More Than a Degree: China's College Expansion and the Marriage Market*

Ming Fang†[1] and Sai Luo‡[2]

[1]University of Maryland, College Park

[2]Shanghai University of Finance and Economics

December 6, 2022

Click here for the latest version.

Abstract

China has experienced a radical expansion of college education that began in 1999. We study the impacts of this college expansion on the Chinese marriage market, with a special focus on its effects on the marriage outcomes of college-educated women and men. The empirical analysis is undergirded by a model featuring educational investment, marriage matching, and reductions in search frictions associated with the expansion. We estimate the effects of the expansion on marriage outcomes by exploiting geographic and birth-cohort variation in exposure to the expansion. Our analysis shows that, consistently with the predictions of the model, the expansion increased the marriage probability of college graduates. The expansion also increased the probability of college-college matches relative to the counterfactual of random matching and reduced the marriage age gap. Our findings highlight the important role of higher education institutions in shaping the marriage market.

JEL Codes: J12, I23, I26

Keywords: marriage market; higher education; college expansion; assortative mating; China

*We thank Katharine Abraham, Judith Hellerstein and Ethan Kaplan for their inspiring and invaluable advice at all stages of this paper. We also thank Caterina Calsamiglia, Mrinmoyee Chatterjee, Shanglyu Deng, Michael Dinerstein, Gabriele Gratton, Run Ge, Melissa Kearney, Weizheng Lai, Camille Landais, Mario Macis, Tianshi Mu, Xincheng Qiu, Ana Reynoso, Anusuya Sivaram, John Soriano, Chenyu Yang, Seth Zimmerman and seminar participants at the University of Maryland for helpful discussions and suggestions. All errors are our own. Ming Fang gratefully acknowledges financial support from the Chiang Ching-kuo Foundation for International Scholarly Exchange.

†mfang@umd.edu

‡luosaiecon@gmail.com

1 Introduction

This paper studies the impact on the marriage market of a radical reform of college education in China. In 1999, the Ministry of Education (MOE) sharply expanded access to higher education. Before the expansion, the rate of college attainment was minimal. The expansion led to a large and continuous increase in college attainment for young men and women. New enrollment in colleges increased by more than 400% in the first eight years after the onset of the expansion (Figure 1).[1] In addition to impacting the labor market and firms as shown in previous studies (Che and Zhang, 2018; Feng et al., 2018; Li et al., 2017), the reform could also have affected the structure of the marriage market in crucial ways. For example, massively expanding access to higher education altered the distribution of college- and non-college-educated women and men and may have reduced search frictions in the college marriage market.

Education not only provides a path to labor-market success but affects marriage outcomes. Returns to education in the marriage market are important influences on educational choices and lifetime welfare (Chiappori et al., 2009, 2018; Lafortune, 2013). Education policies, therefore, may substantially impact the marriage market. Given the nature of the marriage market as a two-sided matching market, such impacts hinge on not only individuals' own education but others' education. When evaluating the impacts of education policies on the marriage market, the general equilibrium effects must be taken into account. This is particularly important in a society whose overall education level has significantly increased, as in many countries in the past few decades (Schofer and Meyer, 2005; World Bank, 2017).

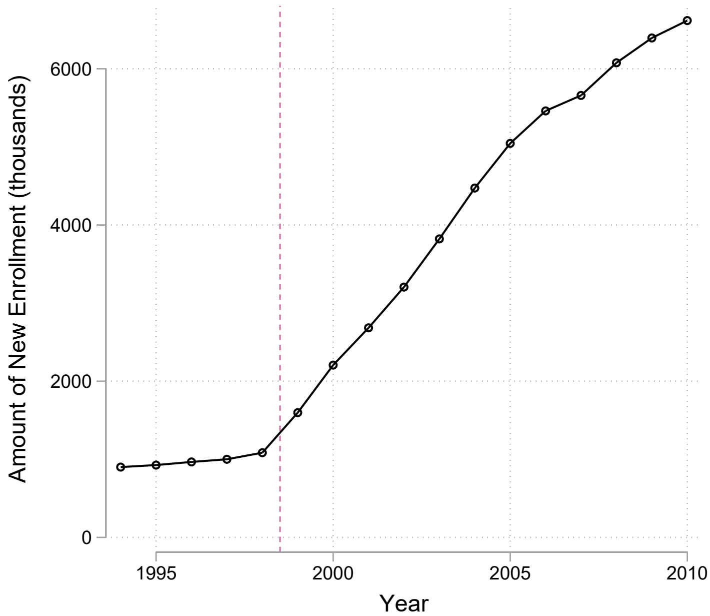
However, it is often difficult to evaluate the causal effects, either from a partial or a general equilibrium perspective, of education on marriage market outcomes because educational investment responds to returns to education in the marriage market. Previous studies have pointed out that marital returns are an important determinant of educational choices (Chiappori et al., 2009; Ge, 2011; Lafortune, 2013; Bruze, 2015; Attanasio and Kaufmann, 2017; Zhang, 2021). China's college expansion provides us with a great opportunity for studying the equilibrium impacts of education policies on the marriage market. First, by exploiting the exogenous timing and the geographic variation of the expansion, we can estimate the causal effects of expanding access to college education on the marriage market. Second, the unprecedented magnitude of the expansion enables us to more easily test its equilibrium effects on the marriage market.

To understand the potential effects of the college expansion on the marriage market and to discipline our empirical analysis, we first build a model of educational investment and marriage matching. In our two-period model, young women and men make choices re-

[1] The dropout rate is extremely low at Chinese universities (Appendix A.2). Therefore, the vast majority of the newly enrolled students became college graduates.

garding college education in the first period, and in the second period they match in the marriage market based on educational-attainment type and idiosyncratic preferences (Choo and Siow, 2006). Anticipation of market prospects in the second period affects educational choices in the first period (Chiappori et al., 2009, 2017).

Figure 1: National Trend of New Enrollment



Notes: This figure reports national new enrollment in colleges for each year based on data from the National Bureau of Statistics of China. The college expansion started in 1999. Both four-year universities and three-year colleges are included.

Our model is further enriched by incorporating the role of colleges in reducing search frictions. Not only might having a college education be an attractive trait in the marriage market, but colleges themselves may serve as local marriage markets. A particular educational institution could directly affect who marries whom by providing a space for social interactions. This channel may operate via various social networks formed during college. Young women and men may meet in class, in student organizations, or via shared friends they met in college. These occasions for social interactions could significantly boost students' chances of getting to know each other and forming romantic relationships.[2] We refer to these potential opportunities as the "local college marriage market" (LCMM).[3]

[2]For example, in Chinese universities, students are typically assigned to a series of classes based on their majors, and majors are determined upon admission to most universities. This arrangement creates chances for the formation of classmate relationships lasting years. Various student associations, meanwhile, are usually formed voluntarily based on common hobbies or aspirations, and they might also provide romantic opportunities based on sorting into the same one.

[3]Recent literature has documented the important role of colleges per se as local marriage markets (Kirkebøen et al., 2021). Indeed, we provide supportive evidence in our context using household survey data (Section 2.2.3).

We model search frictions in the marriage market in a parsimonious way: everyone who enters the market has to pay a fixed cost, which represents search costs. The only exception is that some college students can randomly meet and potentially match. They can thus avoid the search costs yet still get married. The college expansion creates a thicker local marriage market. Therefore, it becomes easier for college students to form marriages without paying the search cost. This reduces search frictions for college graduates and raises their marriage probability. Based on certain assumptions about how the local college marriage market changed in response to the college expansion, we calibrate the model using microdata from before the college expansion. We then simulate the responses of the marriage market to the college expansion using the model. The simulation predicts that the expansion raised the probability of marriage for college women and men.

The college expansion impacted marital outcomes through two channels, according to the model. The first is an adjustment in matching outcomes that was induced by different marginal distributions of education types in the marriage market as also implied in classical matching models. This channel means the expansion potentially reduced college graduates' marriage probability. The second channel is a reduction in average search frictions in the local college marriage market, which is the key innovation of our enriched model. The second channel dominates the first in our model, leading to an overall positive effect of the expansion on college graduates' marriage prospects.

Motivated by the theoretical model, we estimate the causal effects of the college expansion on marriage-market outcomes using a difference-in-differences (DID) design. Exploiting the institutional fact that regions with more pre-existing college resources experienced a larger expansion, we use, as a proxy for college-expansion intensity, a measure that is proportional to historical college enrollment per capita. We compare birth cohorts affected and unaffected by the expansion in places with various levels of this proxy. Under the assumption that in the absence of the expansion, the dynamics of marriage outcomes in locations with different values of the expansion proxy followed parallel trends, our empirical strategy delivers the causal impacts of the expansion on marriage outcomes. We focus on the marriage outcomes of college women and men, but we also look at those of noncollege women and men, as the model also predicts changes in marriage outcomes for these groups because of spillover effects in equilibrium.

The DID estimates show that the college expansion led to a modest increase in marriage probability for both college women and men. When looking at cohorts that went to college five to eight years after the onset of the expansion, a one standard-deviation increase in our expansion-intensity proxy leads to a 2.7 percentage-point increase in marriage probability for college men and a 1.7 percentage-point increase for college-educated women. The results survive a battery of robustness tests. For noncollege groups, we find a relatively small and positive effect of the expansion on the probability of marriage for noncollege men and no effects on noncollege women.

By altering the marriage-market structure, the college expansion may have changed marriage-matching patterns in addition to affecting marriage probabilities. We first look at the effects of the expansion on assortative mating by education level; that is, we examine whether college women and men are now more likely to marry each other. To tease out mechanical effects of the enlarged college population, we construct an index for assortative mating: the difference between the actual probability of college-college matches and the probability of college-college marriages in the hypothetical situation of random matching. We find that the expansion indeed increased the level of assortative mating. Second, we show that the expansion reduced the marriage age gap for college graduates. This finding is also potentially consistent with the story of decreased search frictions in the college marriage market.

This paper contributes to several strands of the literature. First, we contribute to theoretical models of marriage by parsimoniously embedding search frictions in classical matching models. We start with the Choo and Siow (2006) framework, in which marriage matching is transformed into a discrete-choice problem. Chiappori (2017) and Chiappori et al. (2018) add educational choices to this framework. We build into the model a college-specific matching technology. In our model, some people can meet their spouses in college without incurring the search costs that exist for other types of marriage. This reduction in search frictions plays a vital role in explaining what happens in the marriage market following the college expansion. This approach may potentially be used to characterize the marriage market in similar contexts.

Second, our study contributes to a burgeoning literature about how education reforms impact marriage markets. Hener and Wilson (2018) find that a compulsory-schooling reform in the UK reduced the marriage age gap for affected women. André and Dupraz (2019) study a school-construction program in Cameroon and show that a higher level of education leads to a higher likelihood of a polygamous marriage for both men and women. Salcher (2020) finds that girls in Zimbabwe who benefited from an education reform married younger and better educated husbands. Closest to our work is Ge and Huang (2020), which uses China's college expansion as an instrument to estimate the effects of one's own education on marriage and fertility. These studies focus on the partial equilibrium effects of upgrading one's own education, thus neglecting the intrinsic general equilibrium feature of the marriage market as a two-sided matching market. Our study stresses the importance of general equilibrium effects. If such effects matter, then we should interpret the partial equilibrium estimates of the effects of education policies on marriage outcomes with caution. One paper that adopts a similar general equilibrium perspective is Zha (2022), which investigates the effects of the school-construction program in Indonesia on the marriage age gap. Our paper differs from Zha (2022) in two respects. First, we study college education instead of primary education. Second, we show the role of college education in reducing search frictions in the marriage market, which is a novel channel.

Third, our study sheds light on a recent strand of the literature that uncovers the direct role of higher education institutions as marriage markets. Kirkebøen et al. (2021) find that attending a certain college in Norway raises the probability of marrying someone from the same institution. Artmann et al. (2021) document strong assortative mating by field of study in college in the Netherlands and provide causal evidence on the effects of access to specific marriage markets. Using Swedish data, Nybom et al. (2017) show that universities contribute to couples' skill sorting and their children's skill inequality. By exploiting the natural experiment of China's college expansion, we quantitatively show the significance of local college marriage markets when evaluating policies targeted at higher education. A novel finding in our paper is the extensive-margin effect of the expansion on marriage probability. Our paper also confirms the importance of colleges as marriage markets by documenting the effects of the expansion on assortative mating by education and on the marriage age gap. Compared to previous studies in developed economies, our study also provides more relevant insights for less developed countries that are experiencing or will experience expansions in higher education.

Our findings also have implications for some critical issues in China. We find positive effects of the college expansion on the level of assortative mating, which suggests that the expansion could potentially increase inequality and intergenerational persistence of income and social status. Our results are also relevant to China's so-called "leftover women" phenomenon. "Leftover women" is a term used to describe educated women who marry later or are less likely to get married (Fincher, 2016; Magistad, 2013; To, 2015). The reason for the phenomenon is arguably that according to certain social norms, women should "marry up." We find positive effects of the expansion on the marriage probability of college women, suggesting that expanding access to higher education can remove some of the barriers for college women in the marriage market.

The remainder of the paper proceeds as follows. Section 2 develops the model and generates hypotheses. Section 3 describes the context and the data used for our empirical analysis. Section 4 introduces our empirical strategy. Section 5 reports our main results regarding the effects of the college expansion on the probability of marriage. Section 6 presents findings about the effects of the expansion on assortative-mating patterns. Section 7 concludes.

2 Model

In this section, we develop a marriage-matching model with educational investment and fixed search costs.

2.1 Overview of the Model

In the first of two periods, women and men draw their idiosyncratic costs of college education, after which they make the decision whether to attend college. The individual choices determine the distributions of different education types (college or noncollege). Following the education choices, individuals draw their idiosyncratic taste for a spouse, which depends on only the education type of their potential spouse and not the latter's identity. Women and men then match in the marriage market based on education type following a transferable-utility framework (Choo and Siow, 2006).[4]

The model is enriched by including the potential role of colleges in reducing search frictions in the marriage market. This feature of the model incorporates the idea that colleges serve as local marriage markets in which young women and men meet their potential future partners.[5] To enter the marriage market, everyone has to pay a fixed cost of searching for a partner. A subgroup of college women and men, nevertheless, randomly meet each other and form potential matches. If both sides of a potential match agree to marry, then they no longer need to search and therefore do not pay the search cost. We call the set of these meetings the LCMM.

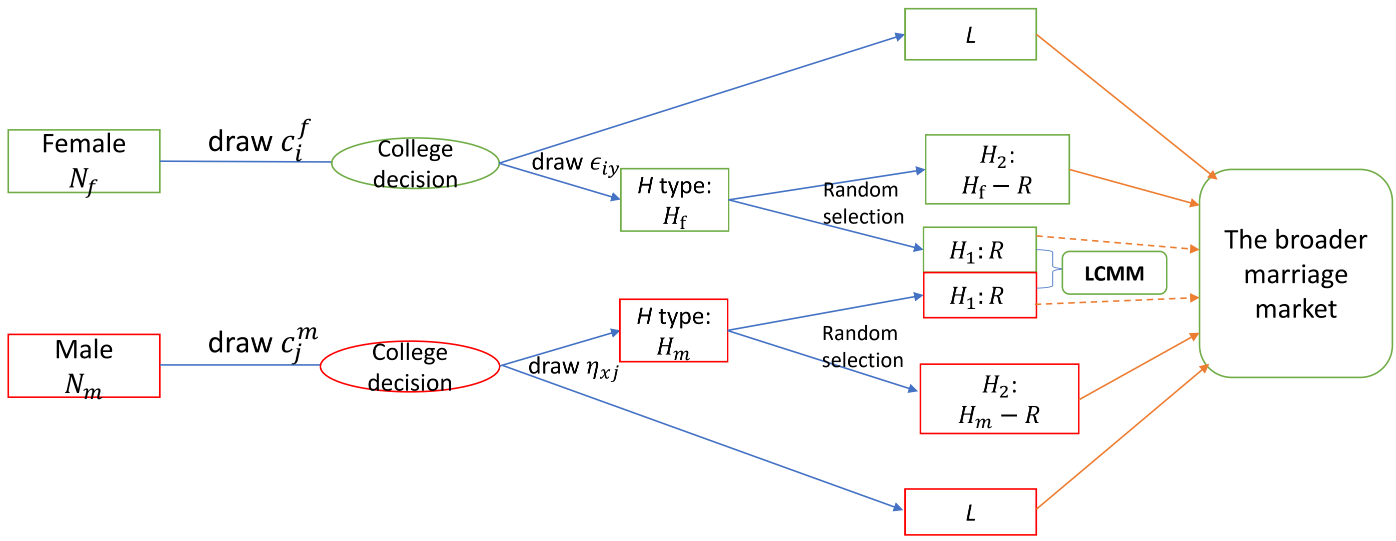
2.2 Model Details

The economy is populated by N_f women and N_m men. We use i as the index for individual woman i and j for man j. There are two periods: $t = 1$ for education choice, and $t = 2$ for marriage-market matching. In Figure 2, we graphically present the structure of the model. We discuss the model details in the rest of the subsection.

[4]In our setting, transferable utility means that husbands and wives send or receive within-marriage utility transfers without loss.

[5]More generally, the social networks developed during college help one find a suitable potential match at one's own college, on a neighboring campus, or even after graduating from college while still in the network.

Figure 2: Model Structure



Notes: Individuals draw their costs of education before choosing whether to get a college education. Before entering the marriage market, they observe their idiosyncratic marital preferences. A certain fraction of college students randomly meet a potential partner in college (type H_1). If they and their potential partners agree to match (and marry) at this stage, they do not need to participate in the costly marriage-search process. If they choose not to match at this stage, they can choose either to stay single or enter the broader marriage market, in which they will have to bear the search cost.

2.2.1 Education Choice

There are two types of educational attainment: college (H) and noncollege (L). At the start of the first period, individuals first draw their idiosyncratic costs of college education and then decide whether to go to college. The cost of education of woman i (man j) is denoted as follows:

$$c_i^f = c_f + \theta_i^f$$
$$c_j^m = c_m + \theta_j^m$$

$c_f(c_m)$ is the average cost of education for the woman (man). θ_i^f (θ_j^m), the individual-specific-shock component of education cost, follows the distribution $G^f(\theta)$ ($G^m(\theta)$).

The group size of college women is denoted by H_f, and that of college men by H_m. We define the college ratios for women and men as $h_f = \frac{H_f}{N_f}$ and $h_m = \frac{H_m}{N_m}$, respectively.

The college expansion is modeled as a reduction in the average-cost parameters (c_f and c_m). We consider the situation in which c_f decreases more than c_m; that is, the education cost decreases faster for women than for men. This situation matches the empirical observation that the college ratio of women increased faster than that of men during the college expansion, as we discuss in more detail in Appendix C.3.2.

2.2.2 Marriage Market in the Second Period

In the second period, conditional on all the individuals' educational choices, women and men meet and match in the marriage market. We adopt the framework developed by Choo and Siow (2006) to characterize the matching process. Individuals in the marriage market match based on their education type. For example, a college man could match with a college woman, match with a noncollege woman, or stay single. Before the matching process starts, each individual draws their idiosyncratic tastes for the education type of their potential partner. This individual taste shock depends only on the potential partner's education type, not on their identity. Moreover, we add to this framework the LCMM to characterize how colleges help reduce search frictions in the marriage market.

2.2.3 Evidence for the Local College Marriage Market

Our assumption about the existence of the LCMM is based on the observation that colleges play an important and direct role in marriage formation in China. We document evidence that supports this view using information from the China Family Panel Studies (CFPS). The CFPS is a nationally representative longitudinal survey. Launched in 2010, the survey covers extensive economic outcomes including family dynamics and relationships. Importantly, the respondents reported how they met with their spouses.

Table 1 reports the fraction of people who met their (first) spouse in school by education type. Comparing the top rows (for college graduates) to the bottom rows (for noncollege individuals), we see that a much higher fraction of college graduates met their spouses in school. On average, more than 20% of college graduates reported that they met their spouses in school. If we look at the statistics before and after the college expansion, 17% of pre-expansion college graduates and 26% of post-expansion college graduates met their spouses in school, and the difference is statistically significant. These statistics provide supportive evidence for the role of colleges as local marriage markets. They also suggest that the importance of LCMMs has increased since the college expansion.[6]

[6]One concern is that individuals in the post-expansion cohorts are younger and mechanically more likely to marry people they met in school. We show in Table B1 that the results are robust to confining the sample to those who married early. The second concern is that the before-versus-after comparison might be contaminated by other secular trends. In Appendix B.1, we provide suggestive causal evidence on the effects of the college expansion by exploiting a DID strategy similar to our baseline econometric specification. Separately, a survey by Wang and Wang (2000) shows that 20.3% of college students in one city were in a romantic relationship in 2000. In a survey by Su et al. (2011), conducted 11 years after the previous survey, this ratio had increased to 46.9%. The numbers suggest that colleges may have played a bigger role in facilitating matching after the college expansion than before.

Table 1: Fraction of People Who Met Their Spouses in School

	Cohorts	Fraction	Observations
College	1975-80 (Pre-expansion)	0.17	313
	1981-88 (Post-expansion)	0.26	1,108
	Difference	0.09 ($p < 0.01$)	
Non-college	1975-80 (Pre-expansion)	0.03	3642
	1981-88 (Post-expansion)	0.05	5169
	Difference	0.02 ($p < 0.01$)	

Source: China Family Panel Studies 2010-2018. All results weighted using the CFPS survey weights.

2.2.4 The Local College Marriage Market

To model the way in which colleges reduce search frictions, we assume that college women and men can meet each other at college. Once they meet, they have the option to get married. We characterize this process using the following meeting function.

Definition 1. The meeting function for the LCMM, $R(H_f, H_m)$, is the number of potential meetings between H_f college women and H_m college men.

$$R = zH_f^a H_m^b$$

Among all the college women and men, R college men *randomly* meet with R college women. Channels for these potential matches, as discussed in the introduction, include various forms of social interactions that provide chances for students of different genders to meet each other. When they meet, they have the option to agree to get married. We assume that the probability of entering the LCMM is independent of student characteristics (including cost of education and marital preferences). Put another way, college students are randomly selected into the LCMM.[7] We denote the college students who are randomly selected into the LCMM (the R college women and men) as type H_1, and the rest of the college students as type H_2.

We assume that the matching function R follows a Cobb-Douglas form, in which z, a, b are constants.[8] We further assume that this matching function exhibits increasing returns to scale ($a + b > 1$). The latter assumption implies that college women and men are more likely to enter the LCMM after the college expansion. The intuition is that as more students enroll in college, the LCMM becomes a thicker market. This implies that with more people

[7]Many of the occasions for social interactions, such as class assignment, are random in Chinese universities.
[8]Similar forms of Cobb-Douglas matching functions are widely used in the literature to characterize the (potential) match between workers and vacancies.

enrolled in college, it becomes more likely for a college student to randomly meet a potential match of the opposite gender via the social interactions in college and, more generally, via the social networks formed in college.

We embed search costs in the model in a parsimonious way by making the simplified assumption that individuals have to pay a fixed cost δ if they choose to enter the marriage market and form a match over staying single. The cost consists of various components, such as money and time spent searching. However, if college woman i and college man j meet in the LCMM and agree to match, they will get married without having to pay the search cost. If they decide not to form a match, then they may choose to either stay single or enter the broader marriage market (Figure 2), for which they will need pay the search cost.[9] We formally describe the payoffs and search cost in the marriage market in the following section.

2.2.5 Payoffs in the Marriage Market

The payoff of marriage is determined by one's own type, the type of one's spouse, and one's unique taste for marriage. For woman i of type x who marries a man of type y, her individual marital payoff is as follows:

$$u_{ixy} = \alpha_{xy} + \tau_{xy} - \delta \mathbb{1}[y \neq 0] \max\{\mathbb{1}[x \neq H_1], \mathbb{1}[y \neq H_1]\} + \epsilon_{iy}$$

The marital payoff to man j of type y is as follows:

$$v_{xyj} = \gamma_{xy} - \tau_{xy} - \delta \mathbb{1}[x \neq 0] \max\{\mathbb{1}[x \neq H_1], \mathbb{1}[y \neq H_1]\} + \eta_{yj}$$

Here, $x, y \in \{H_1, H_2, L, 0\}$. τ_{xy} is the utility transfer within the marriage, which represents how the joint marital surplus is allocated. The value of τ_{xy}, which could be positive or negative, is determined in equilibrium.[10]

Both sides of the match have to pay the fixed cost δ if they decide to enter the marriage market and search for a spouse rather than stay single. α_{xy} and γ_{xy} are the systematic marital payoffs. The only exception is when both sides of a match are in the LCMM *and* agree to marry each other. They can then avoid paying the cost. Their payoffs are thus determined

[9] For example, if a college graduate draws a strong preference for a noncollege partner, then they will reject the potential college-college match even if they randomly enter the LCMM.

[10] We denote the case in which woman i stays single $y = 0$ and that in which the man j stays single $x = 0$. $\tau_{x0} = \tau_{0y} = 0$. In this transferable-utility framework, there is no loss of surplus in the transfer of τ_{xy}. The first subscript always refers to the wife's type, and the second always refers to the husband's type.

as follows:

$$u_{iH_1H_1} = \alpha_{HH} + \tau_{H_1H_1} + \epsilon_{iH}$$
$$v_{H_1H_1j} = \gamma_{HH} - \tau_{H_1H_1} + \eta_{Hj}$$

This is a special case of the payoff functions above. The H_1 and H_2 types generate the same systematic marital returns except that there is no fixed search cost for the H_1H_1 match.[11] The H_1 type and the H_2 type are valued the same way by their potential spouses in terms of idiosyncratic marital preferences.[12]

2.3 Equilibrium

2.3.1 The Marriage Market

Given the preferences and the utility transfer (τ), each individual determines their preferred partner type. The aggregated demand and supply for a given type of match are equal to each other in equilibrium. Specifically, we assume that μ_f^{xy} women of type x choose type-y men and μ_m^{xy} men of type y choose type-x women. Then, in equilibrium,

$$\mu_f^{xy} = \mu_m^{xy} = \mu^{xy}.$$

The equilibrium conditions pin down the transfer τ_{xy}. The matching function μ_{xy} represents the number of type-xy matches in equilibrium. μ_{x0} is the number of type-x women that stay single, and μ_{0y} is the number of single type-y men.

2.3.2 The Educational Choice

Now we return to the educational investment choice. In the first period, after observing idiosyncratic costs of education, each individual decides whether to attend college, anticipating what will happen in the marriage market. Since they do not observe their idiosyncratic preferences for marriage types yet, they only take into account the expected payoff of a given education type (H or L).[13]

For woman *i* who chooses education type H, the expected payoff is $U^H - c_i$. Her expected payoff of choosing education L is U^L. (U^H, U^L), which represent the expected marital payoffs to different education types (before paying the education costs) are determined in the following way. The expected payoff of choosing noncollege (L), U^L, is the maximal payoff

[11] That is, $\alpha_{xH_1} = \alpha_{xH_2} = \alpha_{xH}, \alpha_{H_1y} = \alpha_{H_2y} = \alpha_{Hy}, \gamma_{xH_1} = \gamma_{xH_2} = \gamma_{xH}$, and $\gamma_{H_1y} = \gamma_{H_2y} = \gamma_{Hy}$.

[12] That is, $\epsilon_{iH_1} = \epsilon_{iH_2} = \epsilon_{iH}, \eta_{H_1j} = \eta_{H_2j} = \eta_{Hj}$.

[13] This also rules out the concern of multiple equilibria, which might occur if individuals observe their idiosyncratic marital preferences before the education decision.

determined by one's optimal choice of spouse:

$$U^L = E(u_{iLy}|y = arg \max_{y=0,H,L} u_{iLy})$$

If an individual chooses college education (H), the expected payoff follows the same structure but also depends on whether they randomly enter the LCMM:

$$U^H = \sum_{k=1,2} p^f(H_k)E(u_{iH_ky}|y = arg \max_{y=0,H_k,L} u_{iH_ky})$$

Here, $p^f(H_1) = \frac{R}{H_f}$ is the probability of being selected into the LCMM conditional on attending college.[14] $p^f(H_2)$ is the probability of not being selected into the LCMM conditional on attending college, and $p^f(H_2) = 1 - p^f(H_1)$.

