheta} = \arg\min_{\vartheta} \left\| \widehat{m}(\vartheta) - m^{data} \right\|^2,$$

where single rates are transformed by the logit and match shares by a centered log-ratio transformation. The model moments are generated by solving the marriage market; the four parameters are bounded and optimized from five starting values. Standard errors use the sandwich formula with a Kish-effective-size, delta-method covariance matrix for the empirical moments.

Appendix: constructing the resource distribution W

← back

Step 1. Equivalized household variables from the 2018 household file: `equiv_income` $=$ `fincome`$/$`ES`, `equiv_net_asset` $=$ `fnet_asset`$/$`ES`, with `ES` $= 1 + 0.7(\text{adults} - 1) + 0.5\,\text{children}$.

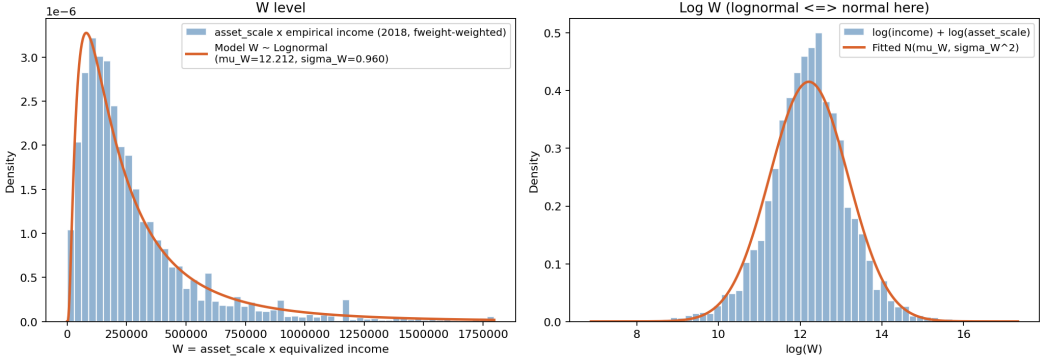
Step 2. Weighted lognormal fit: $\log y_i \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\widehat{\mu}_Y = \frac{\sum w_i \log y_i}{\sum w_i}$, $\widehat{\sigma}_Y^2 = \frac{\sum w_i (\log y_i - \widehat{\mu}_Y)^2}{\sum w_i}$.

Step 3. Scale via weighted medians: `asset_scale` $=$ wmed(`equiv_net_asset`)$/$ wmed(`equiv_income`) $= 5.824$. Thus W is a scaled, equivalized household-income measure interpreted as parents' lifetime resources.

Appendix: household resources distribution

← back

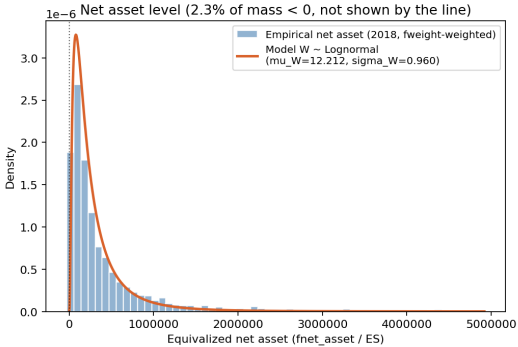
Final W distribution: W_grid = asset_scale x lognormal(2018) income (asset_scale=5.461)

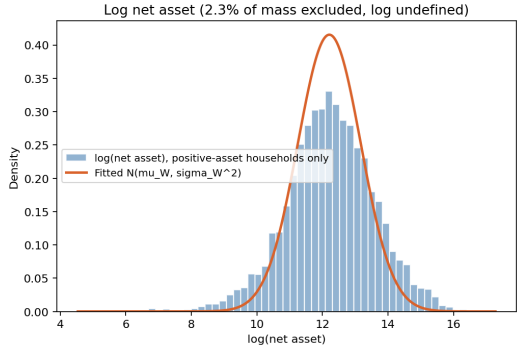


Appendix: household resources distribution (raw)

← back

Model W (lognormal) vs. empirical net asset (2018, asset_scale=5.461)





Appendix: constructing the asset grid a

← back Built from observed positive premarital assets (not lognormal). Let `raw_asset` be the observed/imputed asset, `market_weight` its weight, $W_{\max} = \max(W\text{-grid})$.