Woman i therefore chooses college education (H) based on the following decision rule:

$$U^H - c_i - U^L \geq 0$$

Man j faces a symmetric problem:

$$V^L = E(v_{xLj}|x = arg \max_{x=0,H,L} v_{xLj})$$
$$V^H = \sum_{k=1,2} p^m(H_k)E(v_{xH_kj}|x = arg \max_{x=0,H_k,L} v_{xH_kj})$$

Here, $p^m(H_1) = \frac{R}{H_m}$ is the probability of entering the LCMM and $p^m(H_2) = 1 - p^m(H_1)$. Man j follows a similar decision rule to woman i regarding college education:

$$V^H - c_j - V^L \geq 0$$

We describe the equilibrium of the model in detail in Appendix C.1. The equilibrium is determined by individual educational choices and the marriage market equilibrium. Individual educational choices are determined by expected payoffs in the marriage market. Marriage-market returns are determined by the distribution of education types and by individual marriage choices. The equilibrium is characterized as a fixed point in which individual education choices are consistent with marital returns to education. The existence of the equilibrium is guaranteed by Brouwer's fixed-point theorem.

[14] When making the education choice, individuals do not observe their marital preferences or whether they will enter the LCMM. Therefore, only the expected payoffs matter for their choices.

2.4 Comparative Statics

To illustrate the intuition about the impacts of the college expansion on the marriage market, we report marital outcomes as a function of the distributions of college women (h_f) and college men (h_m). For the full model, both the marital outcomes and education outcomes are functions of the exogenously shifted mean costs of education (c_f and c_m). In Appendix C.4, we report the comparative statics of the full model, including how educational-attainment and marriage outcomes respond to the college expansion.

2.4.1 Calibrating Model Parameters

We start by calibrating the parameters in the model using pre-expansion marriage-market data. Specifically, the systematic-returns-to-marriage parameters are pinned down using marriage patterns on pre-expansion cohorts in the 2005 China mini-census data. We construct a data set for married couples aged between 27 and 40 years old in 2005, which provides us with a snapshot of the marriage market prior to the expansion. Crucial to our model are the parameters for payoffs of marriage. Following the Choo-Siow framework, we can show that for a given type of match xy that is *not* formed via the LCMM (that is, x or y is not type H_1), the joint surplus and the matching function is as follows:

$$\frac{(\mu^{xy})^2}{\mu^{x0}\mu^{0y}} = \exp(\alpha_{xy} + \gamma_{xy} - \alpha_{x0} - \gamma_{0y} - 2\delta) \tag{1}$$

The right-hand side is the exponential form of the joint systematic surplus of marriage over staying single. Intuitively, a higher surplus is associated with a larger measure of the corresponding type of marriage xy. For marriages formed via the LCMM ($x = H_1, y = H_1$), the search cost is avoided. The relationship is as follows:

$$\frac{(\mu^{H_1H_1})^2}{\mu^{H_10}\mu^{0H_1}} = \exp(\alpha_{HH} + \gamma_{HH} - \alpha_{H0} - \gamma_{0H}) \tag{2}$$

The LCMM, however, introduces additional complications in calibrating the parameter for the search costs (δ) because we do not distinguish between type H_1 and H_2 in the data. In order to calibrate δ, we use auxiliary information from the CFPS about the fraction of college-college marriages that are formed via meeting in school. Intuitively, conditional on the matching function R and the marginal distributions of education types, a higher fraction of college-college marriages that are formed by meeting in school should be associated with a higher fixed search cost. We draw from the CFPS a variable λ defined as the fraction of college-college marriages via the LCMM out of all college-college marriages:

$$\lambda \stackrel{\text{def}}{=} \frac{\mu^{H_1 H_1}}{\mu^{HH}}$$

In addition, conditional on the LCMM meeting function R, we have the following definitions of variables:

$$J_f \stackrel{\text{def}}{=} \frac{R - \mu^{H_1 H_1}}{H_f - R - \mu^{H_2 H_2}}$$
$$J_m \stackrel{\text{def}}{=} \frac{R - \mu^{H_1 H_1}}{H_m - R - \mu^{H_2 H_2}}$$

It turns out that

$$\delta = \ln \frac{\lambda}{1-\lambda} - 0.5 \ln J_f J_m. \tag{3}$$

Equations 1, 2, and 3 enable us to identify the search-cost parameter together with payoffs to marriages conditional on the meeting function R. We describe the procedure in detail and prove Equation 3 in Appendix C.2.

2.4.2 Meeting Function for Local College Marriage Market

The function $R = R(H_f, H_m)$ is important for both calibrating model parameters and simulating our model's comparative statics. Without direct information on the matching process, unfortunately, we cannot pin down its functional form. Based on the assumption of increasing returns to scale, we set the function as follows:

$$R = 0.5 H_f^{0.75} H_m^{0.75}$$

The choice of the meeting function is somewhat ad hoc: we choose a functional form that can generate the comparative statics that are largely consistent our main empirical findings. That is, under reasonable assumptions about the meeting function, the simulation based on the theoretical model can explain our empirical results.

2.4.3 Social Norms against Marrying Down among Women

An important feature of China's marriage market is the social norm of aversion to seeing women marry lower-status men ("marrying down"). It is therefore much less likely for highly educated women to marry less educated men than the reverse (Figure C2). This norm is formalized in our model by the marriage-payoff parameters: the marital surplus for college women who marry down is less than that for noncollege women who marry up (the inequality below). This is consistent with the marital-payoff parameters estimated using observed data.

$$\alpha_{HL} + \gamma_{HL} - \alpha_{H0} - \gamma_{0L} < \alpha_{LH} + \gamma_{LH} - \alpha_{L0} - \gamma_{0H}$$

2.5 Simulating the Probability of Marriage

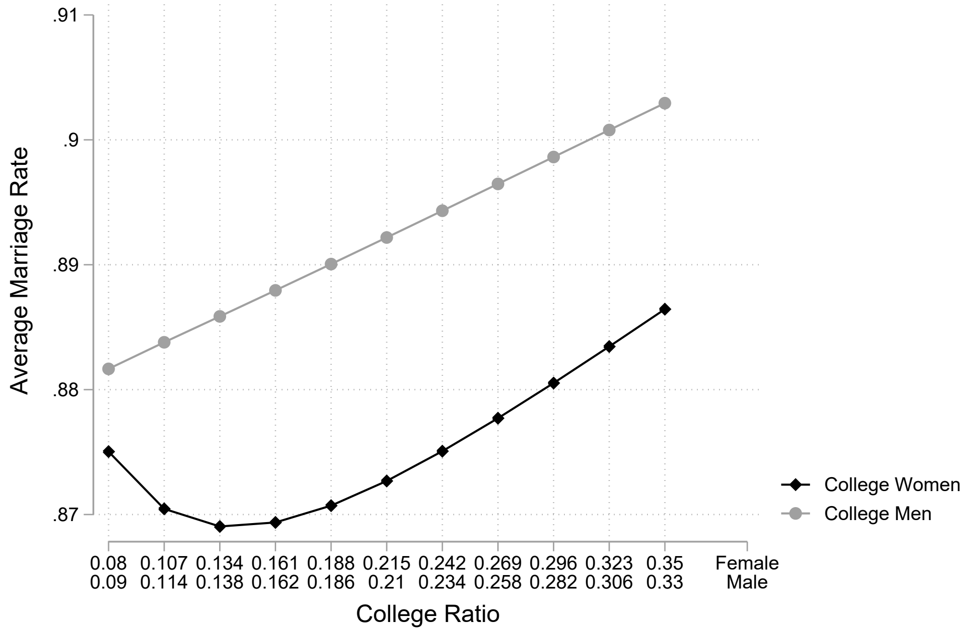
Based on the estimated and calibrated parameters, we simulate marriage-market responses to the college expansion. The college ratio of women is set to be initially lower than that of men based on the data moments for pre-expansion marriages (0.08 and 0.09) used to calibrate the baseline model's parameters. We allow the college ratio of women to increase faster than that of men. Specifically, we set the ending values of the college ratios for women and men as 0.35 and 0.33, respectively. These values reflect the distribution of education types for post-expansion cohorts (1985–88) in high-expansion regions. High-expansion regions are defined based on the value (above the median) of our empirical proxy for the magnitude of the college expansion, which we define in Section 3.2. These data moments also reflect an important feature of the expansion: the college ratio of women increased faster than and overtook that of men (further supporting evidence is discussed in Appendix C.3.2).

The data moments we use for the simulation mainly reflect the temporal variation in the national distribution of education types.[15] For our empirical analysis, however, a before-versus-after comparison does not serve as a reliable identification strategy because it is very likely confounded by other secular trends. Therefore, later in Section 4, we resort to a DID design to empirically estimate the effects of the expansion on the marriage market.

Figure 3 reports the simulated marriage probabilities of college graduates as a function of the college ratios of women and men based on our model. Overall, the simulated results predict an increase in the marriage probability of both college men and women, except for college women early on in the expansion. Our model can help us further disentangle different mechanisms.

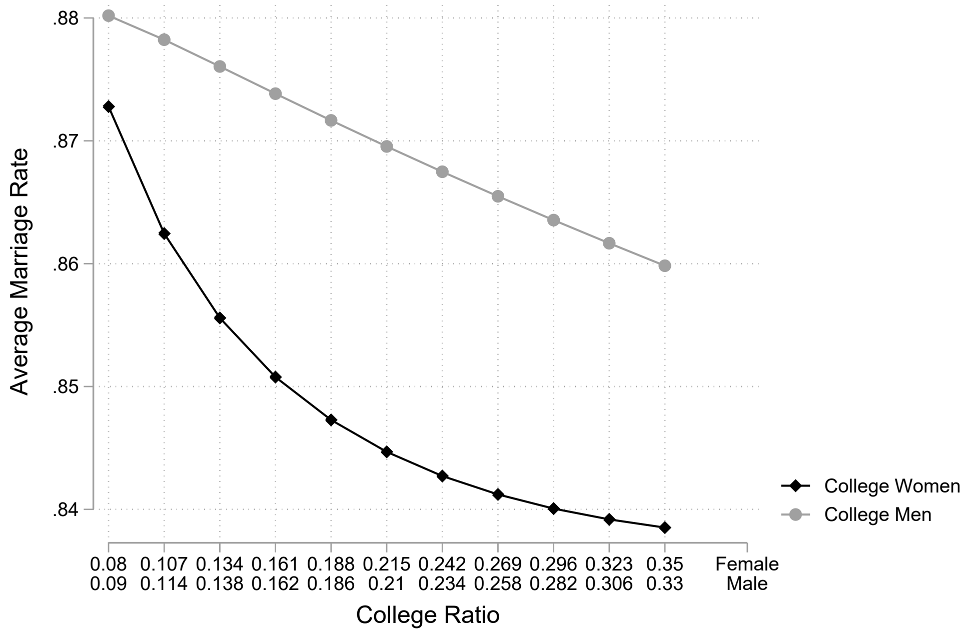
[15] For the post-expansion cohorts, we use data in high-expansion regions. This partially incorporates cross-sectional variation.

Figure 3: College Expansion and Marriage Rates: Simulated Results



Notes: The x-axis displays the college ratios of women and men, which are allowed to evolve simultaneously.

Figure 4: College Expansion and Marriage Rates: Simulated Results without Local College Marriage Market



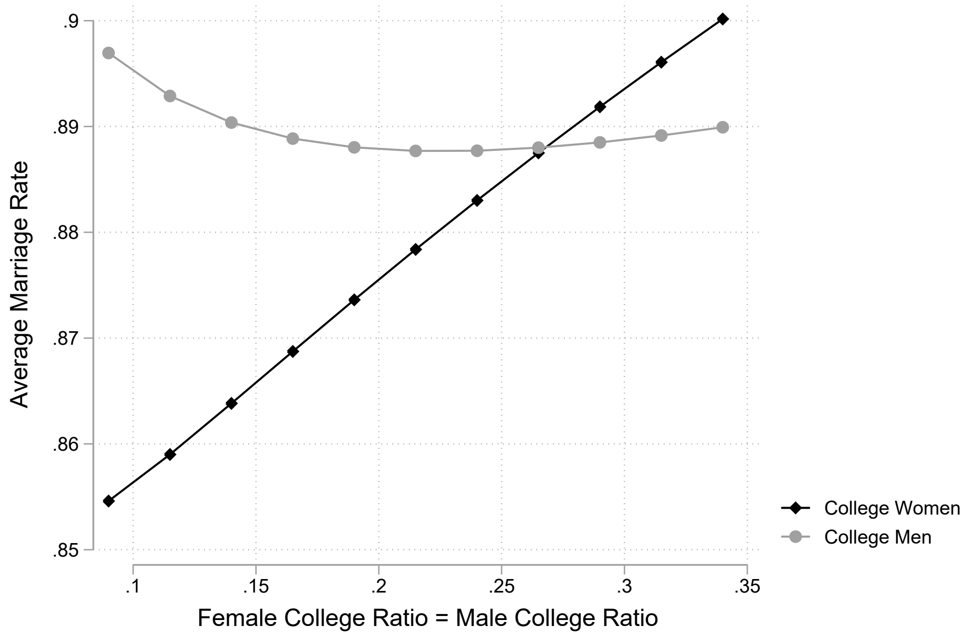
Notes: The x-axis displays the college ratios of women and men, which are allowed to evolve simultaneously.

The Local College Marriage Market Reduces Search Frictions. Through the LCMM channel, the college expansion boosts the marriage probability of college graduates by reducing search frictions. A thicker LCMM makes it easier for college students to find a preferable match within the pool of college-educated individuals. Ceteris paribus, this force pushes up the marriage probability for both college men and women. To evaluate the effects of this channel, we redo the simulation without the LCMM in the model. The results are reported in Figure 4. The marriage probabilities of both college women and men decrease following the college expansion. This is driven by the increasing relative supply of college types in the marriage market: because there are more college women and men, their bargaining power within marriage now decreases relative to that of noncollege types. The reduction in search frictions via the LCMM dominates the effects of changing distributions of education types in the marriage market, causing an overall positive effect as shown in Figure 3.

Change in Distributions of Education Types. As discussed above, a change in the distributions of education types leads to decreasing marriage probabilities for both college women and men (Becker, 1973; Choo and Siow, 2006). Importantly, the education of women increases faster than that of men. As the group size of college women grows faster than that of college men, the former become less scarce and hold less bargaining power in college-college marriages. As a result of this relative change in the group size in the college-specific marriage market, the marriage probability of college women tends to decrease by more than that of college men, as shown in Figure 4. Even in Figure 3, we also observe an initial decrease in the marriage probability of college women. The pattern implies that the effect of the LCMM dominates that of changing education distributions for college women only later, when the college expansion becomes intense enough.

How important is the gender difference in the rate of college expansion? To look into this, we simulate the baseline marriage model by allowing the college ratios of women and men to increase symmetrically. The results, as in Figure 5, show that the effect of the expansion is positive for college women and initially negative but later positive for college men. The reason is the social norms regarding marital preferences (Section 2.4.3). Following an increasing supply of college men, their bargaining power relative to noncollege men decreases. College women, however, rarely marry noncollege men and do not suffer from this devaluation. College men are now willing to transfer more to college women since college men are less attractive to noncollege women. As a result, the marriage rate of college women increases, while that of college men tends to decrease. The LCMM mechanism further contributes to the positive effect on college women and dominates the negative effect on college men later in the expansion.

Figure 5: College Expansion and Marriage Rates: Simulated Results under Gender-Symmetric Expansion



Notes: The x-axis displays the college ratios of women and men, which are allowed to evolve simultaneously.

Table 2: Model Predictions under Different Assumptions

		College Women	College Men
LCMM	Gender Asymmetric Expansion (Faster for Women)	$- \rightarrow +$	$+$
	Gender Symmetric Expansion	$+$	$- \rightarrow +$
No LCMM	Gender Asymmetric Expansion (Faster for Women)	$-$	$-$

Table 2 summarizes the model predictions about the effects of the college expansion on college graduates' marriage probabilities under different assumptions and hypothetical expansion scenarios. Both the LCMM and the change in relative distributions of education types matter for marriage outcomes according to our model and simulation. The enriched model with an LCMM predicts overall positive effects for both college women and men. As we show below, this is consistent with our empirical findings.

In Appendix C.3.3, we discuss the simulation for noncollege groups. In Appendix C.3.4, we report how the within-marriage transfers (τ) change in response to the college expansion in the three hypothetical situations discussed above, respectively. The changes in transfers are consistent with the discussed mechanisms. For example, the college expansion increases the transfer from college men to noncollege women and reduces the transfer from noncollege men to college women in all the three scenarios. When there is no search cost and the college enrollment increases faster for women than for me as we specified in the model, the transfer from college men to college women decreases. When the college ratio increases symmetrically between women and men, the transfer from college men to college women increases.

3 Background and Data

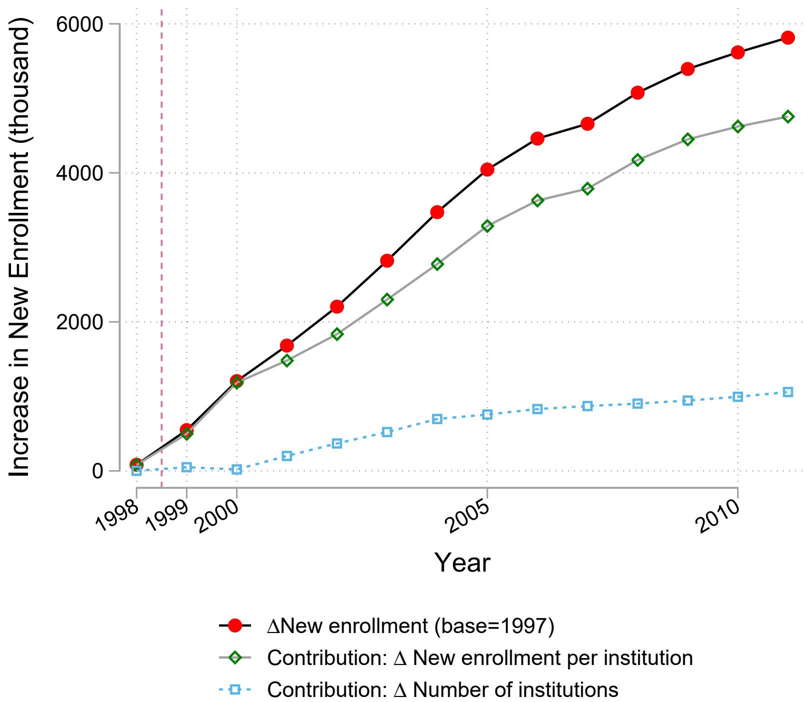
3.1 The Higher Education System and College Expansion in China

Most Chinese universities are public. The process of college admission is strictly controlled by the MOE, which also determines total college enrollment in the country (Feng, 1999). The amount of annual enrollment is set on the basis of the MOE's five-year plans. Each university closely adheres to its assigned quota when setting its admission plans each year. Among the different types of colleges, we focus on the regular college system ("regular colleges"), which consists of four-year and three-year colleges (the latter are analogous to two-year colleges in the United States). Though there is also a part-time postsecondary credential system, which mainly serves adults who are older than regular college students, only regular colleges experienced the expansion. In Appendix A, we provide more details about the college system.

The expansion represented a sharp change in the MOE's plan. The government abruptly decided in 1999 to expand access to college in order to accommodate more youth at risk of unemployment in response to the 1997 Asian financial crisis (Wang, 2014). The policy continued even after the effects of the financial crisis subsided. Figure 1 displays the increase in new college enrollment nationwide before and after the college expansion. The expansion, which doubled the amount of new enrollment within three years after 1999 and even more afterward, has sharply changed the levels in education in relevant cohorts of young women and men.

The MOE implemented the college expansion mainly through scaling up enrollment in existing colleges. In Figure 6, we decompose the rise in total new enrollment into the increase in average enrollment per institution and the increase in the number of institutions. The overall increase is mainly driven by existing colleges. This pattern is also consistent with the assumption that the expansion has led to a thicker college marriage market (Sections 2.2.3 and 2.2.4).

Figure 6: Decomposing the Increase in New Enrollment



Notes: The data come from Chinese Education Yearbooks. We plot (1) the total increase in new enrollment relative to 1997; (2) the increase in new enrollment if the number of institutions did not change and the average new enrollment per institution increased as actually happened; (3) the increase in new enrollment if the average new enrollment did not change and the number of institutions increased as actually happened.

3.2 Measuring Exposure to the College Expansion

To identify the causal impact of the college expansion on the marriage market, we need to account for secular trends in socioeconomic and marriage-market conditions that might create a spurious relationship between being exposed to the expansion and marriage outcomes. We tackle this problem with a DID design that exploits geographic and birth-cohort variation in exposure to the expansion.

The typical age of college enrollment is 18. Because the college expansion started in 1999, we consider cohorts born in and after 1981 as post-expansion cohorts that were directly exposed to the expansion. We compare the marriage outcomes of these cohorts with those of the pre-expansion cohorts (those born before 1981). In our baseline analysis, the 1975–78 cohorts are used as the pre-expansion comparison cohorts. We discuss the choice of post- and pre-expansion cohorts in Section 4.

Critical to our empirical strategy is variation in the intensity of the college expansion across regions. The expansion was implemented mainly via scaling up enrollment in existing universities. As a result, regions with more historical higher education resources naturally benefited more from the expansion. This regional variation has been exploited in various previous studies on the expansion (Feng et al., 2018; Ge and Huang, 2020; Li et al., 2017; Ma, 2020). Motivated by this variation, we construct a proportional-expansion proxy based on the historical abundance of university resources.

In our baseline analysis, we look at local marriage markets at the prefecture level.[16] For each prefecture p, we construct the proportional index as follows:

$$ExpProxy_p = \frac{CollegeEnrollment_p^{1982}}{PopSize_p^{1982}} \tag{4}$$

This proportional-expansion proxy measure is the ratio of college enrollment to population size as of 1982. It is constructed using microdata from the 1982 China census provided by IPUMS International. In Figure 7, we show the geographical distribution of the intensity of the expansion as measured by this proxy. Because there is substantial variation in historical college resources across prefectures, the same is true for the proxy. High-intensity regions (as measured by this proxy) do not show obvious geographic patterns, with the one exception that prefectures in the Northeast tend to have higher intensity.

This proxy is highly predictive of college enrollment in later years. Figure 8 plots the (log) enrollment ratio in 2005 against the (log) expansion proxy. Their correlation can be approximated using a straight fitted line with a slope of 0.67 and an R-squared of 0.44.[17] Therefore, the expansion proxy provides sufficient variation for identifying differing responses of local marriage markets to the expansion.

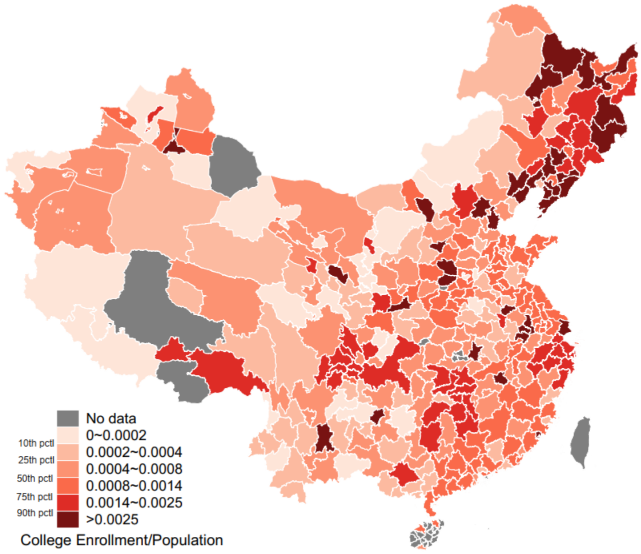
In Appendix D, we provide several additional tests and arrive at two findings. First, the power of the expansion proxy to predict college enrollment in later years is robust across years. Second, the treatment proxy is not associated with economic growth or sex ratio, both of which can affect the marriage market in important ways (Burgess et al., 2003; Chu et al., 2018; Hankins and Hoekstra, 2011; Wei and Zhang, 2011; Ebenstein and Sharygin, 2009). The orthogonality of the treatment proxy in relation to these economic and marriage-market conditions also provides supportive evidence for the parallel-trends assumption of our DID design.[18] Previous studies on China's college expansion have also discussed the validity of exploiting variation in historical college endowments. Feng et al. (2018), Li et al. (2017), and Ma (2020) show that historical college endowment is highly predictive of the number of college graduates after the expansion, and Ma (2020) documents that college endowment in 1982 is not associated with changes in GDP or population.

[16]The prefecture is the subprovincial geo-unit in China. We constructed a consistent set of prefecture units to accommodate historical division changes. There are 340 prefecture units in our sample for analysis.

[17]If college enrollment in later years is perfectly proportional to the initial endowment measured using the 1982 data, the slope of the line should be 1.

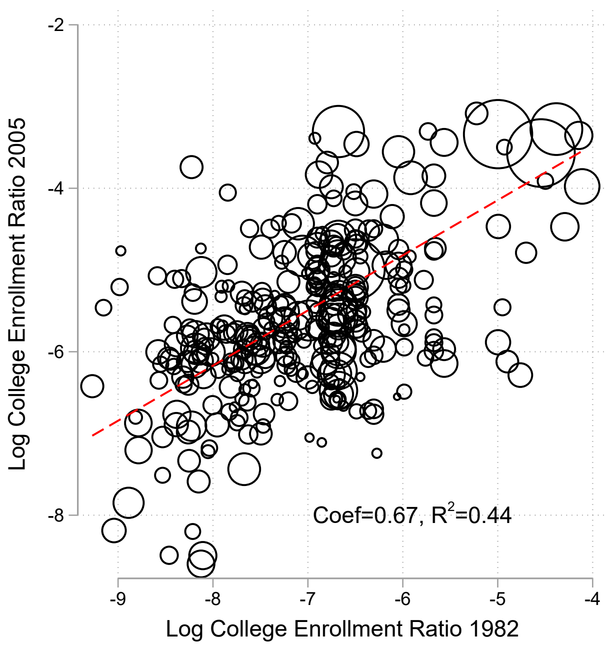
[18]In our empirical analysis, we also provide robustness tests that directly control for local GDP per capita and sex ratio.

Figure 7: Geographical Distribution of the Proxy for College-Expansion Intensity



Notes: The proportional proxy for the magnitude of the college expansion is constructed using 1982 census microdata of China, obtained via IPUMS International.

Figure 8: Initial College-Expansion Proxy and College Enrollment in 2005



Notes: The 1982 ratio of college enrollment to population is constructed using 1982 Chinese-census microdata obtained via IPUMS International. The 2005 ratio of college enrollment to population comes from the 2005 Chinese mini-census. The plots are weighted using the population size of each prefecture.

3.3 Marriage Outcomes

We obtain information about marriage outcomes using the confidential 2010 Chinese census and the confidential 2015 Chinese mini-census. The 2015 mini-census sample that we use is a 0.15% random sample of the population. The 2010 census sample is a 0.35% random sample of the population.[19] For our post-expansion cohorts (born after 1981), we construct their marriage outcomes from the 2015 mini-census data. Corresponding marriage outcomes of the pre-expansion cohorts are constructed using the 2010 census data. The 2010 census data set contains information about individuals' marital history.[20] For example, when we choose the 1975–78 birth cohorts as the pre-expansion cohorts and the 1981–84 birth cohorts as the post-expansion cohorts, we impute the marriage outcomes for the pre-expansion cohorts based on their marital status as of 2009, when they were aged 31–34, the same ages as the post-expansion cohorts (born 1981–84) in 2015.

4 Empirical Strategy

4.1 Baseline Empirical Strategy

In our DID design, we compare cohorts exposed to the expansion (post-expansion cohorts) to cohorts not exposed to it (pre-expansion cohorts) in different local marriage markets. We consider cohorts born after 1981 as the post-expansion cohorts. We divide the post-expansion cohorts into two groups: (1) the 1981–84 birth cohorts ("early post-expansion cohorts"), and (2) the 1985–88 birth cohorts ("late post-expansion cohorts"). The former group was 31–34 years old as of 2015, while the latter was 27–30 years old. Therefore, these two groups capture the impacts of the expansion on different cohorts at different ages.

The choice of pre-expansion cohorts requires further discussion. For two reasons, cohorts born too close to 1981 do not constitute good comparison units. First, the age of college enrollment at 18 is a norm, not a binding constraint. Therefore, cohorts born slightly before 1981 were still partially exposed to the expansion because they enrolled in high school late or retook the college entrance examination. Second, and more importantly, as the marriage market is a two-sided matching market, cohorts that were not directly exposed to the expansion were very likely affected in general equilibrium, as some of their potential partners were exposed to the expansion. Therefore, the ideal pre-expansion cohorts are far enough from the onset of the expansion that they are not subject to such spillover effects. However, choosing control cohorts born too much earlier than the treatment may make the two groups less comparable. This is a concern since China experienced significant cultural, social, and economic shifts when the cohorts that we investigate grew up.

[19]Both data sets are accessed via the Shanghai University of Finance and Economics.

[20]Unfortunately, similar marital-history information does not exist in the 2015 mini-census data set.

For our baseline analysis, we use the 1975–78 cohorts as the pre-expansion comparison group. These cohorts reach college age as close to the expansion shocks as seems sensible given the potential spillovers. These cohorts serve as the comparison group for both the 1981–84 and 1985–88 post-expansion cohorts. To make the marriage outcomes comparable between the post-expansion and the pre-expansion groups, we draw information from marital histories recorded in the 2010 census for the pre-expansion cohorts. The early post-expansion cohorts (born in 1981–84) were observed in 2015, when they were 31–34 years old. For the 1975–78 cohorts, we construct their marriage outcomes as of 2009, when they were similar in age to the post-expansion cohorts in 2015. Meanwhile, the late post-expansion cohorts (born in 1985–88) were 27–30 years old in 2015. Therefore, when we use the 1975–78 cohorts as their comparison group, we look at the marriage outcomes of the earlier cohorts as of 2005, when they were also comparable in age.

We estimate the causal impacts of the college expansion on marriage outcomes using the following specification:

$$y^k_{iapt} = \beta^k_0 + \beta^k_1 ExpProxy_p * Post_t + \beta^k_2 * Post_t + \lambda^k_p + \xi^k_a + \varepsilon^k_{iapt} \tag{5}$$

y^k_{iapt} is the outcome (for example, ever married) of individual i at age a in local marriage market (prefecture) p. The subscript t captures whether the observation belongs to a post-expansion or pre-expansion cohort. $Post_t$ is a dummy variable, with $Post_1 = 1$ indicating post-expansion cohorts. We also control for prefecture (λ^k_p) and age (ξ^k_a) fixed effects. $ExpProxy_p$ is the proxy for local exposure to the expansion (described in Section 3.2).

In the following analysis, we standardize the treatment proxy $ExpProxy_p$ so that one unit represents one standard deviation (SD) across all prefectures. Thus, the coefficient on the interaction term $ExpProxy_p * Post_t$ estimates the change in marital outcomes in a prefecture if we increase the treatment proxy by one SD. A one-SD increase in the treatment proxy is also approximately twice the size of the increase from the 25th percentile to the 75th percentile in the distribution across prefectures. As an example, a one-SD difference in the proxy is approximately the difference between Shanghai, one of the most developed and educated cities in China, and Wuhu, an inland prefecture that ranked 235th in population size and 82nd in GDP in 2015.[21]

4.2 Pre-expansion Cohorts and Graphical Evidence

We choose the 1975–78 cohorts as the pre-expansion cohorts in our baseline analysis in an attempt to resolve the tension between choosing pre-expansion cohorts that are too far from

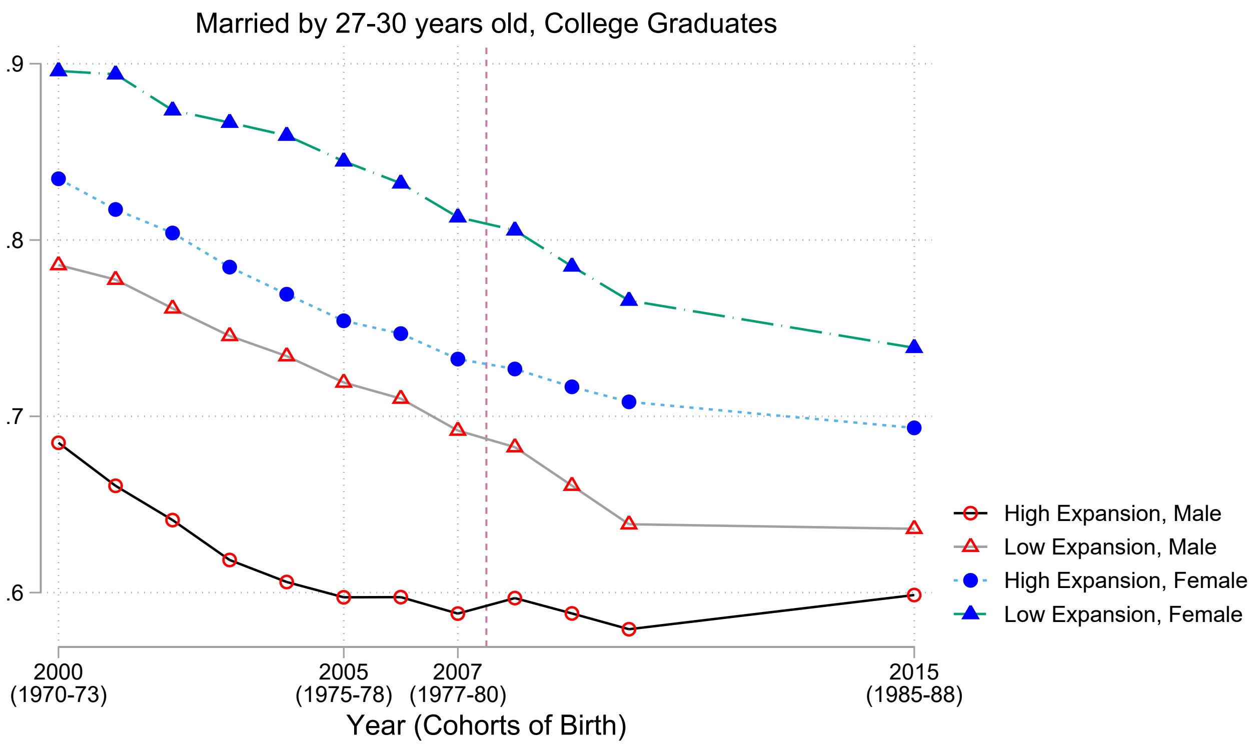
[21] We do not weight using population size when standardizing the treatment proxy across prefectures. Weighting the proxy leads to very similar results: the treatment proxy is only rescaled by a constant factor, and the magnitudes are similar whether weighted or not. We execute all regressions at the individual observation level, effectively weighting the regressions by population size.

the expansion and those too close to it. However, it is helpful to check whether the secular trends in marriage patterns in the pre-expansion cohorts are parallel between regions with higher exposure to the expansion and regions with lower exposure. Largely parallel pre-trends can provide supportive evidence for the parallel-trends assumption. In addition, they may suggest that the tension regarding the choice of pre-expansion cohorts does not greatly affect our findings. In this section, we provide preliminary graphical evidence on the pre-trends.

We divide the local marriage markets (prefectures) into two groups: those with high expansion intensity and those with low intensity. The former group includes prefectures whose proxy for expansion is above the median, and the rest are included in the latter. The fraction of college graduates who were ever married at 27–30 years old is plotted in each year for both groups. The results are displayed in Figure 9 (we report the graphical evidence for parallel pre-trends of noncollege groups in Figure F14). Most college graduates in the early-1970s cohorts were already married at this age (nearly 90% of college women and about 75% of college men). The marriage probabilities declined overall in subsequent cohorts. For the 1985–88 cohorts (late post-expansion cohorts), the numbers are more than 70% for college women and more than 60% for college men. The marriage probabilities have consistently been higher in low-expansion regions than in high-expansion regions.

In Figure 9, left of the vertical line are the pre-expansion cohorts. Looking through these cohorts, it seems that the high- and low-expansion regions followed largely parallel trends before being affected by the expansion. We conduct formal tests regarding parallel pre-trends in Section 5.2, the results of which are consistent with the preliminary graphical evidence. After the expansion, the differences between the high- and low-expansion regions decreased. This decrease suggests the expansion increased marriage probabilities if we assume that the high- and low-expansion regions would have continued the parallel trends in the absence of the expansion.

Figure 9: Trend in Marriage Rates of College Graduates by College-Expansion Intensity



Notes: The sample is divided into two groups based on the value of the expansion treatment proxy below or above the median. Left of the vertical line are samples that consist of only pre-expansion cohorts. The marital-history information in the 2010 census allows us to impute the ever-married fractions before 2010. Using the 2015 mini-census, however, we can only know people's marital status as of 2015.

5 Impacts of the College Expansion on Marriage Probability

5.1 Baseline Results

In this section, we quantify the effects of the college expansion on marriage probability. We focus on college women and men. Our model predicts that their marriage rates increased in response to the expansion. We test this prediction using the DID strategy discussed in Section 4.1.

We report the baseline estimates using the DID specification (Equation 5) in Table 3. The first two columns present estimates (separately for college and noncollege) in which we examine the effects on the late post-expansion cohorts (1985–88) and measure their marriage rates at ages 27–30. Columns (3) and (4) present estimates for the early post-expansion cohorts (1981–84), whose marriage rates are measured at ages 31–34. Here we focus on the estimates for college graduates (Columns (1) and (3)). In Panel A, we present the estimates for men, and in Panel B, we present the estimates for women. The gender gap in the effects on the same cohorts is reported in Panel C.

The results suggest positive and modest impacts of the expansion on the probability of marriage for college graduates. Column (1) shows that when we look at the late post-expansion cohorts (27 to 30 years old in 2015), a one-SD increase in the expansion-treatment proxy raised the marriage probability by 2.7 percentage points for college men and 1.7 percentage points for college women.[22] If we compare a prefecture at the 75th percentile to one at the 25th percentile of the distribution of the expansion proxy, the expansion led to an increase of 1.44 percentage points in the marriage probability of college men from the late post-expansion cohorts and 0.66 percentage points for their college-women counterparts.

The magnitudes of these effects, though not very large, are still meaningful. For example, they are comparable to the changes in the mean marriage probabilities presented in Table 3. For college men in the late post-expansion cohorts (born in 1985-88), the increase in marriage probability driven by a one-SD increase in the treatment proxy (2.7 percentage points) is 88% of the the overall decrease in marriage probability of college men aged 27–30 between the pre-expansion and late post-expansion cohorts (3.1 percentage points). Looking at college women from the late post-expansion cohorts, the effect of a one-SD increase in the treatment proxy (1.7 percentage points) is equivalent to 22% of the overall decline in marriage probability (7.8 percentage points).[23]

The effects are smaller for college graduates from the early post-expansion cohorts (31–34 years old in 2015). Column (3) shows that a one-SD increase in the treatment proxy leads to an increase of 1.24 percentage points in marriage probability for college men and 0.53 percentage points for college women. Moreover, the estimated effect on college women is statistically nonsignificant. The effect on college men is still meaningful, as it is equivalent to 50% of the decrease in marriage probability between the pre-expansion and early post-expansion cohorts (2.5 percentage points).

[22]The magnitudes of these effects are approximately those of college graduates in Shanghai versus Wuhu, as discussed in Section 4.1.

[23]Among all the groups in Table 1, college women aged 27 to 30 experienced the biggest drop in the ever-married rate.

Table 3: Effects of the College Expansion on the Probability of Marriage: Baseline Results

Dependent variable: Ever being married				
	(1)	(2)	(3)	(4)
Post-expansion cohorts	1985-88, 27-30 years old in 2015		1981-84, 31-34 years old in 2015	
Pre-expansion cohorts	1975-78, 27-30 years old in 2005		1975-78, 31-34 years old in 2009	
	College	Noncollege	College	Noncollege
A. Male				
ExpProxy∗Post	0.0273***	0.0059***	0.0124***	0.0008
	(0.0049)	(0.0020)	(0.0039)	(0.0014)
Observations	40196	187259	36486	181105
Marriage rate of				
Pre-expansion cohorts	0.644	0.793	0.883	0.882
Post-expansion cohorts	0.613	0.728	0.858	0.874
B. Female				
ExpProxy∗Post	0.0171***	0.0020	0.0053	-0.0009
	(0.0037)	(0.0034)	(0.0040)	(0.0012)
Observations	38477	182842	33894	176213
Marriage rate of				
Pre-expansion cohorts	0.787	0.916	0.919	0.961
Post-expansion cohorts	0.709	0.873	0.892	0.953
C. Female − Male				
	-0.0102*	-0.0039	-0.0072	-0.0017
	(0.0061)	(0.0039)	(0.00560)	(0.0018)

Notes: The marital outcome of pre-expansion cohorts is constructed using marriage history so that it is comparable to the post-expansion cohorts. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit represents one standard deviation across all prefectures. All regressions control for prefecture fixed effects and age fixed effects. Standard errors clustered at the prefecture level are in parentheses. There are 340 clusters. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The difference between the estimated effects for the early and late post-expansion cohorts may arise for two reasons. First, the magnitude of the expansion was much larger in later years than in the first few years (Figure 1). Therefore, the effects of the expansion are probably larger for the late post-expansion cohorts. Second, we observe the early post-expansion cohorts at older ages (31–34) than the late post-expansion cohorts (27–30). If the expansion reduced the average age of first marriage and did not indicate permanent changes in marriage probability, then the estimated effects should be smaller when we observe the post-expansion cohorts when they were older than when they were younger. This age effect may also (partially) explain the difference between the early and late post-expansion cohorts. However, because of a data limitation, we cannot distinguish between these two explanations: age and cohort are perfectly co-linear in the cross-sectional post-expansion data.[24]

We also observe gender differences in the estimated effects of the expansion, as shown in the last row of the table. Although most of the differences are imprecisely estimated, the point estimates suggest that the expansion had a larger effect on the probability of marriage for college men than women. In the table, we look at women and men from the same cohorts so they were exposed to the same degree of expansion. Because of the marriage age gap (women on average marry men older than themselves), however, women and men from the same cohorts are not in exactly the same marriage market. Some college women in the post-expansion cohorts may marry older men who were not exposed to the expansion or were exposed to a smaller expansion. The effects of the reduction in search frictions are arguably smaller for these women than men from the same cohorts because college enrollment of the women's potential spouses increased by a smaller magnitude. In Section 5.6, we further discuss this issue.

In Columns (2) and (4), we observe no effects of the expansion on the noncollege groups except for noncollege men from the late post-expansion cohorts. Our simulated comparative statics, however, predict that the marriage rate of noncollege men decreased in response to the expansion. The contrast between the theory and the empirical results implies that the model does not capture all key features of the noncollege marriage market. The expansion not only directly changed the distribution of college attainment but generated spillover effects on the distribution of below-college education types. In results not reported in this paper, for example, we find that the expansion increased the rate of high school graduation among those with less than a college education. To the extent that high schools might also serve as local marriage markets, the expansion likely also reduced search frictions in high school marriage markets. The mechanisms, however, are beyond the scope of this paper.

Overall, the results suggest an economically sizable impact of the expansion on college graduates' marriage probability. The positive estimated effects are in line with the simulation results of our model, in which the reduction in search frictions dominates the effects of only altering the relative distribution of education types. In Sections 5.2–5.6, we discuss a series of robustness checks for our main finding.

5.1.1 "Leftover Women"

In recent years, the rising phenomenon of educated women marrying later and at a lower rate, who are referred to as the "leftover women", has attracted public attention and raised policy concerns (Fincher, 2016; Magistad, 2013; To, 2015). The rapid rise of college-educated

[24] With soon-available 2020 census microdata, we will be able to test these two explanations. For example, we will be able to observe the late post-expansion cohorts (1985–88) when they were 31–34 years old and test whether the effect on their marriage probability is still larger than the effect for the early post-expansion cohorts.

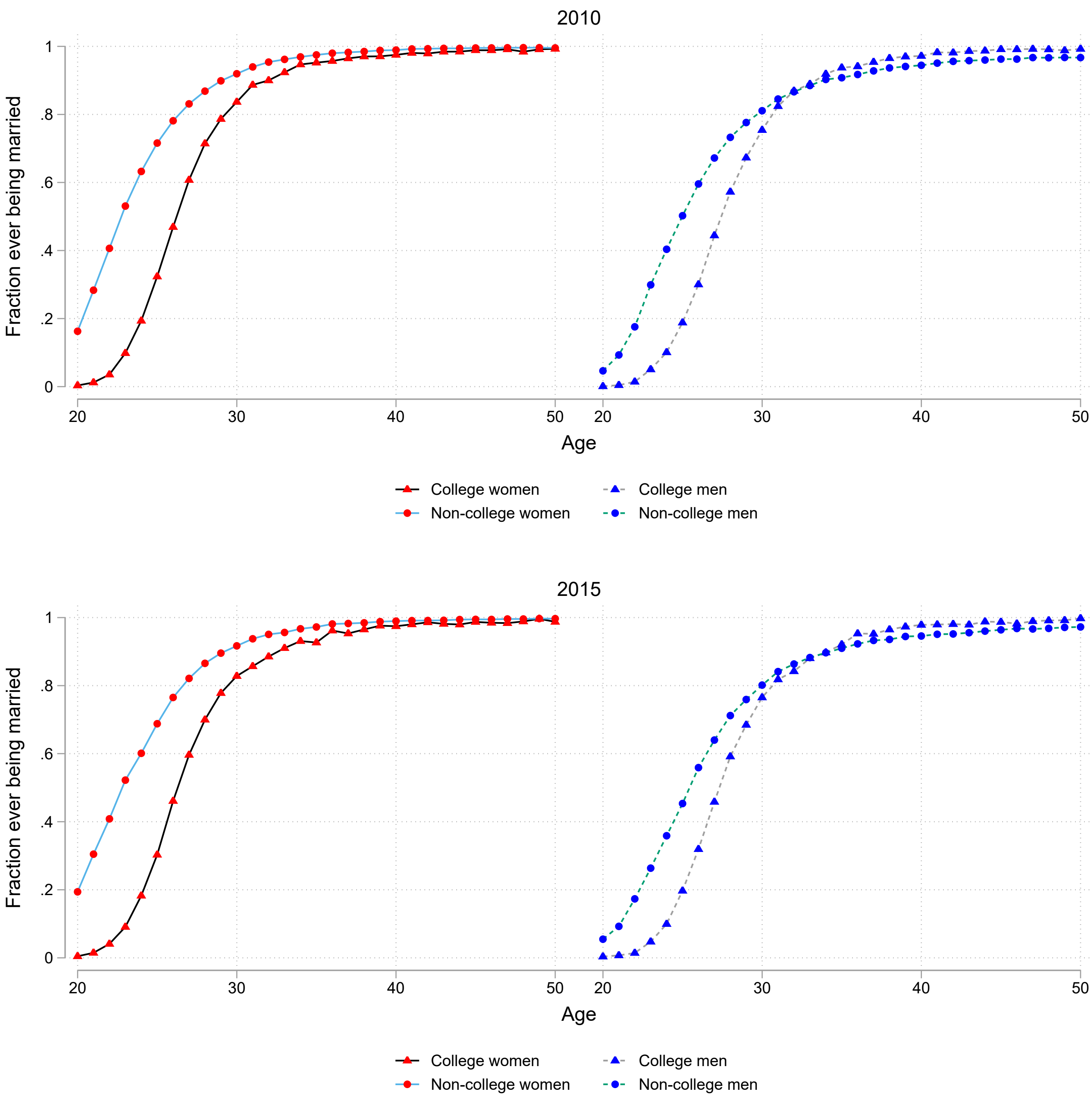
women in combination with traditional gender norms (Section 2.4.3), arguably, is contributing to this trend. The mean values presented in Table 3 indeed show that the marriage probability of college women is lower than that of noncollege women at the same age for all cohorts. Nevertheless, we show that expanding access to college education generates positive spillover effects on the marriage probability of college women. Our findings add to the field's understanding of the effects of college education on women's marriage outcomes and on gender inequality. Education policies that expand access to higher education, our results suggest, may actually raise the marriage prospects of highly educated women.

5.1.2 Effects on Permanent Marriage Rates and Early Marriage

The earliest cohort that was directly exposed to the expansion (born in 1981) was only 34 years old when observed in the 2015 mini-census data. Therefore, we do not observe the marriage outcomes of the post-expansion cohorts at an old-enough age to determine whether the effects indicate permanent changes in marriage outcomes or temporary changes (for example, earlier marriages). Figure 10 reports the age profile of marriage probabilities by gender and college education. As suggested by the cross-sectional age profiles, most people eventually get married. If everyone in the post-expansion cohorts in our sample follows this pattern, then our results probably indicate effects of the expansion on earlier marriage rather than a permanent increase in marriage rates. Nevertheless, the age profile mostly stabilizes after the mid-30s. Therefore, the positive effects on the early post-expansion cohorts, who were 31–34 years old as of 2015, might at least partially reflect an increase in permanent marriage rates.

The profiles depicted in Figure 10 are based on cross-sectional data and do not tell us about the post-expansion cohorts when they grew older. After microdata from the 2020 Chinese census become accessible to us, we will be able to more confidently determine whether the expansion has affected the permanent marriage rates of the post-expansion cohorts.

Figure 10: Age Profiles of Ever-Married Rate



Notes: Data come from random samples of the 2010 census and 2015 mini-census of China. Each plot shows the fraction of people who were ever married by age in the cross-sectional data.

5.2 Falsification Test and Sensitivity to Pre-expansion Cohorts

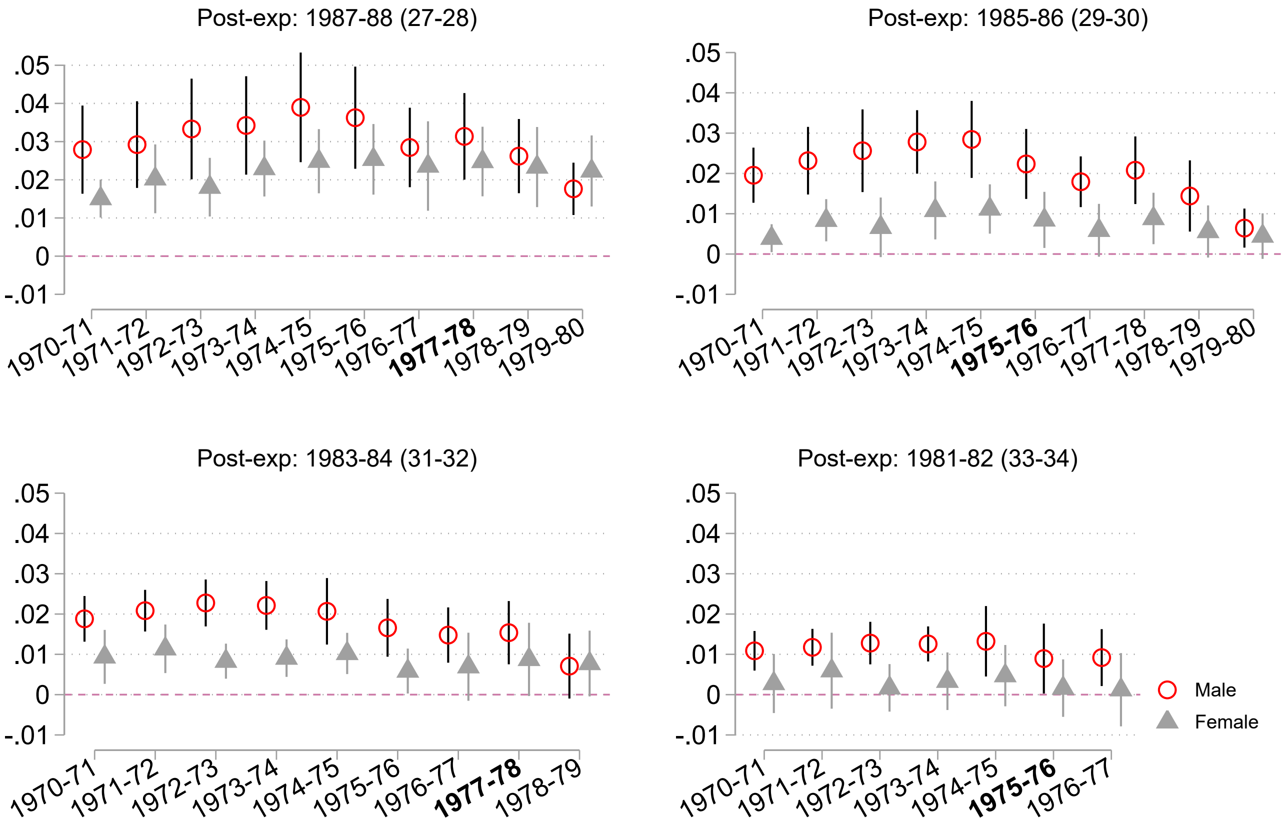
We chose the 1975–78 birth cohorts as our control cohorts for the baseline analysis. The choice, as discussed in Section 4.1, reflects the tension between avoiding potential spillovers in the marriage market and ensuring that the pre-expansion and post-expansion cohorts are comparable. In this section, we investigate whether our findings are sensitive to the choice of pre-expansion cohorts; this exercise also provides suggestive evidence on whether the pre-trends are parallel. To test the sensitivity of estimated results at a more granular level and probe potential spillover effects in the marriage market, we further divide the post-expansion cohorts into four groups: 1987–88 (27–28 years old in 2015), 1985–86 (29–30 years old in 2015), 1983–84 (31–32 years old in 2015), and 1981–82 (33–34 years old in 2015). This also enables us to better track the different treatment effects across post-expansion cohorts.

Let us take the 1987–88 post-expansion cohorts as an example. They were observed at the ages of 27 and 28 in the 2015 census. For the sensitivity test, we use different pre-expansion cohorts and consider their marital status at the same ages (for example, the 1977–78 cohorts in 2005). We then apply our DID model (Equation 5) to these pre-expansion and post-expansion cohorts. Varying the choice of the pre-expansion cohorts, we obtain a series of estimates for the effects of the expansion on the marriage probability of the 1987–88 cohorts in 2015.

The results of the sensitivity test are displayed in Figure 11. We report results for college women and men. Each dot represents the estimated treatment effect for one DID regression. We plot the coefficients and 95% confidence intervals by holding the post-expansion cohorts fixed and varying the pre-expansion cohorts. Two patterns emerge. First, the coefficients are mostly stable. Therefore, our baseline findings are not sensitive to the choice of pre-expansion cohorts.

Second, when estimating the model with pre-expansion cohorts that reach college age just before the college expansion (born in 1979 or 1980), we observe much smaller estimated effects for college men compared to estimated effects using the pre-expansion cohorts as in our baseline analysis (shown in bold on the x-axis in Figure 11). We do not observe a similar pattern for college women. This is consistent with the possibility of spillover effects in the marriage market. Men in China usually marry younger women (the average husband-wife age gap is about two years). Therefore, college men born during 1979–80 were partially treated since their potential spouses were directly exposed to the expansion, while college women born during 1979–80 were much less likely to be treated. This again justifies our choice of the pre-expansion cohorts in our baseline analysis. We further discuss this issue below in a falsification test based on the pre-expansion data.