1. $a_{\min}$ = weighted 1% quantile of positive `raw_asset`.
2. $n_a - n_{\text{tail}}$ empirical nodes at weighted quantiles $q_j = \frac{j-1/2}{n_{\text{emp}}-1}$ ($n_a = 9$); prepend $a_{\min}$.
3. n_{tail} log-spaced nodes from the empirical max to $W_{\max}$.

PCHIP interpolation over the 9 nodes; parent's saving choice by grid search over 2,000 points.

Appendix: education production — exponential shock

← back Education is binary, $h \in \{1, \bar{h}\}$. A random threshold T_g (exam luck, school quality) determines the outcome:

$$h = \begin{cases} \bar{h} & \text{if } e \geq T_g \\ 1 & \text{otherwise} \end{cases} \qquad T_g \sim \text{Exponential}(\lambda_g) \tag{16}$$

Closed-form success probability:

$$p_g(e) = 1 - \exp\left(-\frac{e}{\lambda_g(W)}\right) \tag{17}$$

- $\lambda_g(W)$: gender-specific cost scale, varying with parents' resources; higher $\lambda_g \Rightarrow$ education more costly.
- Estimates imply the education-cost scale **differs by gender**; $\bar{h} = 2$ from the associate-degree income ratio.

Appendix: estimating $\lambda_g(W)$

← back $\lambda_g(W) = \lambda_{g,0}\,(W/W_{\text{ref}})^{\zeta_g}$, W_{ref} = median of the model W-grid. Technology $p_i = 1 - \exp[-e_i/\lambda_g(W_i)]$. $(\lambda_{g,0}, \zeta_g)$ by weighted MLE with $\zeta_g \geq 0$:

$$(\widehat{\lambda}_{g,0}, \widehat{\zeta}_g) = \arg \max_{\lambda_{g,0}>0,\ \zeta_g \geq 0} \sum_i \widetilde{w}_i \left[y_i \log p_i + (1 - y_i) \log(1 - p_i) \right].$$

e_i: mean annual education spending; $y_i = 1\{\texttt{final_education}_i \geq 5\}$. Estimates: $\widehat{\lambda}_{m,0} = 2566$, $\widehat{\zeta}_m = 0$; $\widehat{\lambda}_{f,0} = 2286$, $\widehat{\zeta}_f = 0.238$ (CI $[0.065, 0.412]$); the parent problem uses $\zeta_f = 0.122$. units of λ

Appendix: units of the education cost λ

← back $\lambda_{g,0}$ is estimated on *annual* spending; e in the parents' problem is a one-time outlay out of W. ζ_g is unit-free (W/W_{ref}), but the level needs a conversion s_e: the model uses $s_e\lambda_g(W)$. Calibrate s_e jointly with $(\kappa_m, \kappa_f, \phi_m, \delta, \zeta_f)$, adding the high-education share as a target:

s_e	η_m	η_f	$\eta_f - \eta_m$	p_m	p_f	SR gap
1.0	**0.202**	**0.299**	**0.097**	0.972	0.986	0.046
1.5	0.229	0.362	0.132	0.967	0.980	0.044
2.0	0.256	0.340	0.084	0.953	0.972	0.044
2.5	0.269	0.350	0.081	0.954	0.970	0.040
3.0	0.290	0.365	0.075	0.943	0.963	0.043
4.0	0.338	0.394	0.057	0.905	0.949	0.051
Data	$[-.009, .164]$	$[.096, .285]$	$[.059, .169]$	0.718	0.837	0.050

The elasticities select $s_e = 1$: a larger s_e raises both elasticities and shrinks the gap, while κ adjusts and keeps $p > 0.9$. With the exponential technology, η and p both depend on e/λ; I prioritize the elasticities.

Appendix: IPFP — algorithm

← back **Init:** $s_x^{(0)} = \sqrt{n^m[x]}/2$, $s_y^{(0)} = \sqrt{n^f[y]}/2$.

Iterate until $\max(\|\Delta s_x\|, \|\Delta s_y\|) < \epsilon$:

$$S_x = K s_y, \qquad s_x^{\text{new}} = \frac{-S_x + \sqrt{S_x^2 + 4n^m}}{2}$$

$$S_y = K^\top s_x^{\text{new}}, \qquad s_y^{\text{new}} = \frac{-S_y + \sqrt{S_y^2 + 4n^f}}{2}$$

Recover: $M_{m,\emptyset} = s_x^2$, $M_{f,\emptyset} = s_y^2$, $M[x,y] = s_x K[x,y] s_y$.

Appendix: parameters — estimated

← back

Symbol	Description	Model	95% CI
$\widehat{\mu}_W,\ \widehat{\sigma}_W$	Lognormal W	10.514, 0.960	—
$\widehat{\alpha}$	Education complementarity	-1.131	$[-1.230, -1.033]$
$\widehat{\beta}$	Mismatch penalty	3.795	$[3.663, 3.926]$
$\widehat{\theta}_m$	Married education (M)	-0.039	$[-0.179, 0.101]$
$\widehat{\theta}_f$	Married education (F)	-0.919	$[-1.071, -0.768]$
$\widehat{\lambda}_m$	Education scale (M)	2566	$[2156, 3055]$
$\widehat{\lambda}_f$	Education scale (F)	2286	$[1905, 2742]$
$\widehat{\zeta}_m$	Wealth elasticity (M)	0	$[0, 0.143]$
$\widehat{\zeta}_f$	Wealth elasticity (F)	0.122	$[0.065, 0.412]$

Marriage-surplus CIs: sandwich SEs from GMM re-estimation at $\phi_m = 0.75$; λ_g, ζ_g CIs from weighted-MLE (ζ_m on the boundary).

Appendix: parameters — calibrated / from literature

← back

Symbol	Description	Model
$\widetilde{h}_L,\ \widetilde{h}_H$	Education types (norm.; income ratio)	1, 2
$\widetilde{\kappa}_m,\ \widetilde{\kappa}_f$	Single education return (M, F)	3.45, 5.41
$\widetilde{\phi}_m$	Relative-asset norm, male only	0.75
$\widetilde{\delta}$	Altruism weight	0.40
$\overline{\chi}$	Economies of scale in marriage	0.7
$\overline{\mathrm{SR}}$	Marriage-market sex ratio (imposed)	1.12
$\widetilde{s}_e$	Education-cost unit conversion	1
σ_ε	T1EV scale normalization	1

$\overline{\mathrm{SR}} = 1.12$ is based on Census and imposed on male marriage-market masses ; male education $\times$ asset composition held fixed.

Appendix: fit — education–income elasticity

← back

	Data range	Model
Male elasticity	$[-0.009, 0.164]$	0.202
Female elasticity	$[0.096, 0.285]$	0.299
Female − male	$[0.059, 0.169]$	0.097

Calibration anchor: only-child, no-family-FE $(\eta_m, \eta_f, \eta_f - \eta_m) = (0.150, 0.285, 0.135)$; full-sample benchmark $(0.164, 0.246, 0.082)$. The calibration lands at $\zeta_f = 0.122$; the gap stays inside $[0.059, 0.169]$.