Figure 11: Sensitivity Test for Pre-expansion Cohorts

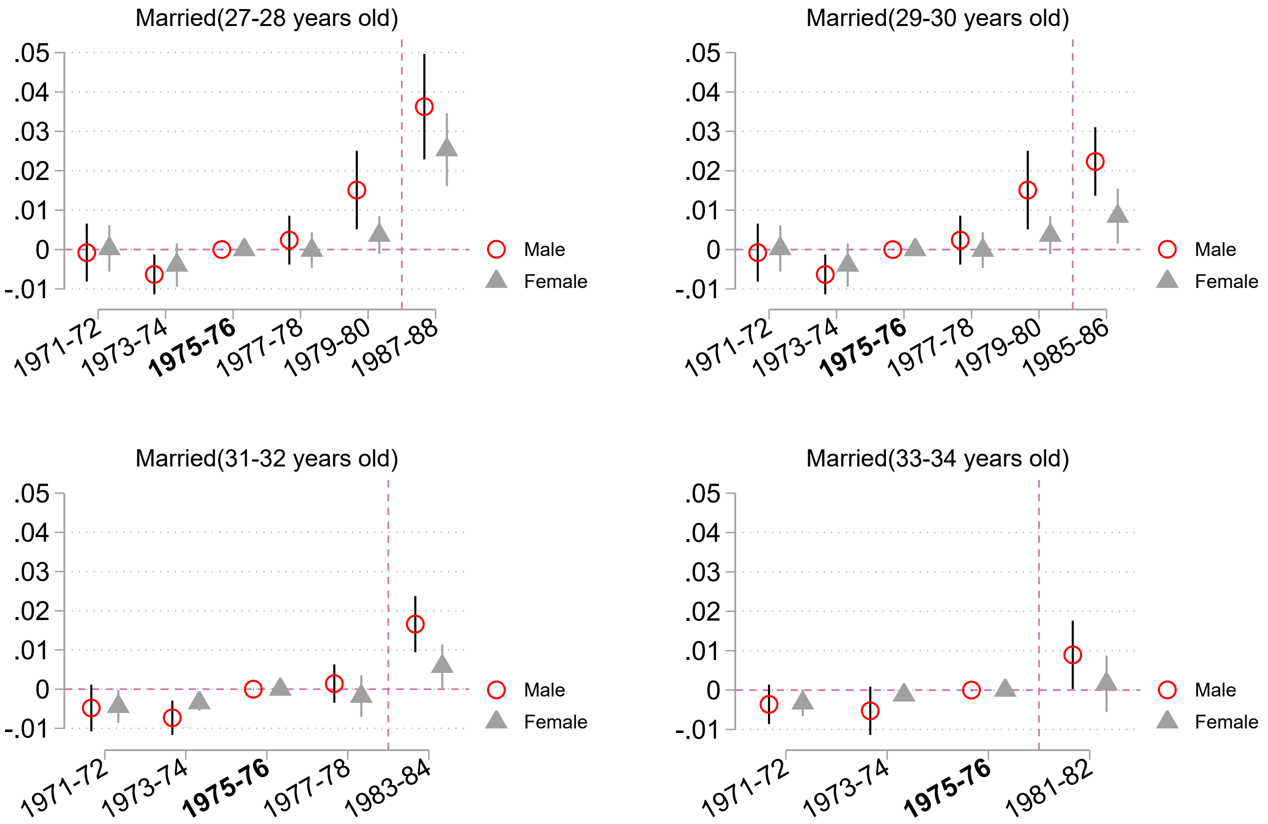


Notes: Each panel plots the difference-in-differences coefficients estimated by combining the post-expansion cohorts in its title and different pre-expansion cohorts on its x-axis. The specification is the same as Equation 5. Point estimates for college women and men and 95% confidence intervals, estimated by clustering at the prefecture level, are plotted. Pre-expansion cohorts in **bold** on the x-axis are the cohorts used for our baseline results.

We also conduct a falsification test considering only the pre-expansion cohorts. In the test, we assume that the 1975–76 cohorts were not exposed to the expansion while other cohorts were, and therefore we take the 1975–76 cohorts as the benchmark pre-expansion cohorts. The placebo treatment effects are then estimated by using different placebo "post-expansion" cohorts and the 1975–76 pre-expansion cohorts.[25] If our observed effects on marital outcomes are indeed caused by the expansion, then the treatment should have no effects in these placebo tests. These tests can help us formally detect whether the parallel-trends assumption is violated in the pre-expansion cohorts.

[25]We choose 1976 as the cutoff because it is the midpoint of our baseline pre-expansion cohorts (1975–78). We can therefore formally conduct the falsification test using the baseline pre-expansion cohorts.

Figure 12: Falsification Test



Notes: Each panel plots the difference-in-differences coefficients. The pre-expansion cohorts are fixed at 1975–76 (in bold). The x-axis displays different falsification "post-expansion" cohorts. The specification is the same as Equation 5. Point estimates for college women and men and 95% confidence intervals, estimated by clustering at the prefecture level, are plotted. Right of the vertical line are estimated coefficients using the actual post-expansion cohorts (cohorts born later than 1980). Because we can only obtain marital-history information from the 2010 census, we cannot use the 1979–80 cohorts when the outcome is ever being married at the ages of 31–32 *or* the 1977–80 cohorts when the outcome is ever being married at the ages of 33–34.

The results are reported in Figure 12. We fix the pre-expansion cohorts at 1975–76 (in bold) and estimate the DID specification using falsification "post-expansion" cohorts after or before them. The estimates are displayed in a similar way as in Figure 11). Consistently with the suggestive patterns from the previous sensitivity test, we largely find no effects of the placebo expansion. One exception is the 1979–80 cohorts. The placebo expansion seems to have positive effects on college men from these cohorts. Considering the marriage age gap, the potential spouses of these men were directly impacted by the (actual) college expansion. This can raise the marriage prospects of these men because of (1) a reduction in search frictions and (2) a relative increase in the number of college women in the marriage market.[26] We only observe positive spillover effects on the marriage probability of college

[26]The college expansion did not directly increase the college ratio of men in the 1979–80 cohorts. However,

men from these pre-expansion cohorts. This pattern further supports the view that our empirical results are driven by the college expansion.

5.3 Continuous-Treatment DID

Recent studies of DID designs have pointed out potential issues with continuous treatment variables. First, the traditional OLS estimator of the two-way fixed-effects model may load negative weights on treatment effects of some units with a multivalued or continuous treatment variable. If there are heterogeneous treatment effects, the negative weights lead to bias in the estimated causal effects (de Chaisemartin et al., 2022). To address this concern, we adopt a heterogeneity-robust DID estimator (de Chaisemartin et al., 2022, 2019). The results are reported in Table 4.

Table 4: Fuzzy-DID Estimator (de Chaisemartin et al., 2022)

Dependent variable: Ever being married				
	(1)	(2)	(3)	(4)
Post-expansion cohorts	1985-88, 27-30 years old in 2015		1981-84, 31-34 years old in 2015	
Pre-expansion cohorts	1975-78, 27-30 years old in 2005		1975-78, 31-34 years old in 2009	
	College	Non-college	College	Non-college
A. Male				
ExpProxy∗Post	0.0532***	0.0066	0.0054	-0.0011
	(0.0133)	(0.0109)	(0.0090)	(0.0067)
	[0.0273]	[0.0059]	[0.0124]	[0.0008]
Observations	40196	187259	36486	181105
B. Female				
ExpProxy∗Post	0.0373***	0.0010	0.0080	-0.0022
	(0.0133)	(0.0080)	(0.0050)	(0.0038)
	[0.0171]	[0.0020]	[0.0053]	[-0.0009]
Observations	38477	182842	33894	176213

Notes: This table reports the estimates using the fuzzy-DID estimator for all college graduates by gender. The marital outcome of pre-expansion cohorts is constructed using marriage history so that it is comparable to the post-expansion cohorts. $ExpProxy$ is the proxy for college expansion, which is standardized such that one unit represents one standard deviation across all prefectures. All regressions control for prefecture fixed effects and age fixed effects. The bottom quintile is used as the benchmark "untreated" group for the estimator. Baseline estimates are in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The alternative estimator delivers largely qualitatively robust findings. For college men and women from the late post-expansion cohorts, the new point estimates are larger than our baseline estimates. For college men from the early post-expansion cohorts (1981–84), however, the new estimate is still positive, but it is smaller and no longer statistically significant.

it raised the college ratio of women who are younger yet potentially in the same marriage market. This tended to raise the marriage prospects of college men by reducing the sex ratio of college-educated people.

The second problem is selection bias associated with different treatment intensities (Callaway et al., 2021). If we interpret the DID estimator as the weighted average of treatment effects when moving between contiguous treatment values, there might be a selection bias unless all the treated units would have shared the same path of outcomes if they had received treatment of the same value. In our case, there could be a positive selection bias if the treatment effect in a high-treatment prefecture would have been higher than that in a low-treatment prefecture if the former had also received the low treatment. Because we can only observe one realized treatment value for each prefecture, there is no easy fix for this potential problem. Such a selection bias is likely to appear, for example, if the treated units can select their treatment values and if units with larger treatment effects tend to select a higher treatment value. Nevertheless, our treatment proxy is based on higher education endowment well before the college expansion. Therefore, prefectures in our sample do not actually select based on treatment effects. This setup alleviates our concern about the unique selection bias associated with continuous treatment variables.

5.4 Measurement Errors in College Types

Parallel to the regular college system in China is a postsecondary credential system (also referred to as adult higher education) that mainly serves adults older than typical college ages. The admission bar to the latter system is very low, and students are not required to regularly study on-site (Kai-Ming et al., 1999; Wang, 2011). The college expansion, as mentioned, has been concentrated on the regular college system. Unfortunately, we cannot distinguish between regular college degrees and other postsecondary credentials in the census data. To address the issue of potential measurement errors in college types, we use CFPS data to correct for the potential errors. The CFPS data set, though much smaller compared to the census, contains detailed information about respondents' college-degree types. We predict whether an individual possesses a regular college degree using basic demographic characteristics in the CFPS data with a logit model.[27] The estimated model is then applied to the census data, and we classify individuals whose predicted probability of having a regular college degree is above 0.5 as regular college graduates. Details about the prediction model are provided in Appendix E.

The estimated effects on the predicted regular college graduates are comparable to our baseline estimates in Columns (1) and (3) of Table 3. The estimates are quantitatively similar to the baseline estimates in Table 3 and also exhibit similar levels of statistical significance. The results show that our baseline findings are robust to considering the potential difference between regular college graduates and other postsecondary-credential holders.

[27] The predictors include interactions between (a) birth cohort and (b) the province of residence, gender, ethnicity, and urban-rural residency status.

Table 5: Results for Predicted Regular College Graduates

Dependent variable: Ever being married				
	(1)	(2)	(3)	(4)
Post-expansion cohorts	1985-88, 27-30 years old in 2015		1981-84, 31-34 years old in 2015	
Pre-expansion cohorts	1975-78, 27-30 years old in 2005		1975-78, 31-34 years old in 2009	
	Male	Female	Male	Female
ExpProxy∗Post	0.0260***	0.0174***	0.0130***	0.00658
	(0.00538)	(0.00369)	(0.00366)	(0.00494)
	[0.0273]	[0.0171]	[0.0124]	[0.0053]
Observations	29039	21703	26757	18628
Marriage rate of				
Pre-expansion cohorts	0.640	0.747	0.881	0.909
Post-expansion cohorts	0.629	0.720	0.859	0.890

Notes: All models are estimated using the sample of predicted regular college graduates. Baseline estimates are in brackets. The marital outcome of pre-expansion cohorts is constructed using marriage history so that it is comparable to the post-expansion cohorts. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit represents one standard deviation across all prefectures. All regressions control for prefecture and age fixed effects. Standard errors clustered at the prefecture level are in parentheses. There are 340 clusters. * $p < 0.1$, ** $p < 0.05$, $^{***\cdot}$ $p < 0.01$. Baseline estimates are in brackets.

5.5 Local Marriage Markets and Migration

We define each prefecture (the subprovincial geo-unit) as a local marriage market. Whether this definition captures actual marriage markets is crucial for interpreting our empirical results. One potential concern is that the marriage market is larger than a single prefecture, leading to mis-specification in the empirical model. We discuss this issue from two aspects.

First, there is good reason to believe that the marriage markets are largely local. The typical prefecture in China is a very large geo-unit. The average population size of a prefecture was four million in 2015.[28] Given the large size, it is probably reasonable to believe that most marriages are formed within prefectures. In a recent work, Chen et al. (2022) document that commuting-based metropolitan areas in China usually do not cross prefectural boundaries. The authors use this information to delineate China's local labor markets, which also likely reflects how localized China's marriage markets are.

Second, we relax the assumption that local marriage markets are confined to prefectures. Instead, we consider the province as the unit of analysis.[29] The treatment proxy is defined at the province level. We re-estimate the baseline DID specification (Equation 5) with the province-level treatment variable and province fixed effects. The results are reported in Table 6. As shown in Columns (1) and (3), we still find that the expansion boosted the marriage probabilities of college women and men. For example, Column (1) shows that

[28]The average population size of a US state, for comparison, was 6.4 million in 2015.

[29]There are 31 provincial-level units in mainland China.

increasing the treatment proxy by one SD across provinces leads to a 4 percentage-point increase in the marriage probability of college men and a 2.4 percentage-point increase for college women in the late post-expansion cohorts. The results are qualitatively consistent with our baseline results.[30] Therefore, our conclusions are robust to taking a province as a local marriage market.

Table 6: College Expansion and Marriage: Province Level Marriage Markets

Dependent variable: Ever being married				
	(1)	(2)	(3)	(4)
Post-expansion cohorts	1985-88, 27-30 years old in 2015		1981-84, 31-34 years old in 2015	
Pre-expansion cohorts	1975-78, 27-30 years old in 2005		1975-78, 31-34 years old in 2009	
	College	Non-college	College	Non-college
A. Male				
ExpProxy∗Post	0.0398***	0.0050	0.0185***	-0.0020
	(0.0047)	(0.0046)	(0.0063)	(0.0030)
Observations	40198	187259	36489	181105
Marriage rate of				
Pre-expansion cohorts	0.644	0.793	0.883	0.882
Post-expansion cohorts	0.613	0.728	0.858	0.874
B. Female				
ExpProxy∗Post	0.0239***	0.0038	0.0104*	0.0005
	(0.0049)	(0.0085)	(0.0056)	(0.0012)
Observations	38478	182842	33897	176213
Marriage rate of				
Pre-expansion cohorts	0.787	0.916	0.919	0.961
Post-expansion cohorts	0.709	0.873	0.892	0.953

Notes: The marital outcome of control cohorts is constructed using marriage history so that it is comparable to the treatment cohorts. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit is one standard deviation of the treatment proxy across provinces. All regressions control for province fixed effects and age fixed effects. Standard errors clustered at the province level are in parentheses. There are 31 clusters. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A related concern is migration between local marriage markets. For two reasons, this turns out to be not as consequential as it might appear. First, while college students in one prefecture do not solely come from that prefecture, this fact does not change the interpretation of our results. Our cross-sectional variation comes from the fact that college enrollment

[30] The results are also quantitatively comparable when defining the marriage market at the prefecture versus province level. The SD of the prefecture-level treatment proxy is 1.57 times that of the province-level proxy. If we rescale the estimates in Table 6 using this factor, for example, the treatment effect on college men in Column (1) is 0.025, while that for college men in Column (1) of Table 3 is 0.027.

increased much more in places with high preexisting higher education resources, regardless of where the additional enrollment came from.

Second, after they graduate, college students do not necessarily stay in the prefecture where they go to college. This might create potential measurement errors if our treatment proxy is not an accurate measure of the stock of college graduates in the local marriage market. But the mobility of college graduates after graduation seems low. For example, based on the 2015 mini-census data, the probability of staying in the same prefecture as five years ago is 90% for college women and 88% for college men aged 27–35 years old.

If the migration decision after graduation is uncorrelated with either the treatment or local marriage-market conditions, that would at most attenuate our estimates by introducing additional measurement errors. The concern would be that college graduates who have a stronger preference for marriage are more likely to move to cities that have more college graduates and can provide more abundant opportunities for marriage. But, given the low rate of migration after college graduation, this potential confounding channel is unlikely to drive our main results. The robustness of the results using provincial-level local marriage markets also alleviates this concern because cross-province migration is less common than cross-prefecture migration.[31]

5.6 Marriage Age Gap

We investigate women and men from the same cohorts in the baseline results so that the analyzed cohorts experienced the same intensity of treatment and were observed at the same ages. However, women on average marry men older than themselves. The average husband-wife age gap in China is about two years. Women and men from the same cohorts, therefore, might not be in exactly the same marriage market. The dynamics are further complicated by the fact that the intensity of the expansion increased over time. When we look at men and women from the same cohorts, the women are actually in a marriage market with older men, and the effects of the expansion on college attainment are smaller for their potential spouses.

[31] The probability of staying in the same province as five years ago is 95% for college women and 93% for college men aged 27–35 years old. Separately, our data set provides information on (1) whether the current prefecture of residence is the same as five years ago and (2) the province of residence five years ago. Based on this, we conduct two additional tests: (1) we re-estimate the baseline model with only individuals that resided in the same prefecture as five years ago; (2) we re-estimate the model that takes provinces as local marriage markets but take the province of residence five years ago as the current province of residence. The results are quantitatively very similar to (and qualitatively consistent with) our original results (Table 3 for the first test and Table 6 for the second test). These tests further suggest that selection into migration is not likely to bias our results.

Table 7: Compare Treatment Effects on Different Genders Allowing for Two-year Age Gap

Dependent variable: Ever being married				
	(1)	(2)	(3)	(4)
	College	Non-college	College	Non-college
A. Male				
Post-expansion cohorts	1985–88, 27-30 years old in 2015		1981–84, 31-34 years old in 2015	
Pre-expansion cohorts	1975–78, 27-30 years old in 2005		1975–78, 31-34 years old in 2009	
ExpProxy*Post	0.0273***	0.0059***	0.0124***	0.0008
	(0.0049)	(0.0020)	(0.0039)	(0.0014)
Observations	40196	187259	36486	181105
Marriage rate of				
Pre-expansion cohorts	0.644	0.793	0.883	0.882
Post-expansion cohorts	0.613	0.728	0.858	0.874
B. Female				
Post-expansion cohorts	1987–90, 25-28 years old in 2015		1983–86, 29–32 years old in 2015	
Pre-expansion cohorts	1977–80, 25-28 years old in 2005		1977–80, 29–32 years old in 2009	
ExpProxy*Post	0.0285***	0.0106**	0.0066**	-0.0011
	(0.00596)	(0.00489)	(0.0033)	(0.0026)
Observations	43630	178921	38302	169558
Marriage rate of				
Pre-expansion cohorts	0.580	0.841	0.864	0.935
Post-expansion cohorts	0.504	0.788	0.825	0.924
C. Female − Male				
	0.00125	0.00467	-0.00589	-0.00186
	(0.00774)	(0.00527)	(0.00511)	(0.00291)

Notes: The marital outcome of control cohorts is constructed using marriage history so that it is comparable to the treatment cohorts. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit represents one standard deviation across all prefectures. All regressions control for prefecture and age fixed effects. Standard errors clustered at the prefecture level are in parentheses. There are 340 clusters. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

To test the effects of the expansion for women and men in the same marriage market, we re-estimate the econometric model by comparing men from the same cohorts as in the baseline model but also comparing women two years younger than the corresponding men. The results are reported in Table 7.[32] Column (1) shows that when we compare college women that are two years younger than college men in the late post-expansion cohorts, the

[32] These results need to be interpreted with caution because age is another important trait to match on in the marriage market. Therefore, the marriage age gap is endogenous to change in marriage-market conditions. In Section 6.4, we show that the college expansion indeed reduced the marriage age gap. The effects, nevertheless, are small relative to the preexisting marriage age gap. Therefore, we take the two-year gap as approximately given in this exercise.

expansion has quantitatively similar effects on their marriage probabilities. If we look at college women who are two years younger than college men from the early post-expansion cohorts, as in Column (3), the impact of the college expansion is statistically significant. The gender difference in treatment effects, however, is still sizable (and nonsignificant). Taken together, these findings suggest that at least part of the gender differences observed in our baseline findings (Table 3) are explained by the positive average husband-wife age gap.

5.7 Potential for Selection and Composition Change

We focus on the equilibrium effects of the expansion on college graduates. One concern, however, is that the expansion may have shifted the composition of college students. Such an endogenous composition change could bias our results if students that enrolled in college as a result of the expansion are very different from other college students in terms of their marriage decisions. For example, if those who went to college because of the expansion were more likely to get married than other college students with or without the expansion, then we may observe a spurious positive impact of the expansion on the marriage probability of an average college graduate.[33]

In our model, the only trait relevant for marriage matching is education type. Marital preferences do not systematically differ for individuals with a given education type. The model, therefore, does not accommodate this potential selection story. From the model perspective, we focus on the expansion's average equilibrium effects that shifted the overall education distribution. Therefore, as long as the marginal students who enrolled as a result of the expansion were not systematically different from other college graduates in terms of marital preferences or other traits in the marriage market, our empirical results will still be largely consistent with the effects that we predict in the theoretical model.

Still, it is helpful to empirically test whether the endogenous composition change threatens our findings. One critical feature of China's college system is that a single score from the college entrance examination (gaokao) is the sole determinant of college admission.[34] Only students whose scores were above a certain threshold had the opportunity to be admitted each year. The college expansion increased the quota for college admission and therefore drew more students from the relatively lower score distribution. If our main findings are driven by a composition change, then we expect that students from the lower score distribution are also more likely to get married.

We use a unique data set, the China Household Income Project (CHIP), to conduct a

[33]It is ex ante unclear in which direction the potential bias will be. For example, our results may reflect a negative bias if the students that went to college because of the expansion came from relatively low-socioeconomic-status families and are less likely to marry. On the other hand, the bias could be positive if these students had less human capital and therefore lower opportunity costs in the labor market from marrying early.

[34]The gaokao includes multiple subjects. It takes place only once a year in each summer.

test on the gaokao score and marriage probability. The CHIP is a nationally representative household survey that covers income and expenditure information of Chinese households.[35] The 2013 wave of CHIP contains information on respondents' gaokao score. Using the sample of college graduates, we estimate the correlation between individuals' score and marriage probability with the following specification:

$$Married_{iars} = \rho Score_{iqrs} + \xi_a + \phi_r + \varphi_s + \epsilon_{iars} \tag{6}$$

The dependent variable is a dummy for being married for individual i of age a. r indicates the region (province) where i took the gaokao exam, and s stands for the year of the exam. $Score$ is i's total exam score.[36] We also control for age, region, and gaokao-year fixed effects when applicable. The sample includes cohorts born between 1975 and 1986, which are roughly the same as the cohorts in our main analysis. These cohorts were at least 27 years old in 2013, so their marriage outcomes are largely comparable to those in our baseline analysis.

The results are shown in Table 8. Columns (1) and (2) show that there is no significant association between the gaokao score and marriage probability. The estimates are very imprecise as a result of the small sample size, but the magnitudes are small. For example, Column (1) suggests that increasing the gaokao score by 100 is only associated with a 0.8 percentage-point higher marriage probability for college men and 0.3 percentage-point higher probability for college women.[37]

We further look into cohorts that were and were not exposed to the expansion. If our results are driven by selection bias, then we should see a more negative association between the score and marriage probability for the post-expansion cohorts. As shown in Columns (3) and (4), we actually observe a negative association for the pre-expansion cohorts and a positive association for the post-expansion cohorts. Based on the point estimates, it seems that the composition-change bias, if anything, goes in favor of our main findings and suggests that our baseline estimates provides a lower-bound of the expansion's positive effect on marriage probabilities. None of the estimates are statistically significant, so they should be interpreted with caution. Nevertheless, these patterns suggest that the composition change is not likely to be driving our results.

[35]The surveys were organized in multiple years by the China Institute for Income Distribution at Beijing Normal University and conducted by the National Bureau of Statistics of China.

[36]The score ranges from 0 to 750, and we divide the score by 100 in the regressions.

[37]One hundred points represent a giant increase in the test score and can catapult a student to a much higher tier of universities.

Table 8: Correlation between College-Entrance-Exam Score and Marriage Probability

Dependent variable: Ever being married				
	(1)	(2)	(3)	(4)
A. Male				
Cohorts	1975-86		Pre-expansion	Post-expansion
$\frac{Score}{100}$	0.0082	0.0032	-0.0175	0.0155
	(0.0143)	(0.0151)	(0.0150)	(0.0257)
Observations	494	468	219	249
Age FE	Yes	Yes	Yes	Yes
Exam Year FE		Yes	Yes	Yes
Province FE		Yes	Yes	Yes
B. Female				
Cohorts	1975-86		Pre-expansion	Post-expansion
$\frac{Score}{100}$	0.0027	-0.0019	-0.0105	0.0098
	(0.0189)	(0.0165)	(0.0185)	(0.0252)
Observations	490	467	193	274
Age FE	Yes	Yes	Yes	Yes
Exam Year FE		Yes	Yes	Yes
Province FE		Yes	Yes	Yes

Notes: The sample consists of college graduates in the 2013 wave of China Household Income Project. The maximum of the original exam score is 750. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.8 Other Robustness Tests

In this section, we report several additional tests that further confirm the robustness of our findings. Our results are robust to looking at four-year or three-year institutions, using an alternative proxy for college-expansion treatment, and controlling for prefecture-level covariates.

5.8.1 Four-Year versus Three-Year Institutions

There are two types of college in China: four-year universities (benke) and three-year colleges (zhuanke). The four-year universities have higher quality and higher admission standards, and both types of college degree bring very high wage returns compared to lower education levels (Zhong, 2011). Both four-year and three-year institutions experienced the expansion. In our baseline analysis, we focus on the whole higher education system and therefore do not distinguish between these two levels of higher education. It might be informative to explore whether the effects of the expansion differ between them. Table F6 reports the effects of the expansion on four-year and three-year institutions' graduates, respectively. While the expansion increased the marriage probabilities of both groups, the effects seem to be larger for four-year university graduates.

5.8.2 Using 1990 Data to Construct the Expansion Proxy

In the baseline analysis, we use the historical endowment of higher education resources as of 1982 as a proxy for treatment intensity. Because variation in this measure was mostly predetermined during the pre-expansion era, this proxy is unlikely to be confounded by the growth of higher education that is associated with subsequent economic development. This benefit, however, comes with a potential cost: the proxy might be less strongly correlated with the actual intensity of the expansion than a proxy based on variation closer to the expansion period. As a robustness check, we construct an analogous treatment proxy using data from the 1990 Chinese census and redo the analysis. The results, reported in Table F7, are qualitatively similar to our baseline findings.

5.8.3 Additional Covariates

Our identification assumption requires that the marriage outcomes follow parallel trends in the absence of the expansion. The parallel-trends assumption could be violated if our treatment proxy is associated with other key factors that influence the marriage market and these factors cause different secular trends. We have shown, in Appendix D, that our treatment variable is not associated with GDP-per-capita growth or local sex ratio. We further control for baseline prefecture characteristics interacted with a time dummy. In the first exercise, we control for GDP per capita and the sex ratio of each prefecture in 2000.[38] The results are reported in Table F8. In the second exercise, we include province-by-time fixed effects, which control for regional differences in economic growth and marriage-market conditions at the province level and exploit only within-province variation to identify the effects of the expansion on local marriage markets. The results are reported in Table F9. In both tests, our main findings are robust, both qualitatively and quantitatively.

6 How the Expansion Affected Matching Patterns

6.1 Effects on Marriage-Matching Patterns

In this section, we further assess the implications of the college expansion for marital sorting. It is perhaps unsurprising that there are more college-college marriages as a result of the in-

[38] We calculate the sex ratio of each prefecture in the age range of 10–30 years old using the 2000 China census data from IPUMS. Data on GDP per capita come from the City Statistics Year Book of China.

creasing supply of college graduates in the marriage market. It is unclear, however, whether this reflects a change in underlying marital sorting—that is, whether educated women and men are more likely to marry each other conditional on the marginal distribution of education types. Research has shown that increasing positive assortative mating by education could amplify household income inequality (Greenwood et al., 2014; Eika et al., 2019). It is therefore important to understand how the underlying sorting patterns in China have changed following the expansion.

We first directly estimate the impacts of the expansion on matching patterns by focusing on the proportions of different types of match. We focus on the simple dichotomy between college and noncollege graduates and estimate the DID model in Equation 7 with married couples:

$$z_{ijpt} = \beta_3 + \beta_4 ExpProxy_p * Post_t + \beta_5 * Post_t + \lambda_p + \varepsilon_{ijpt} \tag{7}$$

The dependent variable is a series of dummies for different match types between wife i and husband j in local marriage market p at time t. These dummies include whether the marriage consists of (1) a college wife and a college husband, (2) a college wife and a noncollege husband, (3) a noncollege wife and a college husband, and (4) a noncollege wife and a noncollege husband. The parameter of interest is the coefficient on the interaction between the expansion proxy and the dummy for post-expansion cohorts ($Post_t$).[39] We also include local-marriage-market fixed effects λ_p.[40]

The results, reported in Table 9, confirm that the college expansion has led to more college-college marriages.[41] For example, a one-SD increase in the treatment proxy caused a 2.1 percentage-point increase in the probability of college-college marriages for the early post-expansion cohorts. Comparing the national averages of different marriage types in the pre-expansion and post-expansion cohorts also leads to similar conclusions. These patterns, however, could partially be mechanical results of more college women and men in the marriage market. In order to distinguish the underlying changes in the matching structure from the mechanical effects, we have to adjust for the changes in the marginal distributions. This is the goal of the next subsection.