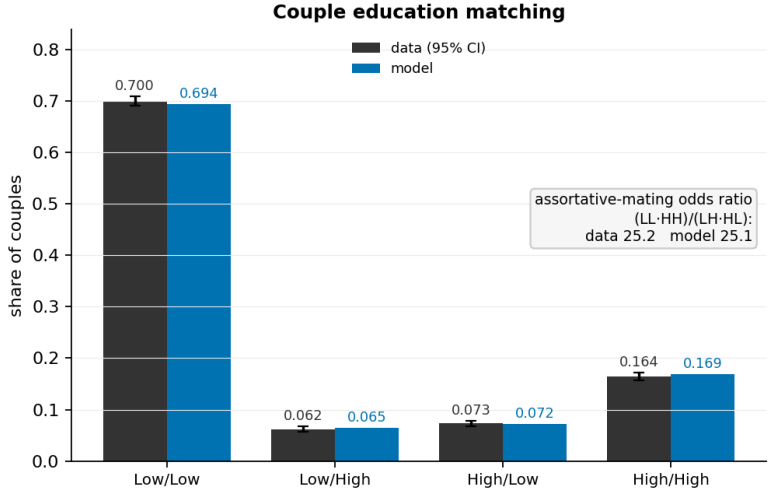
Appendix: fit — saving rate by gender

← back

Gender	Series	mean	$p10$	$p25$	$p50$	$p75$	$p90$
Male	Data	0.348	0.050	0.129	0.325	0.564	0.720
Male	Model	0.330	0.165	0.292	0.427	0.412	0.337
Female	Data	0.298	0.046	0.106	0.270	0.502	0.668
Female	Model	0.284	0.147	0.161	0.373	0.392	0.291

Raw a/W definition. Saving-rate gap $s_m - s_f$: data 0.050 / model 0.046 (mean).

Appendix: fit — matching and selection



← back

Marriage-matching matrix: data versus model.

Appendix: fit — matching and selection (tables)

Matching by education

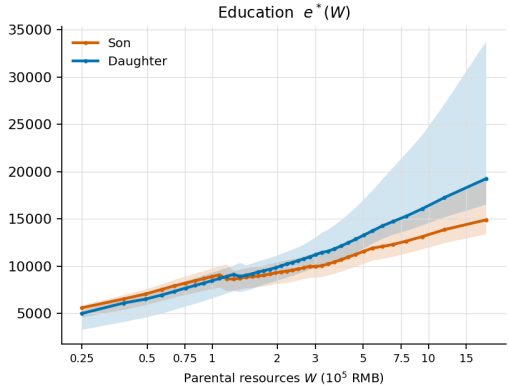
Mh / Fh	Data	Model
Low / Low	0.700	0.694
Low / High	0.062	0.065
High / Low	0.073	0.072
High / High	0.164	0.169

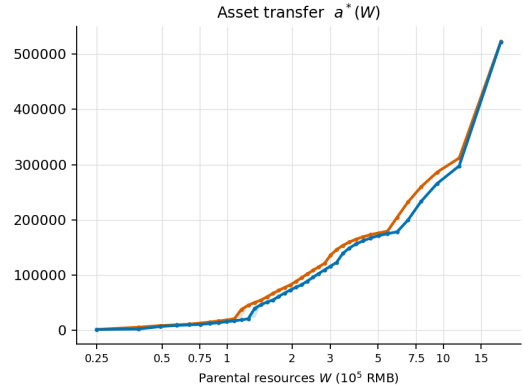
High education, single vs. married

	Data	Model
M single	0.495	0.474
M married	0.237	0.242
F single	0.727	0.737
F married	0.234	0.234

Appendix: policy functions $e_g^*(W)$, $a_g^*(W)$ (with 95% CI)

← back





95% band: parametric Monte Carlo over 150 draws of (λ_g, ζ_g) (from their estimated sampling s.e.'s) and $(\alpha, \beta, \theta_m, \theta_f)$ (from the GMM covariance); the parents' problem is re-solved at each draw, and the band is the 2.5th–97.5th percentile across draws at each W. (zeta_f=0.1225, delta=0.40, phi_m=0.75, feasible=True; 150/150 draws)

Appendix: Ratio Adjustment (RAS / IPFP Idea)

← back Baseline policy $\rightarrow$ masses b^g; counterfactual policy $\rightarrow$ masses c^g. Shift the *empirical* n^g_{emp}:

$$\tilde{n}^g_{ij} = n^g_{\text{emp},ij} \underbrace{\frac{c^g_{ij} / \sum_k c^g_{ik}}{b^g_{ij} / \sum_k b^g_{ik}}}_{\text{shape_ratio}_{ij}\ \text{(asset profile)}} \underbrace{\frac{\sum_k c^g_{ik}}{\sum_k b^g_{ik}}}_{\text{row_ratio}_i\ \text{(education comp.)}} \quad \text{then renormalize to } \sum n^g_{\text{emp}}.$$

- Same algebra as RAS / iterative proportional fitting: seed $\times$ row factor $\times$ column factor.
- Factors are *given* by the model's base$\rightarrow$cf change, not solved to hit target margins.
- No change in policy $\Rightarrow$ ratios $= 1 \Rightarrow \tilde{n}^g = n^g_{\text{emp}}$ exactly.

Appendix: $\lambda_m = \lambda_f$ — by gender and education

← back

Group	Baseline	$\lambda_m = \lambda_f$	Δ (pp)
Men, low edu	0.692	0.735	+4.34
Men, high edu	0.443	0.420	−2.32
Women, low edu	0.862	0.860	−0.21
Women, high edu	0.406	0.426	+1.98

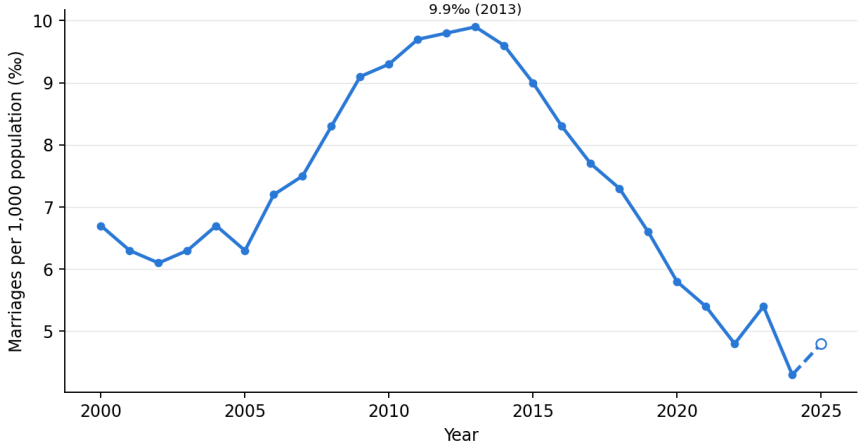
Low-education men marry more: the sons who remain low-educated bring more premarital assets, which the provider norm rewards. High-education men marry less. Among women, the high-education marry more and the low-education slightly less. Women's education composition is unchanged.

Appendix: sizing the flat transfer

← back $T = \theta \times \overline{W}$, $\theta = 20\%$, $\overline{W}$ = mean parental resources at the calibrated point (≈¥331,884), giving T ≈¥66,376.73 — an equal-*transfer-size* comparison (same RMB per recipient), not an equal-total-budget one. RMB reference: median equivalized household income, CFPS 2018 (≈¥38,125/yr), so $T \approx 20.9$ months of income.

- **Parent side.** $W \to W + T$ for every household — additive, not proportional (the same RMB for everyone, matching the flat newlywed-side transfer).
- **Newlywed side.** `marriage_subsidy_S`$= T$ in the surplus resource term $2\ln(a + a' + T) - \ln a - \ln a' - 2\ln(1 + \chi)$, credited only to couples who match; ϕ_m still uses the *premarital* a, a'.

Appendix: crude marriage rate, 2000–2025



← back

Marriage registrations per 1,000 population, Ministry of Civil Affairs. 2025 (open circle) is preliminary. Different unit from the model (share married), so only % changes are compared: 2010–2024 model −18.8%, crude −53.8% ⇒ 18.8/53.8 ≈ 35%.