[39] We adjust the age of couples in the pre-expansion cohorts to make them comparable to the post-expansion cohorts. For example, when we choose the 1981–84 cohorts (who were 31–34 years old when observed in 2015) as the post-expansion cohorts, couples in the pre-expansion cohorts (1975–78) only include those who got married by 2009 (when they were at the same age). The estimates are similar and do not qualitatively alter our conclusion if we do not make such age adjustment.

[40] Controlling for age fixed effects of the husband and wife does not alter the estimates by much and does not affect our qualitative conclusion.

[41] By construction, the DID coefficients of the four columns in the same row add up to zero. In Panel A, the coefficients in Column (1) (Column (2)) and Column (4) (Column (3)) appear exactly opposite due to rounding errors.

Table 9: Effects of the College Expansion on Matching Patterns

Dependent variable (Wife-Husband)	(1) C-C	(2) NC-C	(3) C-NC	(4) NC-NC
A. Early Post-expansion Cohorts				
Post-expansion cohorts	1981–84, 31–34 years old in 2015			
Pre-expansion cohorts	1975–78, 31–34 years old in 2009			
ExpProxy∗Post	0.0210***	0.00165	-0.00165**	-0.0210***
	(0.00537)	(0.00101)	(0.000826)	(0.00609)
Observations	224384	224384	224384	224384
Dep. var. mean of				
Pre-expansion cohorts	0.097	0.043	0.023	0.836
Post-expansion cohorts	0.185	0.048	0.036	0.731
B. Late Post-expansion Cohorts				
Post-expansion cohorts	1985–88, 27–30 years old in 2015			
Pre-expansion cohorts	1975–78, 27–30 years old in 2005			
ExpProxy∗Post	0.0204***	0.00232***	-0.00262***	-0.0201***
	(0.00421)	(0.000826)	(0.000671)	(0.00441)
Observations	202855	202855	202855	202855
Dep. var. mean of				
Pre-expansion cohorts	0.083	0.040	0.020	0.856
Post-expansion cohorts	0.172	0.050	0.039	0.738

Notes: Each observation is a married couple. Dependent variables are dummies for corresponding matching type. C: college. NC: non-college. For example, C-NC refers to a dummy for the wife having a college degree while the husband does not. The marital outcome of pre-expansion cohorts is constructed using marriage history so that it is comparable to the post-expansion cohorts. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit is equivalent to one standard deviation across all prefectures. All regressions control for prefecture fixed effects. Standard errors clustered at the prefecture level are in parentheses. There are 340 clusters. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6.2 Assortative Mating

In order to address the mechanical effects of a larger college-educated population, we construct a new measure for assortative mating: the difference between the actual probability of college-college matching and the probability of college-college matching under the counterfactual of random matching. This measure, which we refer to as the absolute-difference measure, accounts for the fact that there would be more college-college marriages following the college expansion even under random matching. Formally, for the realized matching outcomes in a marriage market (as characterized by a contingency table, Table 10), the index is defined as follows:

Table 10: Contingency Table

		Education of Husband	
		C	NC
Education of Wife	C	k_1	k_2
	NC	k_3	k_4

Notes: The number in each cell k_j is the number of marriages of the corresponding type. $K = k_1+k_2+k_3+k_4$.

$$
\begin{aligned}
AbsDiff &= \frac{k_1}{K} - \frac{k_1+k_2}{K} * \frac{k_1+k_3}{K} \\
&= \frac{k_1k_4 - k_2k_3}{K^2}
\end{aligned}
$$

In the hypothetical situation in which marriage matching by education is totally random, this index will always be zero. An increase in this index implies that college-college marriages are now more likely relative to the benchmark of random matching. An increase also implies a potential increase in income inequality because highly educated individuals are more likely to sort into matches with other highly educated individuals.

To estimate the causal effects of the expansion on assortative mating, we again employ a DID design. We divide the national sample into high- and low-expansion regions based on the value of the expansion proxy. For the pre-expansion and post-expansion cohorts, we estimate the matching indexes in the high- and low-expansion regions, respectively. The estimated assortative-mating indexes are then used for constructing the DID estimate. We also adjust the pre-expansion cohorts so that their age range when their marriage outcomes were observed is comparable to the age range of the post-expansion cohorts, as in Section 6.1.[42]

Table 11 reports estimates for the effects of the expansion on assortative mating by college education. The expansion increased the level of assortative mating. If we look at the late post-expansion cohorts, the expansion increased the probability of college-college marriages relative to random matching by 4.1 percentage points in the high-expansion regions compared to the low-expansion regions. The magnitude of this effect seems large: it is half of the pre-expansion assortative-mating index in the high-expansion regions and more than 80% of the pre-expansion index in the low-expansion regions. We find quantitatively similar effects for the early post-expansion cohorts. The results imply that the expansion potentially increased income inequality across households by increasing assortative mating by college-educated individuals.

[42]If we do not make the age adjustment, the results are similar and do not qualitatively change our conclusion.

Table 11: Effects of the College Expansion on Assortative Mating: Absolute-Difference Index

		Late Post-expansion Cohorts: 1985-88		Early Post-expansion Cohorts: 1981-84	
	Region by Expansion Intensity	High	Low	High	Low
Cohorts	Pre-expansion: 1975-78	0.082	0.050	0.095	0.054
	Post-expansion	0.149	0.075	0.158	0.075
	Diff-in-diff	**0.041*****		**0.042*****	
		(0.002)		(0.002)	

Notes: The index is calculated using the sample of married couples in the random samples of 2010 census and 2015 mini-census. The sample includes married couples with either side falling in the specified cohort range. Ages of couples in pre-expansion cohorts are adjusted to be comparable to those in the post-expansion cohorts. The national sample is divided into high vs. low regions based on whether the treatment proxy is above or below median. Standard errors in parentheses are estimated by boostrapping from the original sample 1,000 times. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Compared to graduates of three-year colleges, graduates of four-year universities on average have more human capital and higher earnings. In Table F10, we estimate how the expansion changed the level of assortative mating by four-year university degrees versus lower education levels. We again find that the expansion had a positive and non-negligible impact on assortative mating.

6.3 Comparison with Previous Measures of Assortative Matching

Many methods have been developed to measure assortative matching. The goal of most previous studies is to compare the underlying tendencies for assortative mating (by education) of two marriage markets (characterized by two contingency tables) that have different marginal distributions of education types. Put another way, the indexes are designed to fully adjust for the difference in marginal distributions. That goal slightly differs from ours because we are curious about the impact of the expansion, which itself is a change in marginal distributions of education, on matching patterns and its implications for inequality. Put another way, we are investigating the effects of changing marginal distributions of education types rather than trying to fully adjust for it. Therefore, the existing measures do not perfectly fit our purpose.

To illustrate the point, we compare our index with several measures that have been widely used and examined: log odds ratio (Siow, 2015; Chiappori et al., 2020; Ciscato and Weber, 2020), minimum distance (Liu and Lu, 2006; Fernández and Rogerson, 2001), and rank correlation. Given the matching outcomes characterized by Table 10, the definitions of these indexes are provided in Table 12.

Table 12: Compare Different Matching Indexes

	Definition	Range	Scale invariance	Symmetry	Monotonicity	Perfect PAM
Log Odds Ratio	$\ln \frac{k_1 k_4}{k_2 k_3}$	$[0, \infty)$	Y	Y	Y	Y
Rank Correlation	$\frac{k_1 k_4 - k_2 k_3}{\sqrt{(k_1+k_2)(k_3+k_4)(k_1+k_3)(k_2+k_4)}}$	$[0, 1]$	Y	Y	Y	Y
Minimum Distance	$\frac{k_1 k_4 - k_2 k_3}{(k_1+min\{k_2,k_3\})*(k_4+min\{k_2,k_3\})}$	$[0, 1]$	Y	Y	Y	Y
Absolute Difference	$\frac{k_1 k_4 - k_2 k_3}{K^2}$	$[0, 0.25]$	Y	Y	Y	N

Notes: $K = k_1 + k_2 + k_3 + k_4$.

Chiappori et al. (2021) assess indexes of assortative mating used in the literature and propose that a satisfactory matching index should have the following properties. First, it should be invariant to the scale of the population. Second, it should be symmetric between the categories (in our case, college versus noncollege). Third, the monotonicity condition requires that it should increase with more people in diagonal cells (k_1 and k_4) when the marginal distributions are held constant. Fourth, the perfect-PAM (positive assortative mating) condition requires that a contingency table under perfectly positive assortative mating ($k_2 = k_3 = 0$, no off-diagonal matches) should exhibit the maximal value of the index. We compare the various indexes' properties in Table 12. In Appendix G, we discuss the properties of these indexes.

Our index (absolute difference) fails only the perfect-PAM condition. We argue that this failure is not a fatal threat to our index. To see why, consider the hypothetical situation of perfect assortative mating. A measure that satisfies the perfect-PAM condition should always achieve the maximum value no matter how the marginal distribution of education changes. However, if access to higher education was initially very low but then expanded to a much higher level, such an expansion would still contribute to inequality, which is the outcome of the joint forces of a high level of assortative mating and the higher education expansion. Therefore, we choose the absolute-difference index as the measure for the level of assortative mating.

Using the alternative indexes, we also explore whether the expansion affected the underlying assortative-mating patterns. The results are reported in Table F11. For none of the three alternative indexes are there any significant estimates for differential trends in the high-expansion versus low-expansion regions after the expansion. The results suggest that the expansion did not significantly change the underlying assortative-matching tendency, although it led to more college-college marriages relative to the benchmark of random matching.

6.4 Marriage Age Gap

In addition to education, age is another important trait for marriage matching. One often-used measure for matching patterns regarding age is the age gap between the husband and the wife. Our model does not incorporate matching on age and therefore does not provide theoretical guidance on how to think of the effects of the college expansion on matching by age. Still, investigating the expansion's impact on matching patterns by age helps us to understand how the expansion changed the marriage market. It may also provide insights about the channels for our main finding about the positive effects of the expansion on the marriage probability of college graduates: if the positive effects are driven by the LCMM, then we should probably expect that the expansion shrank the marriage age gap of college graduates because they increasingly met people from similar backgrounds at college.

In this section, we estimate the effects of the expansion on the marriage age gap using the DID specification in Equation 5. Table 13 reports the results for college women and men. Our estimates suggest that the expansion had a small but statistically significant negative impact on the marriage age gap. For example, Column (1) shows that a one-SD increase in the treatment proxy led to a 0.03-year (0.36 month) drop in the marriage age gap for college women in the late post-expansion cohorts. The results support the story that the expansion reduced search frictions in the LCMM. The expansion also had a negative effect on the marriage age gap of the early post-expansion cohorts: in Column (3), a one-SD increase in the treatment proxy decreased the age gap for college women by 0.07 years (0.84 months).[43]

In Table F12, we report the estimated effects of the expansion on the marriage age gap of noncollege groups. We again find a negative effect, implying that the expansion had spillover effects on the matching patterns of noncollege individuals beyond the (mostly) null effects on their marriage probabilities. We discuss the results and potential explanations in Appendix F.4.

[43] The effect on age gap seems to be larger for the early post-expansion cohorts than for the late post-expansion cohorts. Nevertheless, the effect sizes are very small, so the difference between the late and early post-expansion cohorts may not be economically meaningful.

Table 13: Effects of the College Expansion on the Marriage Age Gap of College Graduates

	(1)	(2)	(3)	(4)
Dependent variable: Age gap (Husband - Wife)				
Post-expansion cohorts	1985-88, 27-30 years old in 2015		1981-84, 31-34 years old in 2015	
Pre-expansion cohorts	1975-78, 27-30 years old in 2005		1975-78, 31-34 years old in 2009	
	Female	Male	Female	Male
ExpProxy∗Post	-0.0284**	-0.0309*	-0.0732***	-0.0866***
	(0.0142)	(0.0171)	(0.0172)	(0.0262)
Observations	20724	18683	22987	24347
Average age gap of				
Pre-expansion cohorts	2.104	1.331	1.995	1.839
Post-expansion cohorts	1.737	0.760	1.733	1.369

Notes: The marital outcome of pre-expansion cohorts is constructed using marriage history so that it is comparable to the post-expansion cohorts. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit represents one standard deviation across all prefectures. All regressions control for prefecture fixed effects and age fixed effects. Standard errors clustered at the prefecture level are in parentheses. There are 340 clusters. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

7 Conclusion

China's college expansion that began in 1999 has provided access to higher education for millions of young women and men. This paper has studied the impacts of that radical education reform on the marriage market. We combined a theoretical model with empirical analysis to uncover equilibrium effects of the expansion and understand the underlying mechanisms.

We first developed a marriage-matching model with educational investment and search frictions in the marriage market. One key assumption is that expanding access to higher education reduces average search frictions in the college marriage market. The main forces at work in the model are changes the relative distributions of different education types in the marriage market and the reduction in search costs in the LCMM due to the expansion. Our model suggests the expansion has had important general equilibrium effects on the marriage market. When the LCMM channel dominates the effects of changing the relative distributions of education types, the expansion raises the marriage probabilities of both college women and men.

Exploiting regional and cross-cohort variation in exposure to the expansion, we empirically estimated causal impacts of the expansion on the marriage market. We showed that the expansion indeed meaningfully increased the marriage probabilities of college women and men. We also found effects of the expansion on marriage-matching patterns. The expansion increased the level of assortative mating by education, as measured by the difference between the probability of actual college-college marriages and the hypothetical probability of such marriages under random matching. It also reduced the marriage age gap. Our empirical findings have important implications for critical issues in contemporary China such as the "leftover women" phenomenon and income inequality.

This paper adds to our understanding of the equilibrium effects of education institutions and education reforms on marriage outcomes. The findings of this paper are potentially important for policy makers when considering the impacts of education policies on critical lifetime outcomes, and are especially relevant for countries that are expanding or will expand access to higher education.

References

André, Pierre, and Yannick Dupraz. 2019. "Education and hypergamy: evidence from Cameroon."

Artmann, Elisabeth, Nadine Ketel, Hessel Oosterbeek, and Bas van der Klaauw. 2021. "Field of study and partner choice." *Economics of Education Review* 84 102149.

Attanasio, Orazio P, and Katja M Kaufmann. 2017. "Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations." *Journal of Economic Behavior & Organization* 140 35–55.

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

Bruze, Gustaf. 2015. "Male and female marriage returns to schooling." *International Economic Review* 56 (1): 207–234.

Burgess, Simon, Carol Propper, and Arnstein Aassve. 2003. "The role of income in marriage and divorce transitions among young Americans." *Journal of Population Economics* 16 (3): 455–475.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with a continuous treatment." *arXiv preprint arXiv:2107.02637*.

de Chaisemartin, Clément, Xavier D'Haultfoeuille, Félix Pasquier, and Gonzalo Vazquez-Bare. 2022. "Difference-in-Differences Estimators for Treatments Continuously Distributed at Every Period." *arXiv preprint arXiv:2201.06898*.

de Chaisemartin, Clément, Xavier D'Haultfœuille, and Yannick Guyonvarch. 2019. "Fuzzy differences-in-differences with stata." *The Stata Journal* 19 (2): 435–458.

Che, Yi, and Lei Zhang. 2018. "Human capital, technology adoption and firm performance: Impacts of China's higher education expansion in the late 1990s." *The Economic Journal* 128 (614): 2282–2320.

Chen, Ting, Yizhen Gu, and Ben Zou. 2022. "Delineating China's Metropolitan Areas Using Commuting Flow Data." *Available at SSRN*.

Chiappori, Pierre-André. 2017. *Matching with transfers: The economics of love and marriage*. Princeton University Press.

Chiappori, Pierre-Andre, Monica Costa Dias, and Costas Meghir. 2021. "The measuring of assortativeness in marriage: a comment."

Chiappori, Pierre-André, Monica Costa Dias, and Costas Meghir. 2018. "The marriage market, labor supply, and education choice." *Journal of Political Economy* 126 (S1): S26–S72.

Chiappori, Pierre-André, Monica Costa Dias, and Costas Meghir. 2020. "Changes in assortative matching: Theory and evidence for the us."

Chiappori, Pierre-Andre, Murat Iyigun, and Yoram Weiss. 2009. "Investment in schooling and the marriage market." *American Economic Review* 99 (5): 1689–1713.

Chiappori, Pierre-André, Bernard Salanié, and Yoram Weiss. 2017. "Partner choice, investment in children, and the marital college premium." *American Economic Review* 107 (8): 2109–67.

Choo, Eugene, and Aloysius Siow. 2006. "Who marries whom and why." *Journal of political Economy* 114 (1): 175–201.

Chu, Junhong, Haoming Liu, and Ivan PL Png. 2018. "Nonlabor income and age at marriage: evidence from China's heating policy." *Demography* 55 (6): 2345–2370.

Ciscato, Edoardo, and Simon Weber. 2020. "The role of evolving marital preferences in growing income inequality." *Journal of Population Economics* 33 (1): 307–347.

Ebenstein, Avraham Y, and Ethan Jennings Sharygin. 2009. "The consequences of the "missing girls" of China." *The world bank economic review* 23 (3): 399–425.

Eika, Lasse, Magne Mogstad, and Basit Zafar. 2019. "Educational assortative mating and household income inequality." *Journal of Political Economy* 127 (6): 2795–2835.

Feng, Shuaizhang, Xiaoyu Xia et al. 2018. "Endogenous skill-biased technology adoption: Evidence from China's college enrollment expansion program."

Feng, Yuan. 1999. "National college entrance examinations: The dynamics of political centralism in China's elite education." *Journal of education* 181 (1): 39–57.

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

Fincher, Leta Hong. 2016. *Leftover women: The resurgence of gender inequality in China*. Bloomsbury Publishing.

Ge, Run, and Jialin Huang. 2020. "Has the college expansion affected marriage and fertility?" *China Journal of Economics (in Chinese)* 3 168–201.

Ge, Suqin. 2011. "Women's college decisions: how much does marriage matter?" *Journal of Labor Economics* 29 (4): 773–818.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos. 2014. "Marry your like: Assortative mating and income inequality." *American Economic Review* 104 (5): 348–53.

Hankins, Scott, and Mark Hoekstra. 2011. "Lucky in life, unlucky in love? The effect of random income shocks on marriage and divorce." *Journal of Human Resources* 46 (2): 403–426.

Hener, Timo, and Tanya Wilson. 2018. "Marital age gaps and educational homogamy-evidence from a compulsory schooling reform in the uk."

Kai-Ming, Cheng, Jin Xinhuo, and Gu Xiaobo. 1999. "From training to education. Lifelong learning in China." *Comparative Education* 35 (2): 119–129.

Kirkebøen, Lars, Edwin Leuven, and Magne Mogstad. 2021. "College as a Marriage Market."

Lafortune, Jeanne. 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, Hongbin, Yueyuan Ma, Lingsheng Meng, Xue Qiao, and Xinzheng Shi. 2017. "Skill complementarities and returns to higher education: Evidence from college enrollment expansion in China." *China Economic Review* 46 10–26.

Liu, Haoming, and Jingfeng Lu. 2006. "Measuring the degree of assortative mating." *Economics Letters* 92 (3): 317–322.

Ma, Xiao. 2020. "College Expansion, Trade, and Innovation: Evidence from China."

Magistad, Mary Kay. 2013. "China's 'leftover women', unmarried at 27." *BBC News* 21.

Marioulas, Julian. 2017. "China: A world leader in graduation rates." *International Higher Education* (90): 28–29.

Nybom, Martin, Erik Plug, Bas Van der Klaauw, and Lennart Ziegler. 2017. "Skills, Parental Sorting, and Child Inequality."Technical report, Working Paper.

Salcher, Isabelle. 2020. "Schooling and Marriage Markets How an Education Reform Impacted Matching Patterns: Evidence from Zimbabwe."

Schofer, Evan, and John W Meyer. 2005. "The worldwide expansion of higher education in the twentieth century." *American sociological review* 70 (6): 898–920.

Siow, Aloysius. 2015. "Testing Becker's theory of positive assortative matching." *Journal of Labor Economics* 33 (2): 409–441.

Su, Puyu, Jiahu Hao, Zhaohui Huang, Xiao Limin, and Fangbiao Tao. 2011. "A study on intimate partner violence based on 2575 college students (in Chinese)." *Chinese Journal of Epidemiology* 32 (4): 346–351.

To, Sandy. 2015. *China's Leftover Women: Late marriage among professional women and its consequences*. Routledge.

Wang, Naixia. 2011. "Adult higher education in China: problems and potential." In *Higher Education Reform in China*, 44–61, Routledge.

Wang, Qinghua. 2014. "Crisis management, regime survival and "guerrilla-style" policy-making: the June 1999 decision to radically expand higher education in China." *The China Journal* (71): 132–152.

Wang, Shumei, and Likang Wang. 2000. "A study on health education about the HIV based on 2974 college students in Shanghai (in Chinese)." *Chinese Journal of School Doctor* 14 (1): 20–21.

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

World Bank. 2017. *World development report 2018: Learning to realize education's promise*. The World Bank.

Wu, Daguang, Manman Teng, and Duanmiao Li. 2016. "College Graduation Rate and the Quality of Higher Education: A Comparative Study Based on China and the US (in Chinese)." *Journa of Higher Education* 12.

Zha, Danyan. 2022. "Schooling Expansion and the Female Marriage Age: Evidence from Indonesia."

Zhang, Hanzhe. 2021. "An investment-and-marriage model with differential fecundity: On the college gender gap." *Journal of Political Economy* 129 (5): 1464–1486.

Zhong, Hai. 2011. "Returns to higher education in China: What is the role of college quality?" *China Economic Review* 22 (2): 260–275.

Appendix

A More Details about The Higher Education System in China and the College Expansion

A.1 College Types

There are different types of colleges in China. We focus on regular colleges in this paper. Regular colleges admit students through the national college entrance examination (a.k.a "Gaokao"), and the exam score is the sole determinant of which colleges the students are admitted into. Students are required to be fulltime and on-site. There are two types of regular college education. One is four-year universities (*Benke*), the other is junior colleges (*Zhuanke*), which usually requires 3 years to accomplish. Four-year universities are better funded by the central and the local government, have more teachers, and are considered as having higher quality and prestige. Four-year universities can be regarded as Tier 1 and junior colleges can be regarded as Tier 2. Students admitted into four-year universities have much higher scores. Most students take the Exam right after high school, around the age of 18.

There is also a post-secondary credential system (sometimes referred to as *adult higher education*) in China. The major difference of this system is that students do NOT need to take the same college entrance exam as the regular college students and the admission bar is close to open-enrollment. Students are not required to study regularly on-site. Instead, they may attend classes on a part-time basis (e.g. at night or over weekend), and they may also study remotely (e.g. online). The degrees obtained are different from those of regular colleges, and the requirements for graduation are also much lower compared to those of regular colleges. Unsurprisingly, the degrees are not rewarded as much as regular college degrees in the labor market (and possibly also in the marriage market).

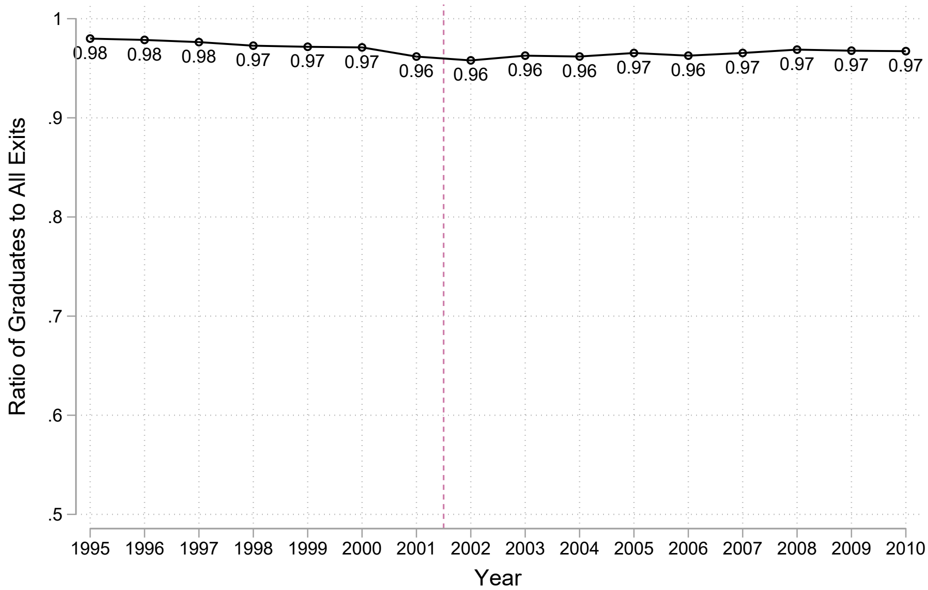
One related issue in the China census data (and most household surveys) is that when respondents report their educational attainment, they only report whether they have a college degree, and whether it is *Benke* or *Zhuanke*, but do not distinguish between regular and special college degrees. The only census year that this information is collected in 2000. CFPS data collects this information. We rely on CFPS data to show the relative importance of different colleges in China's higher education system, and over the period of college expansion.

A.2 Dropout in Chinese Universities

Unlike in some other contexts (e.g. the US), the college dropout rate is extremely low in China. For example, Marioulas (2017) states that "China has one of the lowest college dropout rates in the world, with sources from the ministry of education, who state that less than 1% of students fail to complete their degrees." Using confidential data from a random sample of 646 universities, Wu et al. (2016) documented that the average graduation rate was 96.91% and only 2.63% among these universities had a graduation rate that was below 90%. Moreover, the graduation rate did not systematically differ by college quality. For example, the graduation rate in Project 985 universities (the top 39 universities in China) was 95.51%, only slightly lower than the average value.

To check if the pattern of dropout changed after the college expansion, we plot in Figure A1 the ratio of graduates to all exits from higher education institutions in each year using data from the Chinese Education Yearbooks.[1] Ideally, we would like to calculate the rate of graduation for students that newly enrolled in a given year. However, the Yearbooks do not distinguish between four-year universities and three-year colleges when reporting the total number of graduates each year. Nevertheless, Figure A1 shows that most college students successfully graduated and that pattern did not change following the college expansion.

Figure A1: The Ratio of Graduates to All Exits from Higher Education Institutions



Source: The Chinese Education Yearbooks. This figure reports the ratio of graduates to all exits from higher education institutions in each year. Both four-year universities and three-year colleges are included. The college expansion started in 1999. Three-year-college students who were newly enrolled in 1999 normally graduated in 2002 (as indicated by the vertical line), and four-year-university students in 2003.

[1]Other potential reasons for exiting include completion without a degree, being suspended, being expelled, quitting, and death.

B College Local Marriage Markets: Suggestive Evidence from the CFPS

We investigate the role of colleges as local marriage markets using the CFPS data. In the CFPS, married respondents were asked one question about how they met their (first and current) spouse. Among other options, they were asked whether they met their spouse (both current and first spouse) "at school" by themselves. The specific question is as follows:

How did you get to know your current/first spouse? [Select only one response]

1. Knew each other at school by ourselves

2. Knew each other at workplace by ourselves

3. Knew each other at place of residence by ourselves

4. Met each other at other places by ourselves

5. Introduced by friends/relatives

6. Through marriage agency

7. Arranged by parents

8. Through the Internet

77. Other [Please specify]

We take the answer of "meeting spouse at school" as a proxy for how school/college serves as a local marriage market.[2] Using the answers to this information, we define a dummy variable for married individuals meeting their spouse on campus, which equals to 1 if the answer to the question is "met her/him at school by myself". For each cohort, we can then look at what how many marriages arose from acquaintance in college.

In Table 1, we have shown the fractions of people who met spouses in school by college education and before versus after the expansion. One potential concern is that the observed couples in the post-expansion cohorts were younger than the pre-expansion cohorts. It could be that mechanically younger couples are more likely those who meet on campus. To adjust for this potential bias, we further restrict the sample to those who got married by the age of 27 (i.e. those who married early) in both the pre- and the post-expansion cohorts. The results are reported in Table B1. After adjusting for early marriages, we still observe similar patterns: college graduates are much more likely to meet their spouses in school and the fraction increased after the college expansion.

[2]One advantage of this proxy is that we directly observe how individuals met each other, while previous literature use only their common experience to gauge the role of common school/workplace in marriage matching.