Appendix: growth counterfactual — GDP path table

← back

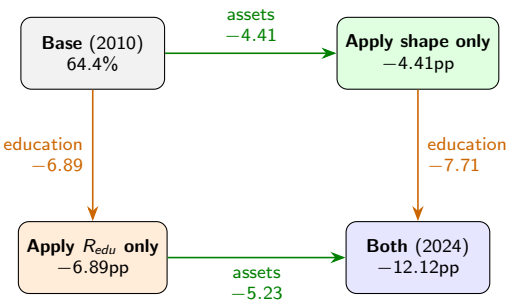
Year	Parental resources G_t	Marriage rate (model)	Change since 2010			Share explained
			Model (pp)	Model (%)	Crude marriage rate (%)	
2010	1.000	0.644	0.00	0.0	0.0	—
2015	1.464	0.632	−1.20	−1.9	−3.2	0.58
2020	1.937	0.574	−6.94	−10.8	−37.6	0.29
2022	2.169	0.548	−9.62	−14.9	−48.4	0.31
2024	2.400	0.523	−12.12	−18.8	−53.8	**0.35**
2025	2.520	0.508	−13.57	−21.1	−48.4†	0.44†
2030	3.049	0.462	−18.17	−28.2	—	—
2035	3.586	0.433	−21.06	−32.7	—	—

Crude marriage rate in 2010: 9.3‰. G_t: parental resources relative to 2010 (IMF WEO April 2026; projected after 2025); education cost fixed. Model (pp) is the change shown in the main figure; Model (%) is the same change relative to 2010, compared with the crude rate. Share explained = Model (%) / Crude (%); 2024 is the latest official year. †2025 crude rate is preliminary.

how the change is decomposed

Appendix: decomposition method (GDP, 2010 → 2024)

← back



Two orders, two answers ⇒ average them.

Step 1. Split the model's change in each cell

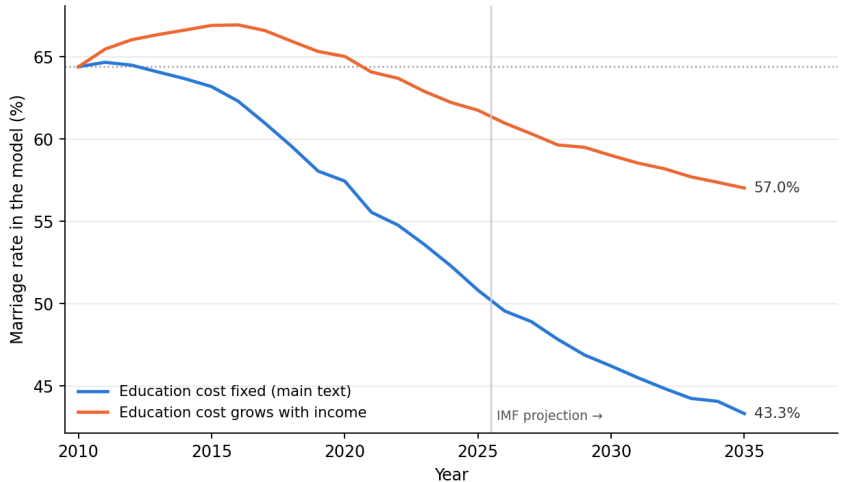
$$\frac{c(h,a)}{b(h,a)} = \underbrace{R_{edu}(h)}_{\text{education composition}} \times \underbrace{shape(a \mid h)}_{\text{asset profile within education}}$$

Step 2. Apply to the observed market
Multiply the data's type masses by one or both factors (sex ratio fixed) and re-solve the market: four corners.

Step 3. Shapley: average over the two orders
Education: $\frac{1}{2}(-6.89 - 7.71) = \mathbf{-7.30}$
Assets: $\frac{1}{2}(-4.41 - 5.23) = \mathbf{-4.82}$
Sum: -12.12pp = total change.

An accounting of where the change shows up, not separate mechanisms; other splits are possible.