Table B1: Fraction of People Who Met Their Spouses in School: Early Marriages

	Cohorts	Fraction	Observations
College	1975-80 (Pre-expansion)	0.25	183
	1981-88 (Post-expansion)	0.32	643
	Difference	0.07 ($p < 0.1$)	
Non-college	1975-80 (Pre-expansion)	0.04	2449
	1981-88 (Post-expansion)	0.06	4038
	Difference	0.02 ($p < 0.01$)	

Source: China Family Panel Studies 2010-2018. All results weighted using the CFPS survey weights. The sample includes only individuals who were married by 27.

B.1 Suggestive Causal Evidence

We show that the probability of meeting one's spouse in college has increased after the college expansion. However, this before-after comparison could be contaminated by other secular trends. We provide some suggestive causal evidence in this section by exploiting a DID design analogous to our baseline econometric model. We know individuals' provinces of residence. Therefore, we estimate the following DID model:

$$meetschool_{ipb} = \rho_1 ExpProxy_p * Post_b + \rho_2 * Male_i + \lambda_p + \xi_b + \varepsilon_{ipb} \tag{B1}$$

The dependent variable is a dummy variable for meeting one's spouse in school for individual i of cohort of birth b in province (local marriage market) p. $ExpProxy_p$ is the same as the expansion proxy used in our baseline model but measured at the province level (Section 5.5). $Post_b$ is a dummy for the post-expansion cohorts (1981–88).[3] We further control for gender, cohort fixed effects, and province fixed effects. The results are reported in Table B2. Column (1) shows the results for college graduates, which suggest the college expansion indeed increased the probability of meeting one's spouse in school. We do not find effects of the college expansion on non-college individuals, as reported in Column (2).

[3] For the DID model, we include pre-expansion cohorts in 1975–78 and post expansion cohorts in 1981–88 that are consistent with our baseline econometric model.

Table B2: Effects of the College Expansion on Meeting Spouses in School

	(1) College	(2) Non-college
ExpProxy∗Post	0.0281**	-0.0176
	(0.0119)	(0.0151)
Bootstrapped p-value	[0.154]	[0.375]
N	1690	7658

Dependent variable is a dummy for meeting one's spouse in school. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit represents one standard deviation across all provinces. Standard errors clustered at the province level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Boostrapped p-values are reported in brackets. Results are weighted using the CFPS survey weights.

C Appendix to the Model and Simulation

C.1 Characterization of the Full Model Equilibrium

C.1.1 Educational Choices

In this subsection, we characterize the full model including educational choices and the marriage matching process. The full equilibrium of the model is determined by the fixed point of educational choices and marriage market matching outcomes. To capture this, let's first reconsider the first step of the model: educational choices. Recall that the costs of college education are assumed to follow:

$$c_i^f = c_f + \theta_i^f$$
$$c_j^m = c_m + \theta_j^m.$$

Consider the choice problem of female i (we use i for individual woman and j for individual man). If she chooses not to go to college, then the expected marital payoff conditional on education type L is U^L, which is determined as below (E is the Euler's constant ≈ 0.577):[4]

$$\begin{aligned} U^L =& Emax\{u_{iLH}, u_{iLL}, u_{iL0}\} \\ =& E(u_{iLy} | y = arg \max_{y=0,H,L} u_{iLy}) \\ =& E + \ln(\sum_{y=0,H,L} \exp(\alpha_{Ly} - \delta + \tau_{Ly})). \end{aligned} \tag{C2}$$

The equation is derived based on the type-I extreme value distribution of the idiosyncratic

[4]In the choice of an L type, she does not distinguish between H_1 and H_2 because these two are identical in her preferences.

marital preferences ϵ_{iy} (Choo and Siow, 2006).

Similarly, denote her expected marital payoff (without paying the education cost) conditional on choosing education type H is U^H. U^H follows a similar structure but also depends on whether she randomly enters the LCMM or not.

$$\begin{aligned}U^H &= \sum_{k=1,2} p^f(H_k)E(u_{iH_ky}|y = arg \max_{y=0,H_k,L} u_{iH_ky}) \\ &= p(H_1)[E + \ln(\exp(\alpha_{HH} + \tau^1_{HH}) + \sum_{y=0,L} \exp(\alpha_{Hy} - \delta + \tau_{Hy}))] \\ &+ (1 - p(H_1))[E + \ln(\exp(\alpha_{HH} - \delta + \tau^2_{HH}) + \sum_{y=0,L} \exp(\alpha_{Hy} - \delta + \tau_{Hy}))]. \end{aligned} \tag{C3}$$

$p^f(H_1) = \frac{R(H_f,H_m)}{H_f}$ is the probability of entering the LCMM.

Woman i compares the expected payoff of college education ($U^H - c^f_i$) and that of not choosing college education (U^L). Conditional on the expected payoffs U^H and U^L, the distribution of c^f_i determines the distribution of education. The college ratio of women is determined by

$$h_f = G^f(U^H - U^L - c_f). \tag{C4}$$

Similarly, for the education choice problem of men, V^L is the expected martial payoff conditional on being the noncollege type, and V^H is the expected martial payoff conditional on college education (without paying the education cost), we have:

$$\begin{aligned}V^L =& E(v_{xLj}|x = arg \max_{x=0,H,L} v_{xLj}) \\ =& E + \ln(\sum_{x=0,H,L} \exp(\gamma_{xL} - \delta - \tau_{xL})). \end{aligned} \tag{C5}$$

$$\begin{aligned}V^H &= \sum_{k=1,2} p^m(H_k)E(v_{xH_kj}|x = arg \max_{x=0,H_k,L} v_{xH_kj}) \\ &= p^m(H_1)[E + \ln(\exp(\gamma_{HH} - \tau^1_{HH}) + \sum_{x=0,L} \exp(\gamma_{xH} - \delta - \tau_{xH}))] \\ &+ (1 - p^m(H_1))[E + \ln(\exp(\gamma_{HH} - \delta - \tau^2_{HH}) + \sum_{x=0,L} \exp(\gamma_{xH} - \delta - \tau_{xH}))]. \end{aligned} \tag{C6}$$

$$h_m = G^m(V^H - V^L - c_m). \tag{C7}$$

C.1.2 Marriage Market Choices

Conditional on educational choices and educational distribution, men and women make their choices of desired partners. The choices are made given the systematic marital returns, marital transfer (utility price) τ, and individual taste shocks (which follow type-I extreme value distribution). This can be transformed into a discrete choice problem, where each individual chooses their preferred spouse education type (including staying single). Assume the measure of women (men) belonging to type x (y) is μ_f^x (μ_m^y), and among them, μ_f^{xy} (μ_m^{xy}) choose the type y (x) spouse. $y = 0$ ($x = 0$) denotes staying single. This discrete choice problem leads to the following conditions for individual choices (**?**).

For women type H_1 ($\mu_f^{H_1} = \mu_f^{H_1H_1} + \mu_f^{H_1L} + \mu_f^{H_10}$):

$$\frac{\mu_f^{H_1H_1}}{\mu_f^{H_1}} = \frac{\exp(\alpha_{HH} + \tau_{HH}^1)}{\exp(\alpha_{HH} + \tau_{HH}^1) + \exp(\alpha_{HL} + \tau_{HL} - \delta) + \exp(\alpha_{H0})} \tag{C8}$$

$$\frac{\mu_f^{H_1L}}{\mu_f^{H_1}} = \frac{\exp(\alpha_{HL} + \tau_{HL} - \delta)}{\exp(\alpha_{HH} + \tau_{HH}^1) + \exp(\alpha_{HL} + \tau_{HL} - \delta) + \exp(\alpha_{H0})} \tag{C9}$$

$$\frac{\mu_f^{H_10}}{\mu_f^{H_1}} = \frac{\exp(\alpha_{H0})}{\exp(\alpha_{HH} + \tau_{HH}^1) + \exp(\alpha_{HL} + \tau_{HL} - \delta) + \exp(\alpha_{H0})} \tag{C10}$$

For women type H_2 ($\mu_f^{H_2} = \mu_f^{H_2H_2} + \mu_f^{H_2L} + \mu_f^{H_20}$):

$$\frac{\mu_f^{H_2H_2}}{\mu_f^{H_2}} = \frac{\exp(\alpha_{HH} + \tau_{HH}^2 - \delta)}{\exp(\alpha_{HH} + \tau_{HH}^2 - \delta) + \exp(\alpha_{HL} + \tau_{HL} - \delta) + \exp(\alpha_{H0})} \tag{C11}$$

$$\frac{\mu_f^{H_2L}}{\mu_f^{H_2}} = \frac{\exp(\alpha_{HL} + \tau_{HL} - \delta)}{\exp(\alpha_{HH} + \tau_{HH}^2 - \delta) + \exp(\alpha_{HL} + \tau_{HL} - \delta) + \exp(\alpha_{H0})} \tag{C12}$$

$$\frac{\mu_f^{H_20}}{\mu_f^{H_2}} = \frac{\exp(\alpha_{H0})}{\exp(\alpha_{HH} + \tau_{HH}^2 - \delta) + \exp(\alpha_{HL} + \tau_{HL} - \delta) + \exp(\alpha_{H0})} \tag{C13}$$

For women type L ($\mu_f^L = \mu_f^{LH} + \mu_f^{LL} + \mu_f^{L0}$):

$$\frac{\mu_f^{LH}}{\mu_f^L} = \frac{\exp(\alpha_{LH} + \tau_{LH} - \delta)}{\exp(\alpha_{LH} + \tau_{LH} - \delta) + \exp(\alpha_{LL} + \tau_{LL} - \delta) + \exp(\alpha_{L0})} \tag{C14}$$

$$\frac{\mu_f^{LL}}{\mu_f^L} = \frac{\exp(\alpha_{LL} + \tau_{LL} - \delta)}{\exp(\alpha_{LH} + \tau_{LH} - \delta) + \exp(\alpha_{LL} + \tau_{LL} - \delta) + \exp(\alpha_{L0})} \tag{C15}$$

$$\frac{\mu_f^{L0}}{\mu_f^L} = \frac{\exp(\alpha_{L0})}{\exp(\alpha_{LH} + \tau_{LH} - \delta) + \exp(\alpha_{LL} + \tau_{LL} - \delta) + \exp(\alpha_{L0})} \tag{C16}$$

Similarly, we have 9 equations governing the choices of men.

For men type H_1 $(\mu_m^{H_1} = \mu_m^{H_1H_1} + \mu_m^{LH_1} + \mu_m^{0H_1})$:

$$\frac{\mu_m^{H_1H_1}}{\mu_m^{H_1}} = \frac{\exp(\gamma_{HH} - \tau_{HH}^1)}{\exp(\gamma_{HH} - \tau_{HH}^1) + \exp(\gamma_{LH} - \tau_{LH} - \delta) + \exp(\gamma_{0H})} \tag{C17}$$

$$\frac{\mu_m^{LH_1}}{\mu_m^{H_1}} = \frac{\exp(\gamma_{LH} - \tau_{LH} - \delta)}{\exp(\gamma_{HH} - \tau_{HH}^1) + \exp(\gamma_{LH} - \tau_{LH} - \delta) + \exp(\gamma_{0H})} \tag{C18}$$

$$\frac{\mu_f^{0H_1}}{\mu_m^{H_1}} = \frac{\exp(\gamma_{0H})}{\exp(\gamma_{HH} - \tau_{HH}^1) + \exp(\gamma_{LH} - \tau_{LH} - \delta) + \exp(\gamma_{0H})} \tag{C19}$$

For men type H_2 $(\mu_m^{H_2} = \mu_m^{H_2H_2} + \mu_m^{LH_2} + \mu_m^{0H_2})$:

$$\frac{\mu_m^{H_2H_2}}{\mu_m^{H_2}} = \frac{\exp(\gamma_{HH} - \delta - \tau_{HH}^2)}{\exp(\gamma_{HH} - \delta - \tau_{HH}^2) + \exp(\gamma_{LH} - \tau_{LH} - \delta) + \exp(\gamma_{0H})} \tag{C20}$$

$$\frac{\mu_m^{LH_2}}{\mu_m^{H_2}} = \frac{\exp(\gamma_{LH} - \tau_{LH} - \delta)}{\exp(\gamma_{HH} - \delta - \tau_{HH}^2) + \exp(\gamma_{LH} - \tau_{LH} - \delta) + \exp(\gamma_{0H})} \tag{C21}$$

$$\frac{\mu_m^{0H_2}}{\mu_m^{H_2}} = \frac{\exp(\gamma_{0H})}{\exp(\gamma_{HH} - \delta - \tau_{HH}^2) + \exp(\gamma_{LH} - \tau_{LH} - \delta) + \exp(\gamma_{0H})} \tag{C22}$$

For men type L $(\mu_m^L = \mu_m^{HL} + \mu_m^{LL} + \mu_m^{0L})$:

$$\frac{\mu_m^{HL}}{\mu_m^L} = \frac{\exp(\gamma_{HL} - \tau_{HL} - \delta)}{\exp(\gamma_{HL} - \tau_{HL} - \delta) + \exp(\gamma_{LL} - \tau_{LL} - \delta) + \exp(\gamma_{0L})} \tag{C23}$$

$$\frac{\mu_m^{LL}}{\mu_m^L} = \frac{\exp(\gamma_{LL} - \tau_{LL} - \delta)}{\exp(\gamma_{HL} - \tau_{HL} - \delta) + \exp(\gamma_{LL} - \tau_{LL} - \delta) + \exp(\gamma_{0L})} \tag{C24}$$

$$\frac{\mu_m^{L0}}{\mu_m^L} = \frac{\exp(\gamma_{0L})}{\exp(\gamma_{HL} - \tau_{HL} - \delta) + \exp(\gamma_{LL} - \tau_{LL} - \delta) + \exp(\gamma_{0L})} \tag{C25}$$

C.1.3 Marriage Market Equilibrium

The marriage market equilibrium must satisfy the following conditions for consistent choices between women and men (equaling "demand" and "supply").

$$\mu_f^{H_1H_1} = \mu_m^{H_1H_1} \tag{C26}$$

$$\mu_f^{H_2H_2} = \mu_m^{H_2H_2} \tag{C27}$$

$$\mu_f^{H_1L} + \mu_f^{H_2L} = \mu_m^{HL} \tag{C28}$$

$$\mu_f^{LH} = \mu_m^{LH_1} + \mu_m^{LH_2} \tag{C29}$$

$$\mu_f^{LL} = \mu_m^{LL} \tag{C30}$$

In addition, the quantities must satisfy the following accounting identities.

$$\mu_f^{H_1H_1} + \mu_f^{H_1L} + \mu_f^{H_10} = R \text{ ...} H_1 \text{ female} \tag{C31}$$
$$\mu_f^{H_2H_2} + \mu_f^{H_2L} + \mu_f^{H_20} = H_f - R \text{ ...} H_2 \text{ female} \tag{C32}$$
$$\mu_m^{H_1H_1} + \mu_m^{LH_1} + \mu_m^{0H_1} = R \text{ ...} H_1 \text{ male} \tag{C33}$$
$$\mu_m^{H_2H_2} + \mu_m^{LH_2} + \mu_m^{0H_2} = H_m - R \text{ ...} H_1 \text{ male} \tag{C34}$$
$$\mu_f^{LL} + \mu_f^{LH} + \mu_f^{L0} = N_f - H_f \text{ ...} L \text{ female} \tag{C35}$$
$$\mu_m^{LL} + \mu_m^{HL} + \mu_m^{0L} = N_m - H_m \text{ ...} L \text{ male} \tag{C36}$$

C.1.4 Equilibrium Conditions

Given the exogenous variables (education costs parameters and population size), the final set of endogenous variables are:

$$U^H, U^L, H_f, \mu_f^{H_1H_1}, \mu_f^{H_1L}, \mu_f^{H_10}, \mu_f^{H_2H_2}, \mu_f^{H_2L}, \mu_f^{H_20}, \mu_f^{LH}, \mu_f^{LL}, \mu_f^{L0}$$
$$V^H, V^L, H_m, \mu_m^{H_1H_1}, \mu_m^{H_1L}, \mu_m^{H_10}, \mu_m^{H_2H_2}, \mu_m^{H_2L}, \mu_m^{H_20}, \mu_m^{LH}, \mu_m^{LL}, \mu_m^{L0}$$
$$R, \tau_{HH}^1, \tau_{HH}^2, \tau_{HL}, \tau_{LH}, \tau_{LL}$$

We have 30 endogenous variables. To pin down the system, we have:

(1) Expression of U, V as a function of utility parameters and educational distributions (Equation C2, C3, C5 & C6): 4 equations.

(2) Discrete marriage choices (Equation C8–Equation C25): 18 equations. However, 6 out of these 18 equations are redundant because we do not explicitly list μ_f^x and μ_m^y as the final endogenous variables. To see this, let's substitute the expression $\mu_f^{H_1} = \mu_f^{H_1H_1} + \mu_f^{H_1L} + \mu_f^{H_10}$ into Equation C8–Equation C10, then one of the three equations (e.g. Equation C10) is redundant. Therefore, we effectively have 12 equations based on the discrete choice conditions.

(3) Marriage market clearing conditions and account identities (Equation C26–Equation C36: 11 equations.

(4) Definition of the LCMM meeting function

$$R = R(H_f, H_m). \tag{C37}$$

(6) Individual educational choices that determines the distribution of education types

$$H_m = N_m * h_m = N_m * G^m(V^H - V^L - c_m) \tag{C38}$$
$$H_f = N_f * h_f = N_f * G^f(U^H - U^L - c_f) \tag{C39}$$

We have 30 equations for 3 endogenous variables that totally pin down the equilibrium educational choices and marriage matching functions.

C.2 Estimate/Calibrate the Marital Payoffs and Search Cost Parameter

This subsection describes how we estimate or calibrate important payoff and cost parameters in the marriage market. Suppose we observe the following statistics:
(1) The distribution of different types of marriages (but cannot distinguish between H_1 and H_2).
(2) Within the type HH marriage, the fraction of H_1H_1 marriages (estimated using auxiliary information from the CFPS).

From the individual choice functions, we can derive that:

$$2\ln\frac{\mu^{H_1H_1}}{\sqrt{\mu^{H_10}\mu^{0H_1}}} = \alpha_{HH} + \gamma_{HH} - \alpha_{H0} - \gamma_{0H} \tag{C40}$$

$$2\ln\frac{\mu^{H_2H_2}}{\sqrt{\mu^{H_20}\mu^{0H_2}}} = \alpha_{HH} + \gamma_{HH} - \alpha_{H0} - \gamma_{0H} - 2\delta \tag{C41}$$

$$\frac{1}{2}\ln\frac{\mu^{H_1L}\mu^{H_2L}(\mu^{H_1L}+\mu^{H_2L})^2}{\mu^{H_10}\mu^{H_20}(\mu^{0L})^2} = \alpha_{HL} + \gamma_{HL} - \alpha_{H0} - \gamma_{0L} - 2\delta \tag{C42}$$

$$\frac{1}{2}\ln\frac{\mu^{LH_1}\mu^{LH_2}(\mu^{LH_1}+\mu^{LH_2})^2}{\mu^{0H_1}\mu^{0H_2}(\mu^{L0})^2} = \alpha_{LH} + \gamma_{LH} - \alpha_{L0} - \gamma_{0H} - 2\delta \tag{C43}$$

$$2\ln\frac{\mu^{LL}}{\sqrt{\mu^{L0}\mu^{0L}}} = \alpha_{LL} + \gamma_{LL} - \alpha_{L0} - \gamma_{0L} - 2\delta \tag{C44}$$

These are equivalent to Equation 1 & 2 and can be derived using the individual choice functions following Choo and Siow (2006). For the simulation, we need to at least identify the model parameters at the right-hand-side of these equations. However, these values are not readily available as we do not distinguish between H_1 and H_2.

C.2.1 Identify Parameters Associated with Noncollege Types

It turns out this does not matter for the martial surplus parameters involving noncollege types. Combining Equation C9, C10, C12, C13, we can easily show that:

$$\frac{\mu^{H_1L}}{\mu^{H_10}} = \frac{\mu^{H_2L}}{\mu^{H_20}}. \tag{C45}$$

As a result:

$$\frac{\mu^{H_1 L}}{\mu^{H_1 0}} = \frac{\mu^{H_2 L}}{\mu^{H_2 0}} = \frac{\mu^{H_1 L} + \mu^{H_2 L}}{\mu^{H_1 0} + \mu^{H_2 0}} = \frac{\mu^{HL}}{\mu^{H0}}. \tag{C46}$$

Plug Equation C46 back to Equation C42:

$$2 \ln \frac{\mu^{HL}}{\sqrt{\mu^{H0} \mu^{0L}}} = \alpha_{HL} + \gamma_{HL} - \alpha_{H0} - \gamma_{0L} - 2\delta. \tag{C47}$$

The intuition is that noncollege types do not care about the distinct between H_1 and H_2. This is because the search costs do not affect the trade off between H and L or between L and L: H_1 and H_2 are equally valuable to an L type individual. Analogously, we have:

$$2 \ln \frac{\mu^{LH}}{\sqrt{\mu^{L0} \mu^{0H}}} = \alpha_{LH} + \gamma_{LH} - \alpha_{L0} - \gamma_{0H} - 2\delta. \tag{C48}$$

Therefore, we can identify the marital surplus (net of the search cost) for HL, LH, and LL types of marriages.

C.2.2 Calibrate Search Cost and College-College Marital Surplus

Because we cannot distinguish between H_1 and H_2, additional information is required for knowing the surplus to HH marriages with and without paying the search cost.

Assume we can observe $\lambda = \frac{\mu^{H_1 H_1}}{\mu^{HH}}$ (estimated using the CFPS data). We have the following condition for the marriage market equilibrium which is analogous to Equation C45 :

$$\frac{\mu^{LH_1}}{\mu^{0H_1}} = \frac{\mu^{LH_2}}{\mu^{0H_2}} \tag{C49}$$

Recall that we defined the following quantities:

$$\lambda \stackrel{\text{def}}{=} \frac{\mu^{H_1 H_1}}{\mu^{HH}}$$
$$J_f \stackrel{\text{def}}{=} \frac{R - \mu^{H_1 H_1}}{H_f - R - \mu^{H_2 H_2}}$$
$$J_m \stackrel{\text{def}}{=} \frac{R - \mu^{H_1 H_1}}{H_m - R - \mu^{H_2 H_2}}.$$

Rearrange Equation C45 , we get:

$$\frac{\mu^{H_1L}}{\mu^{H_2L}} = \frac{\mu^{H_10}}{\mu^{H_20}} = \frac{\mu^{H_1L} + \mu^{H_10}}{\mu^{H_2L} + \mu^{H_20}} = \frac{R - \mu^{H_1H_1}}{H_f - R - \mu^{H_2H_2}} = J_f$$

Similarly, we can get

$$\frac{\mu^{LH_1}}{\mu^{LH_2}} = \frac{\mu^{0H_1}}{\mu^{0H_2}} = J_m$$

Note that $\mu^{H_1H_1} = \lambda\mu^{HH}$, $\mu^{H_2H_2} = (1-\lambda)\mu^{HH}$

These conditions imply:

$$\mu^{H_10} = \frac{J_f}{1+J_f}\mu^{H0},\ \mu^{H_20} = \frac{1}{1+J_f}\mu^{H0}$$
$$\mu^{0H_1} = \frac{J_m}{1+J_m}\mu^{0H},\ \mu^{0H_2} = \frac{1}{1+J_m}\mu^{0H}$$

The point of these equations above is that we can transform terms that we do not observe (the LFS) to things we can observe under certain functional form assumptions (RHS).

Plug these equations back to equations C40 and C41, we can get

$$\alpha_{HH} + \gamma_{HH} - \alpha_{H0} - \gamma_{0H} - 2\delta = 2\ln\frac{\mu^{HH}}{\sqrt{\mu^{H0}\mu^{0H}}} + 2\ln[(1-\lambda)\sqrt{(1+J_f)(1+J_m)}]$$
$$2\delta = 2\ln\frac{\lambda}{1-\lambda} - \ln J_f J_m$$

To pin down the value of these model parameters, we need to known the functional form of R. As noted in the paper, the form of R is not identified. Instead, we choose a Cobb-Douglas function that satisfies increasing returns to scale. We can then identify these values based on observed data moments.

C.2.3 Estimated/Calibrated Model Parameters

The value of λ is estimated based on the CFPS data: among the college-college marriages (pre-expansion cohorts), the fraction of couples that reported they met in school. We have $\lambda = 0.3$. Based on the value of λ and the conditions described above, we can estimate the parameters governing marital surplus and the search cost. The parameter values are reported in Table C3. Based on the the marriage model, we cannot separately identify α_{x0} and γ_{0y}. We normalize them as zero.[5]

[5] This does not affect our results as only the marital surplus matters in the model for individual decisions.

Table C3: Model Parameter Values

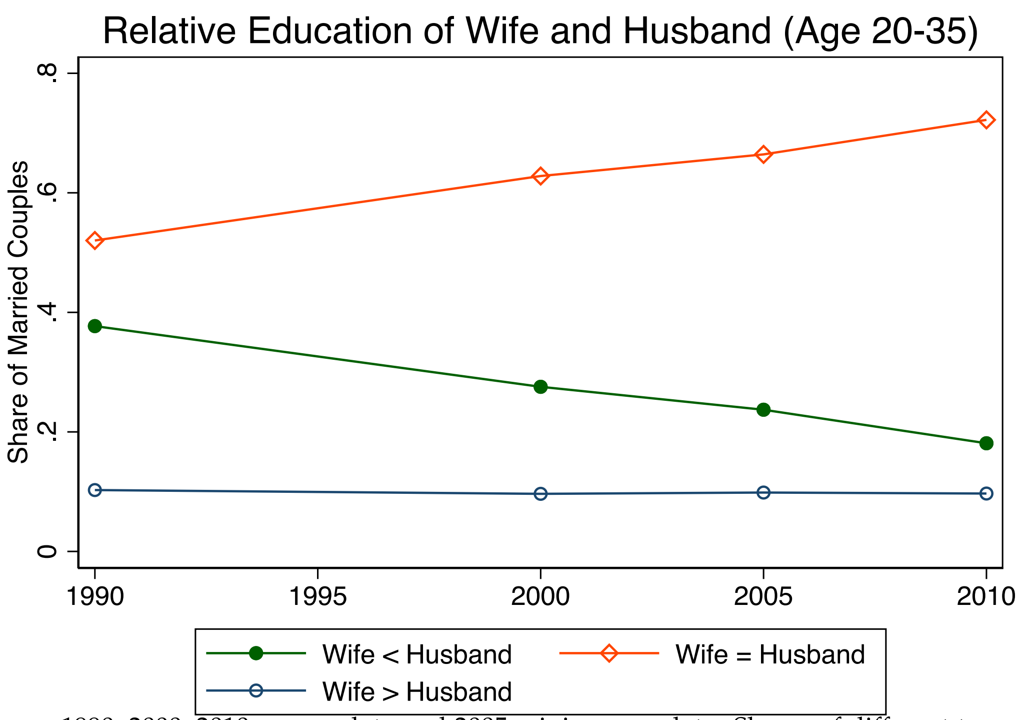
Marriage Types	Parameter	Estimated Value
HH	$\alpha_{HH} + \gamma_{HH} - \alpha_{H0} - \gamma_{0H}$	6.04
HL	$\alpha_{HL} + \gamma_{HL} - \alpha_{H0} - \gamma_{0L}$	2.32
LH	$\alpha_{LH} + \gamma_{LH} - \alpha_{L0} - \gamma_{0H}$	4.29
LL	$\alpha_{LL} + \gamma_{LL} - \alpha_{L0} - \gamma_{0L}$	8.34
	δ	1.88

Using these parameter values, we can then simulate (1) the marriage matching functions conditional on the distribution of education types and (2) the change in education choices and marriage matching functions following the change in the average costs of college education.

C.3 Additional Results for the Marriage Model

C.3.1 Distribution of Relative Education Within Married Couples

Figure C2: Relative Education of Wife and Husband

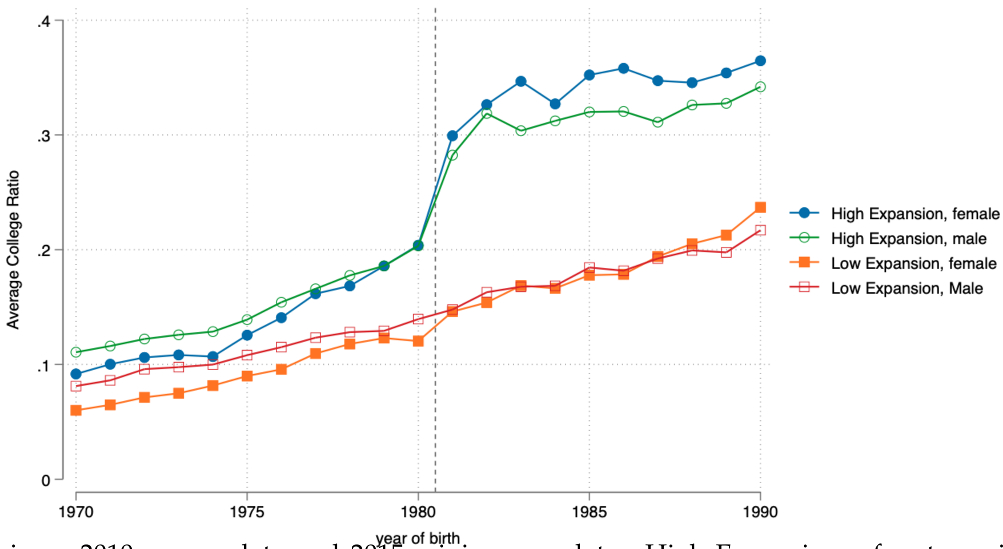


Source: Chinese 1990, 2000, 2010 census data and 2005 mini-census data. Shares of different types of couples by relative education are plotted.

C.3.2 Female College Ratios Overtook Male College Ratios in the College Expansion

Figure C3 reports the college ratio of women and men by cohort of birth. Post-expansion cohorts are those who were born after 1980. We divide the national sample into two groups based on the value of the college expansion proxy (Section 3.2). Three important patterns emerge from the figure. First, college ratios increased by more in high expansion regions relative to low expansion regions after the college expansion. This provides validation for our expansion proxy. Second, before the college expansion, the college ratio of men is higher than that of women. This gender difference has been reversed after the expansion. Third, the college ratio of women overtook that of men earlier and by a larger extent in high expansion regions. This lends supportive evidence to the effects of the college expansion on the reverse of the gender education gap.

Figure C3: College Ratio by Gender and College Expansion Intensity

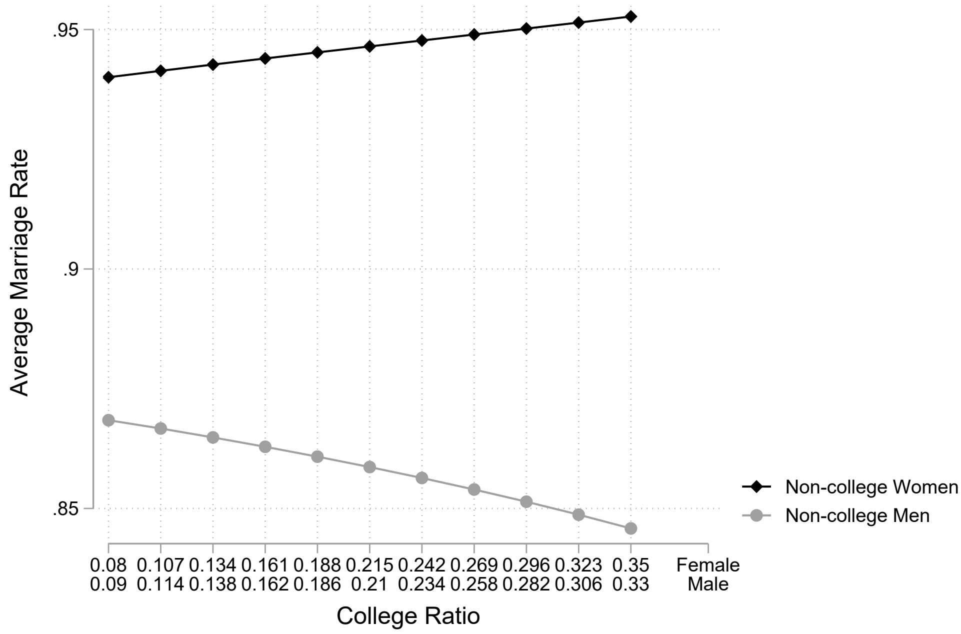


Source: Chinese 2010 census data and 2015 mini-census data. High Expansion refers to regions with the expansion intensity proxy above median. Low Expansion refers to regions with the expansion proxy below median.

C.3.3 Simulation Results for Noncollege Groups

Figure C4 presents the simulated comparative statics for the noncollege groups. The spillover effects in the marriage market also matter for the noncollege individuals. Although their search frictions are not affected in the model, their marriage outcomes are affected by the college expansion through changes in marginal distributions in the marriage market. The change in the relative distribution is such that both noncollege women and noncollege men become more scarce. However, to the extent that noncollege men are less likely than noncollege women to marry up, noncollege men will face stronger competition from the increasing supply of college men. As a result, the model predicts that the marriage prospect of noncollege men will decrease. At the same time, the increasing relative demand for noncollege women from both college men and noncollege men drives up their marriage prospects.

Figure C4: College Expansion and Marriage Rates: Simulated Results



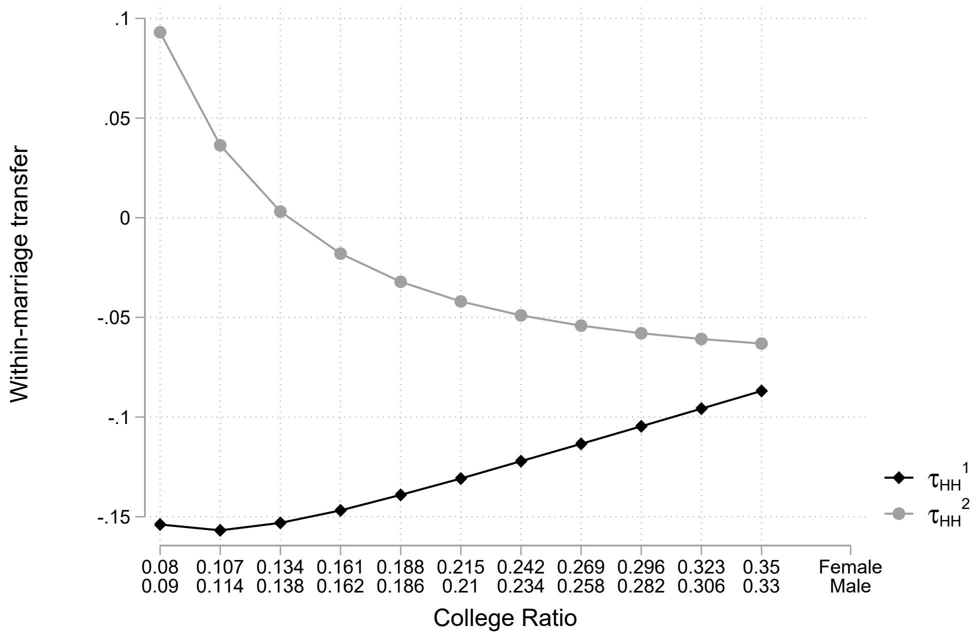
Notes: The x-axis displays the college ratios of women and men, which are allowed to evolve simultaneously.

C.3.4 Change in Within-marriage Transfers

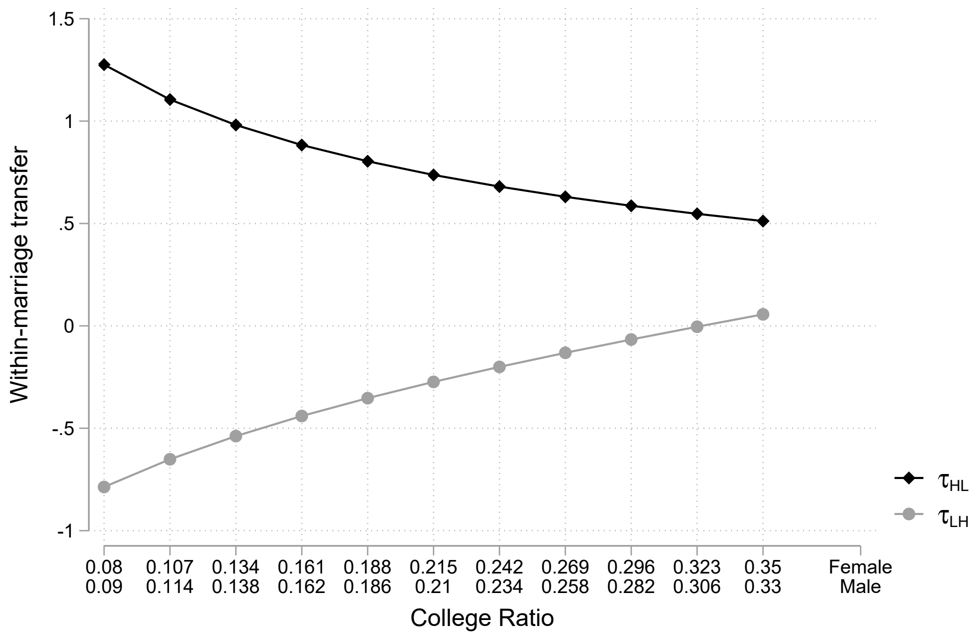
We report the change in simulated within-marriage transfers following the college expansion in C5. Panel B shows that, consistent with the intuition, college men pay noncollege women a higher transfer and college women receive from noncollege men a lower transfer following the college expansion. The case for within-college type marriages is more complicated. As the college expansion reduces the relative "marriageability" of college men by more than that of the college women (because college men are more likely to marry down), college men are willing to pay a higher transfer to college women. On the other hand, as the college ratio increases faster for college women than for college men, creating a relative "over-supply" of college women, they lose some bargaining power and might receive a lower transfer from college men. To see this more clearly, in Figure C7 we plot the within marriage transfers under the hypothetical situation of gender symmetric expansion (Figure 5). In this case, we can observe a clear increase in the within-marriage transfer from college-educated husband to college-educated wife. In Figure C6, we plot the change in transfers if there was no search frictions in the model. In this case, the within-college-marriage transfer decreases following the college expansion. This implies that the increasing supply of college women relative to that of college men plays a more important role in this no-friction scenario.

Figure C5: College Expansion and Within-marriage Transfer: Simulated Results

A. Transfers in within-college type marriages



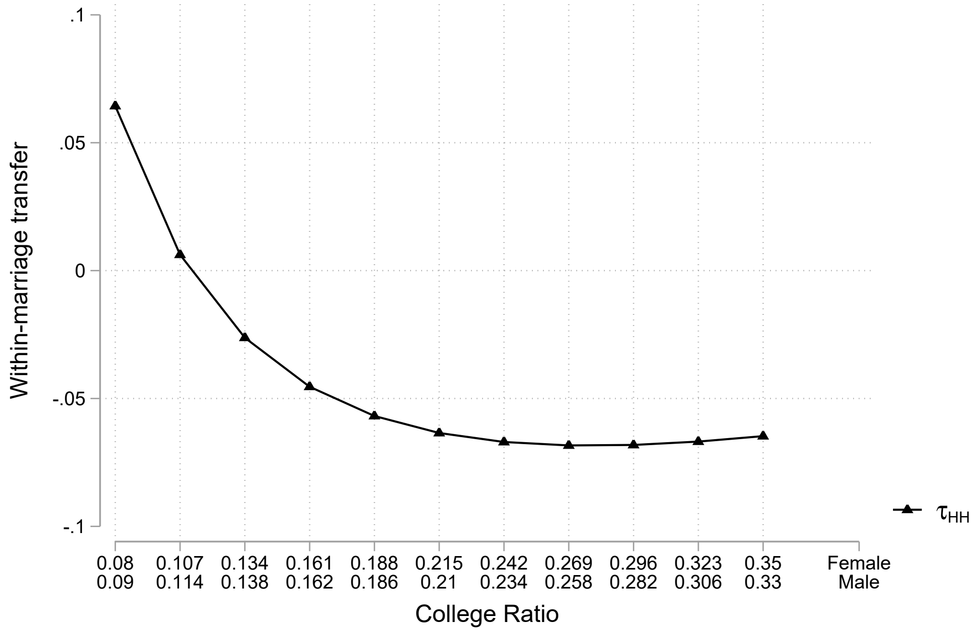
B. Transfers in cross-education type marriages



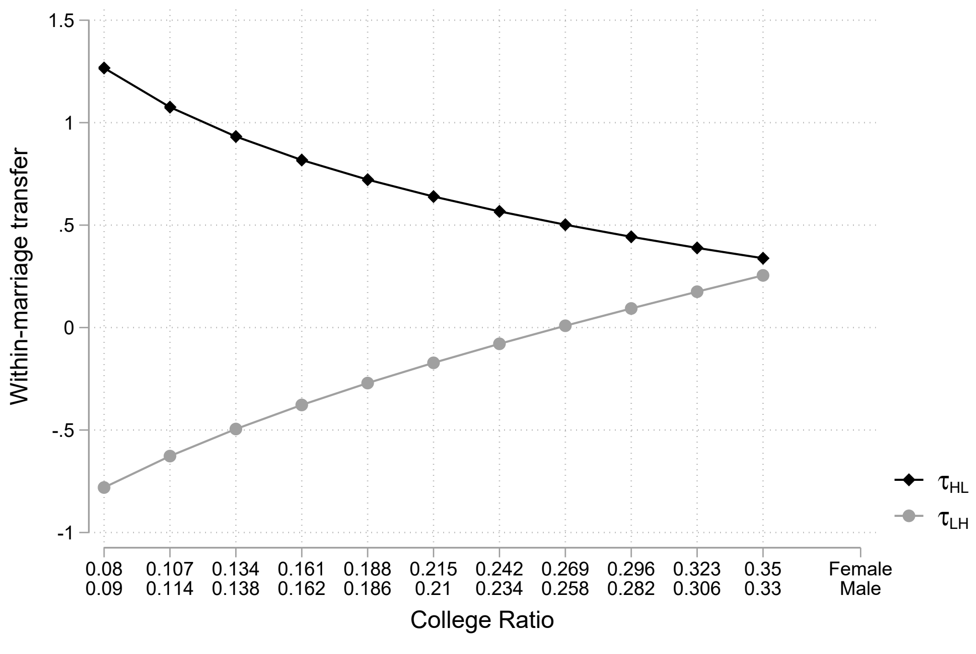
Notes: The x-axis displays the college ratios of women and men, which are allowed to evolve simultaneously.

Figure C6: College Expansion and Within-marriage Transfer: Simulated Results Without Local College Marriage Market

A. Transfers in within-college type marriages



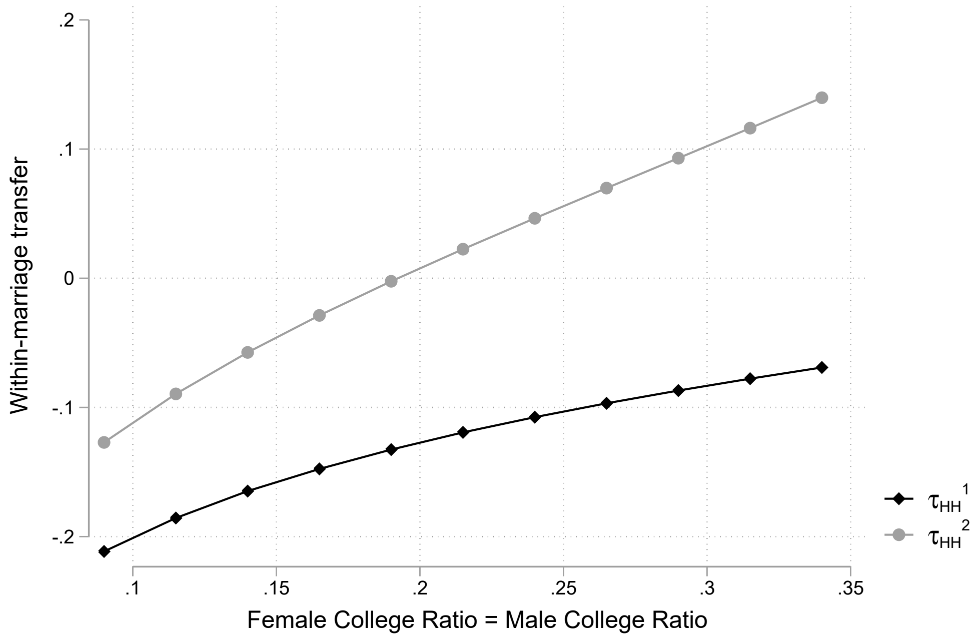
B. Transfers in cross-education type marriages



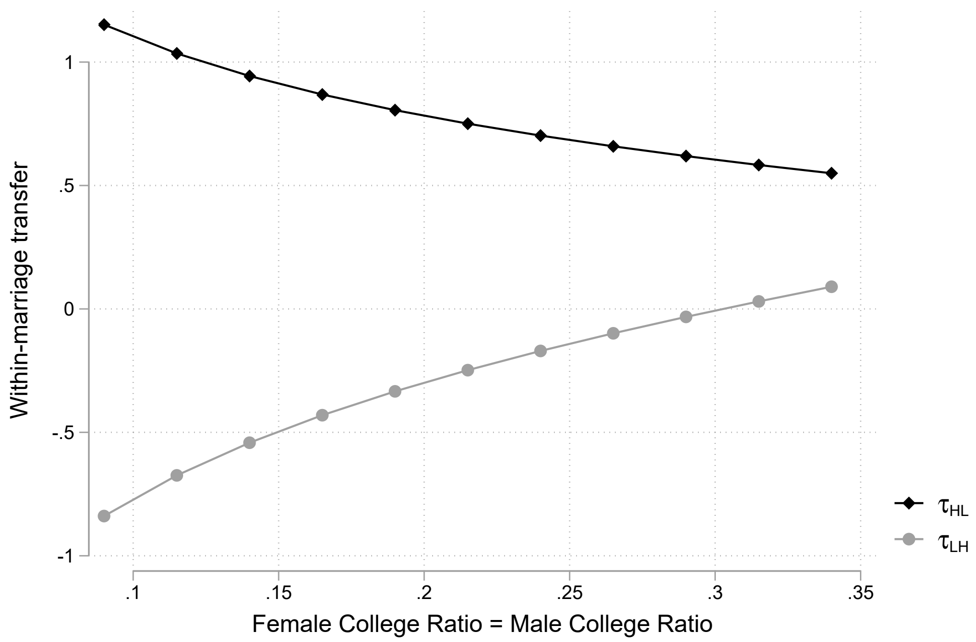
Notes: The x-axis displays the college ratios of women and men, which are allowed to evolve simultaneously.

Figure C7: College Expansion and Within-marriage Transfer: Simulated Results Under Gender-symmetric Expansion

A. Transfers in within-college type marriages



B. Transfers in cross-education type marriages



Notes: The x-axis displays the college ratios of women and men, which are allowed to evolve simultaneously.

C.4 Simulate the Full Model

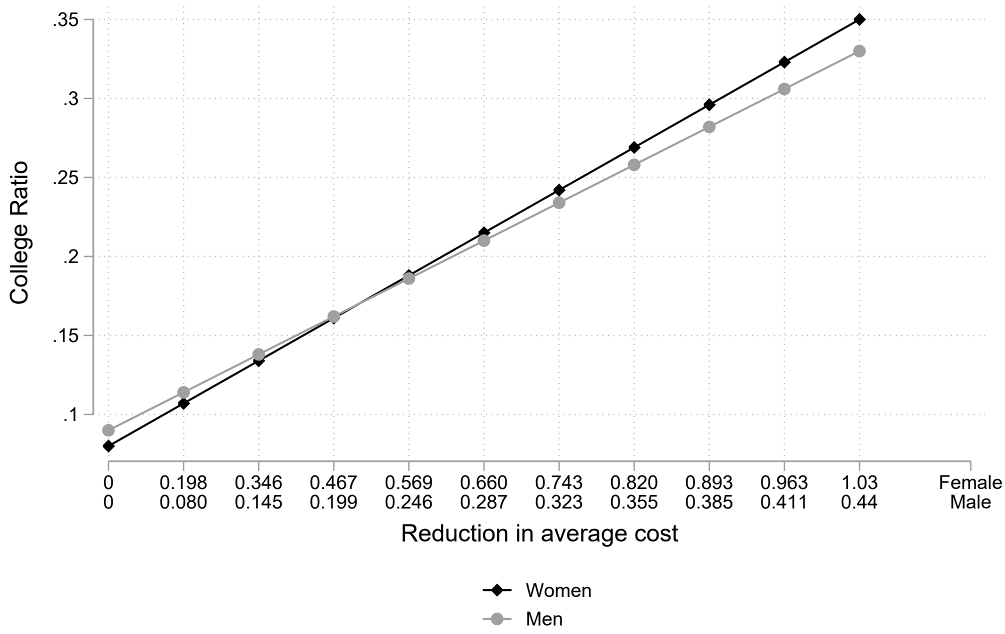
This subsection simulates the full model by changing the education cost parameters. To fully describe individual choices, we need to know the values of preferences parameters α_{xy} and γ_{xy}. We have identified the joint surplus from the empirical matching functions. However, we cannot separately identify these specific preference parameters. Without loss of generality, we make the following normalization:
(1) $\alpha_{x0} = \gamma_{0y} = 0$;[6] (2) $\alpha_{xy} = \gamma_{xy}$.[7]

We can then pin down the expected marital returns to education U^H, U^L, V^H, V^L (Equation C2, C3, C5 & C6). The change in average education cost parameters (c_f, c_m) are set to be consistent with the education distributions in our baseline simulation exercise (Section **??**). Without loss of generality, we assume that the distributions of idiosyncratic the education cost for both men and women follow a standard normal distribution. In Figure C8, we report the change in average educational attainment by gender following the change in the cost parameters. Following the college expansion, the college attainment increases faster for women than men. Figure C9 further display the marriage rates of college women and college men as a function of the reduction in the average education cost parameters. Consistent with our simulated results in the paper, the college expansion has an overall positive effects on the marriage probabilities of both college women and college men.

[6]With this normalization, α_{xy} and γ_{xy} represent the net systematic surplus from marriage.

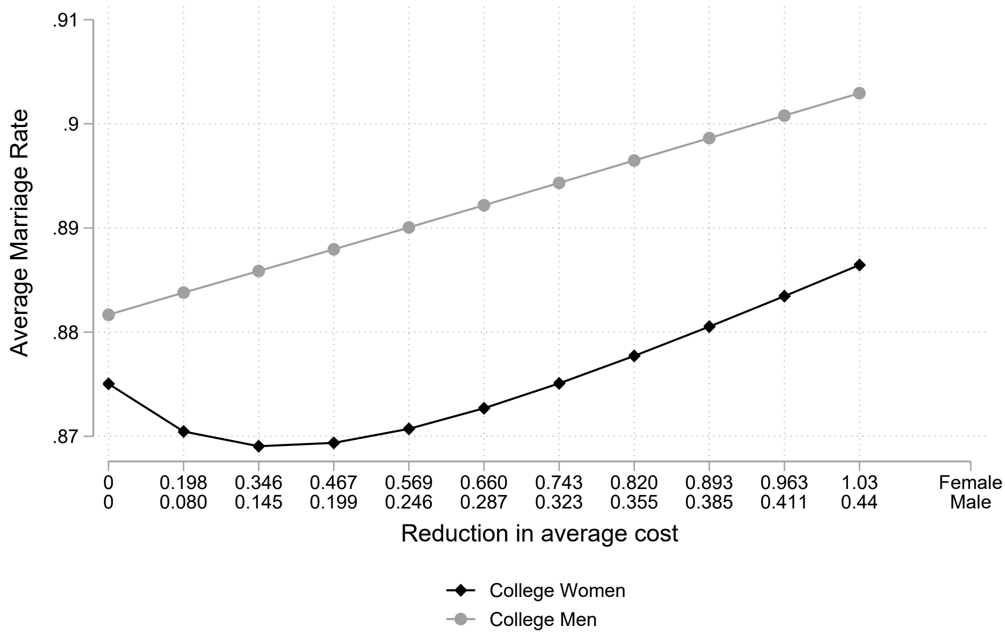
[7]Because we cannot separately identify $alpha$, γ, and τ, this assumption simply states that by default the wife and the husband share the joint surplus from marriage. The within-marriage transfer τ determines the actual division of the joint surplus.

Figure C8: Reduction in Education Cost and College Attainment: Simulated Results



Notes: The x-axis displays the reduction in the mean college education cost, which are allowed to evolve simultaneously for both women and men.

Figure C9: Reduction in Education Cost and Marriage Rate: Simulated Results



Notes: The x-axis displays the reduction in the mean college education cost, which are allowed to evolve simultaneously for both women and men.

D Additional Tests for the College Expansion Proxy

In this section, we run several additional tests that shows: (1) the college expansion proxy is relevant for increase in college enrollment following the college expansion, and (2) the college expansion proxy is exogenous to economic growth and local sex ratio.

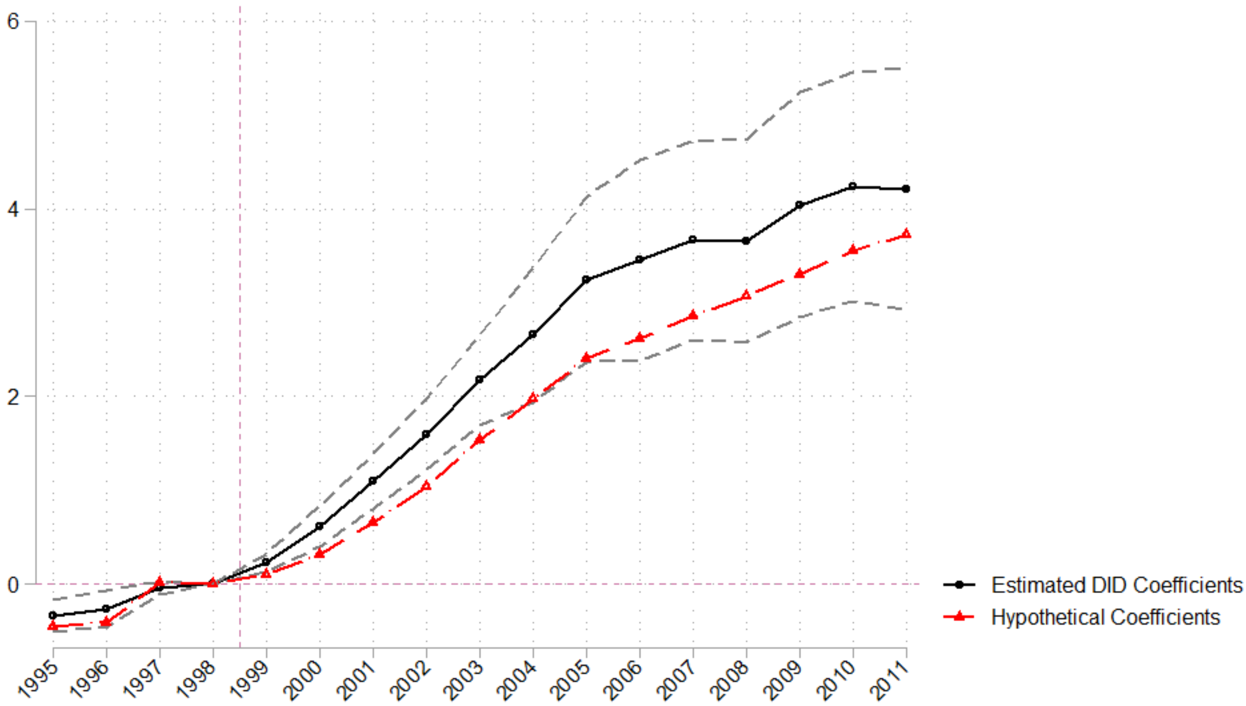
To examine whether our proxy predicts the local intensity of the college expansion, we run the following dynamic DID model:

$$(\frac{CollegeEnroll}{PopSize})_{pt} = \sum_{\tau} \alpha_{\tau} ExpProxy_p * (t = \tau) + \theta_p + \mu_t + \epsilon_{pt} \tag{D50}$$

The dependent variable is the fraction of college enrollment over population in prefecture (city) p. θ_p and μ_t stand for prefecture and year fixed effects, respectively. The data is at each prefecture-year level, which is collected from City Year Books. In Figure D10, we plot the dynamic coefficients on the expansion proxy. The results are consistent with the notion that the college expansion drastically increased college degrees in regions with a larger proxy. One unit of the estimated coefficient implies that for one more college student per capita in 1982, the college enrollment per capita in a prefecture will increase by one between 1998 and the given year in the x-axis. The triangle dots in Figure D10 are the hypothetical dynamic coefficients under the assumption that the college expansion has been perfectly proportional to the initial college enrollment in 1982.[8] The actual dynamic coefficients and the predicted coefficients closely trace each other, with the former being larger after the college expansion. This pattern seems to suggest that the regional inequality of higher education has been further amplified relative to the benchmark measured using the 1982 data.