Appendix: growth with rising education cost



← back

If the education cost grows one-for-one with income ($\lambda_g \to G_t \lambda_g$), the marriage rate first rises (+2.5 pp in 2016) and falls only 2.2 pp by 2024. The education-cost channel is −9.9 of the −12.1 pp decline by 2024 (−13.7 pp by 2035).

References I

Banerjee, A., Meng, X., Porzio, T., and Qian, N. (2023). Aggregate fertility and household savings in China. Technical report, Working paper.

Banerjee, A., Meng, X., and Qian, N. (2010). The life cycle model and household savings: Micro evidence from urban China. Technical report, Working paper.

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

Boar, C. (2021). Dynastic precautionary savings. *The Review of Economic Studies*, 88(6):2735–2765.

Braxton, J. C., Chikhale, N., Herkenhoff, K. F., and Phillips, G. M. (2024). Intergenerational mobility and credit. NBER Working Paper 32031, National Bureau of Economic Research.

Chambers, C., Goldman, B., and Winkelmann, J. (2026). Bachelors without bachelor's: Gender gaps in education and declining marriage rates. NBER Working Paper 35179, National Bureau of Economic Research.

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

Choo, E. (2015). Dynamic marriage matching: An empirical framework. *Econometrica*, 83(4):1373–1423.

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

Du, Q. and Wei, S.-J. (2013). A theory of the competitive saving motive. *Journal of International Economics*, 91(2):275–289.

Fang, M., Huang, Z., Wang, Y., and Yang, Y. (2026). Sex and the city: Demographics, spatial sorting, and the marriage market. *International Economic Review*.

Gao, Y., Alessie, R., and Angelini, V. (2023). Parental housing wealth and children's marriage prospects in china—evidence from charls. *Review of Economics of the Household*, 21(2):615–644.

Geng, F. and Qin, X. (2019). Women's dominance in education: An analysis of the influence of housing price increase on children's educational expenditure. *Journal of Finance and Economics*, 45(4):54–67.

Goldin, C., Katz, L. F., and Kuziemko, I. (2006). The homecoming of American college women: The reversal of the college gender gap. *Journal of Economic Perspectives*, 20(4):133–156.

Greenwood, J., Guner, N., Kocharkov, G., and 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.

McGarry, K. (2016). Dynamic aspects of family transfers. *Journal of Public Economics*, 137:1–13.

Nie, G. (2020). Marriage squeeze, marriage age and the household savings rate in China. *Journal of Development Economics*, 147:102558.

Nie, H. and Xing, C. (2019). Education expansion, assortative marriage, and income inequality in china. *China Economic Review*, 55:37–51.

Qian, N. (2008). Missing women and the price of tea in china: The effect of sex-specific earnings on sex imbalance. *The Quarterly Journal of Economics*, 123(3):1251–1285.

References II

Rüttenauer, T. and Ludwig, V. (2023). Fixed effects individual slopes: Accounting and testing for heterogeneous effects in panel data or other multilevel models. *Sociological Methods & Research*, 52(1):43–84.

Shimer, R. and Smith, L. (2000). Assortative matching and search. *Econometrica*, 68(2):343–369.

Sun, A. and Zhang, Q. (2020). Who marries whom in a surging housing market? *Journal of Development Economics*, 146:102492.

Tian, F. F. and Davis, D. S. (2019). Reinstating the family: Intergenerational influence on assortative mating in China. *Chinese Sociological Review*, 51(4):337–364.

Wei, S.-J. and 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.

Xu, J. (2023). Intergenerational transfers in China: What are the patterns of the transfers and when do the transfers occur? *International Studies of Economics*.

Yang, L., Guo, Z., Liang, H., Samuel, P., Dong, L., Peng, Y., and Li, S. (2023). Male marriage squeeze and a sociodemographic portrait of involuntary bachelors in rural China. *China Population and Development Studies*, 7:342–365.

Zhou, S. and Wang, S. (2024). Gender imbalance in the marriage market and housing demand: Evidence from China. *Sustainability*, 16(14):5861.