[8] That is, we assume that the college enrollment in each prefecture is the product of the national enrollment in a given year and the fraction of the prefecture's enrollment to national enrollment in 1982.

Figure D10: Dynamic Effects of the College Expansion on College Enrollment



Notes: This figure plots the dynamic coefficients (α_τ) from EquationD50 and associated 95% confidence intervals. The red triangle dots are dynamic coefficients assuming that the college expansion has been perfectly proportional to the initial college enrollment in 1982.

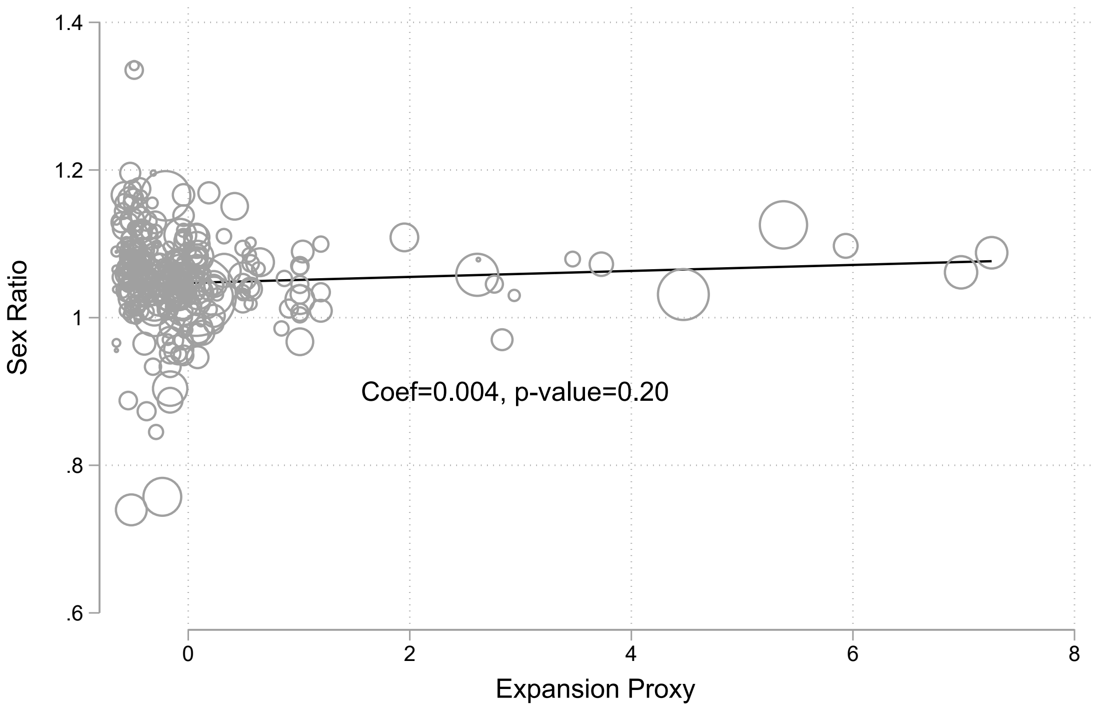
In order for our DID strategy to plausibly identify the causal impacts of the expansion, the expansion proxy must not be associated with other omitted factors that affect the marriage market. One concern is that the expansion proxy might be correlated with changes in regional economic performance. That potential correlation would bias our results if income directly affects marriage outcomes (Burgess et al., 2003; Chu et al., 2018; Hankins and Hoekstra, 2011). In Figure D11, however, we show that the expansion proxy is not systematically associated with the growth of GDP per capita in subsequent years. The growing confidence interval is also consistent with the notion that regional economic inequality has been enlarged in China. Nevertheless, the increasing regional inequality is not associated with our proxy for the college expansion.

Figure D11: Dynamic Effects of the College Expansion on GDP per capita

Notes: The dependent variable is the log GDP per capita of each prefecture obtained from City Statistical Yearbooks. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit is equal to one SD of the treatment proxy across prefectures. This figure plots the dynamic coefficients and the 95% confidence intervals.

Another factor is the sex ratio, which is potentially important given that sex imbalance has become increasingly severe and has affected China's marriage market in various ways (Wei and Zhang, 2011; Ebenstein and Sharygin, 2009). We test the correlation between the expansion proxy and local sex ratio for the cohorts of birth between 1970–1990 in Figure D12. Based on the univariate regression between sex ratio and the expansion proxy, we find no evidence that the proxy is significantly associated with sex ratio.

Figure D12: Correlation between College Expansion and Sex Ratio



Notes: The dependent variable is sex ratio of cohorts of birth between 1970 and 1990. Data is from the micro-data of 2000 China census obtained via IPUMBS International. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit is equal to one SD of the treatment proxy across prefectures. Each dot represents a prefecture weighted using population size. The solid line is fitted line of the univariate linear regression. P-values are computed based on robust standard errors.

E Predicting Regular College Degrees

E.1 Regular college vs. other post-secondary credentials

There is one specific form of measurement errors in the census data used for this research: we do not distinguish between regular college students and other post-secondary credentials that serve adults older than regular college ages in the census data. As the college expansion was all about the regular college system, introducing this form of measurement errors in our analysis might confound our estimates. Intuitively, this would tend to attenuate our estimates if the measurement errors in classifying the relevant group are not correlated with the treatment of the policy. To further rule out the confounding effects, we use a prediction exercise to address this concern. We predict the status of regular college graduates using information available in another smaller dataset where we can distinguish different types of college graduates.

We resort to the China Family Panel Studies (CFPS). The CFPS is a representative household survey of China. Among rich information about individuals and their families, the CFPS contains detailed information on educational attainment including whether the college degree comes from a regular college or the post-secondary credential system. We use the education and demographic information to predict regular college degree status, and then apply that prediction model to census data to obtain a sample of individuals who are (more likely) regular college system graduates. The rationale of this exercise is that it could reduce the measurement errors in college types by picking a group of individuals who are much more likely regular college graduates.

E.2 The Prediction Model

To apply the prediction model to census data, we need information shared by the CFPS and the census data. The common variables include: year of birth, province of residence, gender, ethnics, rural-urban residency registration (*Hukou*) status,

in the CFPS data, the sample size is very small for people with a college degree in particular. Therefore, for a given cohort-of-birth-by-province cell, there are very likely very few or even no observations. To deal with that, we group the years of birth into cohort groups of birth in the following way:

In the first step, we estimate the following Logit model:

$$\begin{aligned} Prob[Regular_i = 1] = F(&\beta_0 + \beta_1 Male_i + \beta_2 Urban_i + \beta_3 Han_i + \gamma_b + \eta_{Prov} + \gamma_b * \eta_{Prov} \\ &+ (Male_i * \gamma_b)\Pi_1 + (Han_i * \gamma_b)\Pi_2 + (Urban_i * \gamma_b)\Pi_3) \end{aligned}$$

$Male_i$, $Urban_i$, and Han_i are dummy variables for being male, having urban residency, being of the Han ethnicity. γ_b represents dummies for the cohort of birth, η_{Prov} is dummies for the province of residence. We also include the interactions between the first three demographic variables and cohort dummies as well as the interactions between cohort dummies and province dummies. The model is estimated using the CFPS data, where we can observe whether a respondent holds a regular college degree or just a post-secondary credential if they reports having a post-secondary degree. The estimated model coefficients are then applied to respondents who have a post-secondary degree in the census data. If the predicted probability is above 50%,

E.3 Internal Accuracy in the CFPS Data

We first evaluate the accuracy of the prediction model when it is applied to the original CFPS data (the dataset used to "train" the prediction model). We use two measures that are typically used to assess machine learning models: the *precision* rate and the *recall* rate. The precision rate captures, among those who are predicted to have a regular degree, what fraction is truly regular college graduates. The recall rate is the fraction of regular college graduates who are predicted as so. Table E4 reports these numbers: the precision rate is 78.8% and the recall rate is 92.5%. For our empirical analysis, this implies that when we look at the predicted regular college graduates, the vast majority of them are indeed regular college graduates.

We further assess the performance of this prediction model using a K-fold cross validation. Specifically, we divide the sample into five random sub-samples. Each time, we use four of them as the training dataset for the prediction model, and evaluate the accuracy of the prediction using the remaining sub-sample as the test dataset. As shown in Table E4, the average precision rate is 76.3%, and the average recall rate is 88.2%. Although we only include in the predict model a limited set of basic demographic information, it seems to predict the regular college degree status pretty accurately.

Table E4: Internal Accuracy in the CFPS Data

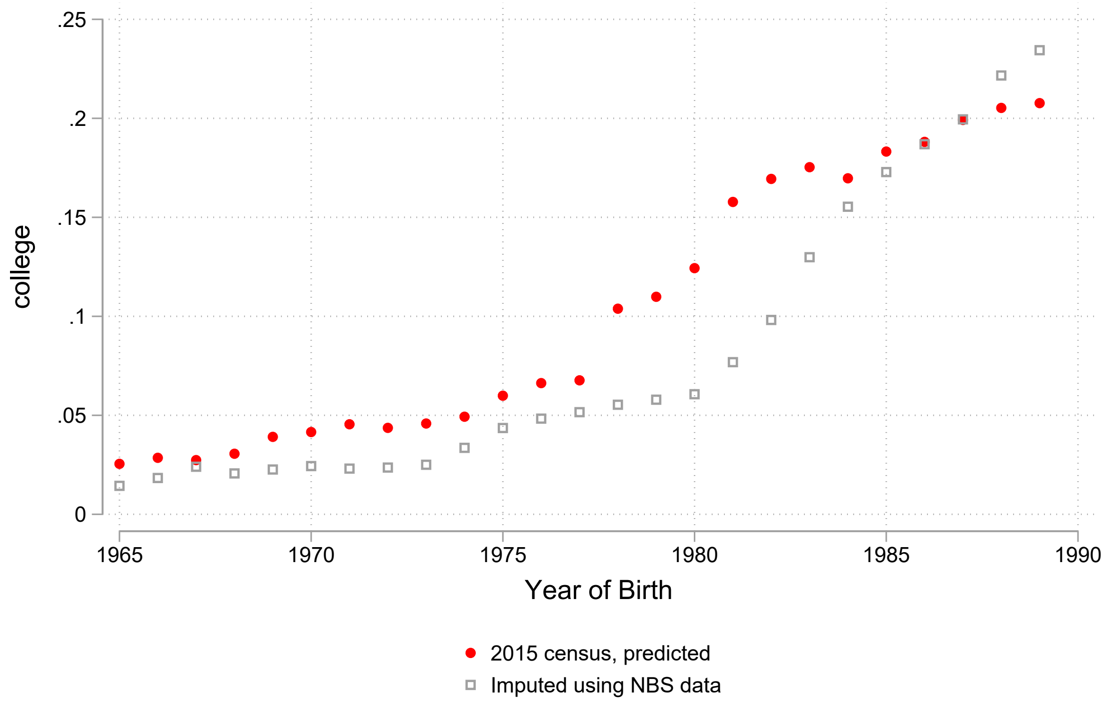
	Full Sample	5-Fold Cross-validation
Precision Rate	77.8%	76.5%
Recall Rate	93.7%	89.2%

Note: The precision rate is the ratio of actual regular college graduates in the predicted sample. The recall rate is the ratio of predicted regular college graduates in the sample of actual regular college graduates.

E.4 External Accuracy in the Census Data

We cannot directly test the accuracy of this prediction model in the census data. To provide some suggestive evidence on how accurately the prediction fits the actual data, we first compare the average regular college ratio of the predicted sample to the actual college ratio that we can gauge using public data. For each cohort of birth, we impute the ratio of college enrollment using public information on the number of births and the number of college enrollment, assuming that everyone goes to college at the age of 18. We then compare this ratio to the predicted regular college ratio by the cohort of birth. The results are reported in Figure E13.

Figure E13: Compare Predicted and Actual College Enrollment Rates



Note: *2015 census, predicted* is the regular college enrollment ratio based on the predicted results. *Imputed using NBS data* is the regular college enrollment ratio calculated using yearly new enrollment data and birth cohort size published by the National Bureau of Statistics of China.

In the second exercise, we evaluate the performance of this model in the 2000 census micro-data which contains information on whether a college degree is regular or just a post-secondary credential. Applying the prediction model parameters estimated with the CFPS data to the 2000 census, we are able to compare the predicted regular degree to the actual regular degree. The results are reported in Table E5. We can still obtain a meaningfully large precision rate.

Table E5: External Accuracy in 2000 Census

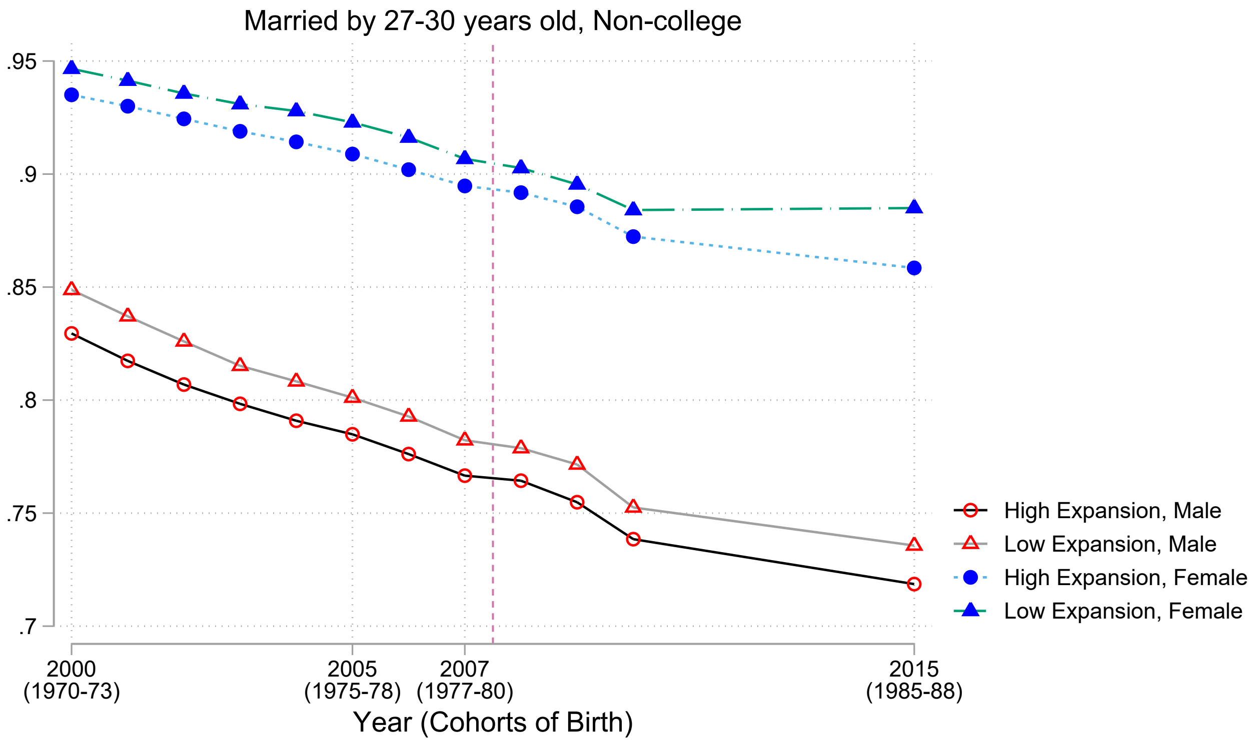
	2000 Census College Subsample
Precision Rate	73.9%
Recall Rate	69.2%

F Additional Results

In this section, we report additional empirical results that supplement main results presented in the paper.

F.1 Parallel Pre-trends for Non-college Groups

Figure F14: Trend in Marriage Rates of College Graduates by College Expansion Intensity



Notes: The sample is divided into two groups based on the value of the expansion treatment proxy below or above the median. The marital history information in the 2010 census allows us to impute the ever-married fractions before 2010. Using the 2015 mini-census, however, we can only know people's marital status as of 2015.

Table F6: Heterogeneity by College Types: Four-year Universities vs. Three-year Colleges

Dependent variable: Ever being married				
Post-expansion cohorts	1985-88, 27-30 years old in 2015		1981-84, 31-34 years old in 2015	
Pre-expansion cohorts	1975-78, 27-30 years old in 2005		1975-78, 31-34 years old in 2009	
	Four-year	Three-year	Four-year	Three-year
A. Male	(1)	(2)	(3)	(4)
ExpProxy∗Post	0.0277^{***}	0.0235^{***}	0.0161^{***}	0.00708^{**}
	(0.00469)	(0.00521)	(0.00465)	(0.00286)
Observations	18153	22043	17377	19109
Marriage rate of				
Pre-expansion cohorts	0.587	0.688	0.874	0.890
Post-expansion cohorts	0.575	0.647	0.857	0.860
B. Female				
ExpProxy∗Post	0.0198^{***}	0.0146^{***}	0.00701	0.00217
	(0.00396)	(0.00401)	(0.00524)	(0.00274)
Observations	17375	21102	15911	17983
Marriage rate of				
Pre-expansion cohorts	0.745	0.816	0.906	0.928
Post-expansion cohorts	0.667	0.749	0.887	0.899

Notes: The marital outcome of pre-expansion cohorts is constructed using marriage history so that it is comparable to the post-expansion cohorts. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit represents one standard deviation across all prefectures. All regressions control for prefecture fixed effects and age fixed effects. Standard errors clustered at the prefecture level are in parentheses. There are 340 clusters. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

F.2 Additional Robustness Checks

Table F7: Effects of College Expansion on Marriage Probability: Treatment Proxy using 1990 Census

Dependent variable: Ever being married.				
Post-expansion cohorts	1985-88, 27-30 years old in 2015		1981-84, 31-34 years old in 2015	
Pre-expansion cohorts	1975-78, 27-30 years old in 2005		1975-78, 31-34 years old in 2009	
	College	Non-college	College	Non-college
A. Male	(1)	(2)	(3)	(4)
ExpProxy$*$Post	0.0220***	-0.0009	0.0132***	-0.0006
	(0.0080)	(0.0028)	(0.0045)	(0.0017)
	[0.0273]	[0.0059]	[0.0124]	[0.0008]
Observations	40196	187259	36486	181105
Marriage rate of				
Pre-expansion cohorts	0.644	0.793	0.883	0.882
Post-expansion cohorts	0.613	0.728	0.858	0.874
B. Female				
ExpProxy$*$Post	0.0139**	-0.0008	0.0062	-0.0007
	(0.0056)	(0.0034)	(0.0041)	(0.0012)
	[0.0171]	[0.0020]	[0.0053]	[-0.0009]
Observations	38477	182842	33894	176213
Marriage rate of				
Pre-expansion cohorts	0.787	0.916	0.919	0.961
Post-expansion cohorts	0.709	0.873	0.892	0.953

The marital outcome of pre-expansion cohorts is constructed using marriage history so that it is comparable to the post-expansion cohorts. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit represents one standard deviation across all prefectures. All regressions control for prefecture fixed effects and age fixed effects. Standard errors clustered at the prefecture level are in parentheses. There are 340 clusters. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Baseline estimates are in brackets.

Table F8: Results Controlling for Baseline City Characteristics

Dependent variable: Ever being married.				
Post-expansion cohorts	1985-88, 27-30 years old in 2015		1981-84, 31-34 years old in 2015	
Pre-expansion cohorts	1975-78, 27-30 years old in 2005		1975-78, 31-34 years old in 2009	
	College	Non-college	College	Non-college
A. Male	(1)	(2)	(3)	(4)
ExpProxy∗Post	0.0299***	0.00806***	0.0122***	0.00108
	(0.00552)	(0.00224)	(0.00357)	(0.00146)
	[0.0273]	[0.0059]	[0.0124]	[0.0008]
Observations	36988	160626	33678	154910
Marriage rate of				
Pre-expansion cohorts	0.641	0.793	0.882	0.882
Post-expansion cohorts	0.615	0.728	0.858	0.874
B. Female				
ExpProxy∗Post	0.0176***	0.00239	0.00414	-0.000987
	(0.00382)	(0.00314)	(0.00342)	(0.00124)
	[0.0171]	[0.0020]	[0.0053]	[-0.0009]
Observations	35561	157529	31519	151635
Marriage rate of				
Pre-expansion cohorts	0.784	0.916	0.916	0.961
Post-expansion cohorts	0.709	0.873	0.892	0.953

Notes: The marital outcome of control cohorts is constructed using marriage history so that it is comparable to the treatment cohorts. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit represents one standard deviation across all prefectures. City level control variables include log GDP per capita and sex ratio, both measured in 2000 and interacted with $Post$ dummy. All regressions control for prefecture and age fixed effects. Standard errors clustered at the prefecture level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Baseline estimates are in brackets.

Table F9: Results Controlling Province by Year Fixed Effects

Dependent variable: Ever being married				
Post-expansion cohorts	1985-88, 27-30 years old in 2015		1981-84, 31-34 years old in 2015	
Pre-expansion cohorts	1975-78, 27-30 years old in 2005		1975-78, 31-34 years old in 2009	
	College	Non-college	College	Non-college
A. Male	(1)	(2)	(3)	(4)
ExpProxy∗Post	0.0239***	0.00717***	0.0069***	0.0033**
	(0.00448)	(0.00253)	(0.00222)	(0.00153)
	[0.0273]	[0.0059]	[0.0124]	[0.0008]
Observations	40196	187259	36486	181105
Marriage rate of				
Pre-expansion cohorts	0.644	0.793	0.883	0.882
Post-expansion cohorts	0.613	0.728	0.858	0.874
B. Female				
ExpProxy∗Post	0.0216***	0.00149	-0.00064	-0.00188
	(0.00305)	(0.00145)	(0.00269)	(0.00191)
	[0.0171]	[0.0020]	[0.0053]	[-0.0009]
Observations	38477	182842	33894	176213
Marriage rate of				
Pre-expansion cohorts	0.787	0.916	0.919	0.961
Post-expansion cohorts	0.709	0.873	0.892	0.953

The marital outcome of control cohorts is constructed using marriage history so that it is comparable to the treatment cohorts. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit represents one standard deviation across all prefectures. All regressions control for province fixed effects interacted with the $Post$ dummy. All regressions control for prefecture and age fixed effects. Standard errors clustered at the prefecture level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

F.3 Assortative Mating

Table F10: Effects of the College Expansion on Assortative Mating: Four-year Universities vs. Below

		Post-expansion: 1985-88		Post-expansion: 1981-84	
	Region by Expansion Intensity	High	Low	High	Low
Cohorts	Pre-expansion: 1975-78	0.032	0.014	0.040	0.016
	Post-expansion	0.075	0.024	0.094	0.029
	Diff-in-diff	**0.033*****		**0.040*****	
		(0.002)		(0.002)	

Notes: The index is calculated using the sample of married couples in the random samples of 2010 census and 2015 mini-census. The sample include married couples with either side falling in the specified cohort range. Ages of couples in pre-expansion cohorts are adjusted to be comparable to those in the post-expansion cohorts. The national sample is divided into high vs. low regions based on the treatment proxy being above or below median. Standard errors in parentheses are estimated by boostrapping from the original sample 1000 times. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table F11: Effects of the College Expansion on Assortative Mating: Various Indexes

		Log Odds Ratio	Minimum Distance	Correlation	Absolute Difference
A. Post-expansion cohorts: 1985–1988, Pre-expansion cohorts: 1975–78					
1975–78	High	4.347	0.774	0.707	0.082
	Low	4.623	0.769	0.679	0.050
1985–88	High	4.030	0.757	0.741	0.149
	Low	4.264	0.744	0.693	0.075
Diff-in-diff		**0.042**	**0.007**	**0.020***	**0.041*****
		(0.086)	(0.013)	(0.011)	(0.002)
B. Post-expansion cohorts: 1981–1984, Pre-expansion cohorts: 1975–78					
	High	4.308	0.780	0.719	0.095
	Low	4.524	0.764	0.678	0.054
1981–84	High	4.202	0.784	0.762	0.158
1975–78	Low	4.488	0.768	0.715	0.075
Diff-in-diff		**-0.071**	**0.001**	**0.007**	**0.042*****
		(0.083)	(0.012)	(0.010)	(0.002)

Notes: See Table 12 for the definition of different indexes. The indexes are calculated using the sample of married couples in the random samples of 2010 census and 2015 mini-census. The sample include married couples with either side falling in the specified cohort range. Ages of couples in pre-expansion cohorts are adjusted to be comparable to those in the post-expansion cohorts. The national sample is divided into high vs. low regions based on the treatment proxy being above or below median. Standard errors in parentheses are estimated by boostrapping from the original sample 1000 times. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

F.4 Marriage Age Gap for Non-college Groups

In Table F12, we show that the college expansion also reduced the marriage age gap of non-college women and non-college men. The magnitude of these effects is small. For example, Column (1) shows that a SD higher treatment proxy caused 0.06 years drop in the marriage age gap of non-college women. Nevertheless, this is unlike the mostly null effects on their marriage probabilities. Our model, which does not incorporate matching on age (and the interaction between education and age in matching), is unable to provide a theoretical guidance. We discuss some possible intuitions here.

Assume non-college women have preferences for both younger age (physical desirability) and higher income (or human capital). Before the college expansion, an older age is possibly a proxy for a higher income among the large pool of non-college men. Therefore, non-college women who have a stronger preferences for spouses' income may accept a larger age gap. Following the college expansion, those women with a stronger preferences for income are more likely to match with increasingly available college men. This is consistent with Column (2) of Table 9 that shows a small but positive increase in non-college women marrying college men. On the other hand, before the expansion, one useful strategy for non-college men that have a preference for younger wives might be to accumulate more human capital and wait longer before they marry. This tends to create a larger marriage age gap. However, following the college expansion, with non-college women that prefer a higher income increasingly marry college men, the marginal benefit of waiting longer for human capital accumulation depreciates for non-college men. As a result, they choose to marry earlier and accept a smaller age gap. This story might explain our findings. However, it is just a conjecture that is not formally test either theoretically or empirically. It can be a fruitful future line of research to dig into the interaction between (college) education and age in the marriage market.

Table F12: Effects of the College Expansion on Marriage Age Gap of Non-college Groups

	(1)	(2)	(3)	(4)
Dependent variable: Age gap (Husband - Wife)				
Post-expansion cohorts	1985-88, 27-30 years old in 2015		1981-84, 31-34 years old in 2015	
Pre-expansion cohorts	1975-78, 27-30 years old in 2005		1975-78, 31-34 years old in 2009	
	Female	Male	Female	Male
ExpProxy∗Post	-0.0586**	-0.0344**	-0.0504*	-0.0373*
	(0.0232)	(0.0135)	(0.0271)	(0.0221)
Observations	123683	105990	128557	117749
Average Age Gap				
Pre-expansion cohorts	2.143	1.296	2.121	1.566
Post-expansion cohorts	2.294	1.082	2.264	1.626

Notes: The marital outcome of pre-expansion cohorts is constructed using marriage history so that it is comparable to the post-expansion cohorts. $ExpProxy$ is the proxy for college expansion, which is standardized so that one unit represents one standard deviation across all prefectures. All regressions control for prefecture fixed effects and age fixed effects. Standard errors clustered at the prefecture level are in parentheses. There are 340 clusters. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

G Additional Discussions on Assortative Mating Index

We start by summarzing whether these indices satisfy the properties proposed in (Chiappori et al., 2021).

Properties	Absolute Difference	Log Odds Ratio	Min Distance	Correlation
Scale invariance	Y	Y	Y	Y
Symmetry	Y	Y	Y	Y
Monotonicity	Y	Y	Y	Y
Perfect PAM	N	Y	Y	Y

The first two properties are intuitive and we can easily observe that the two indices satisfy them.[9]

Monotonicity means that, when we hold the identical marginal distributions of college women ($\frac{k_1}{k_1+k_2}$) and college men $\frac{k_1}{k_1+k_3}$, then adding more people in each diagonal cell ($k_1$or k_4) will always increase assortativeness. The two indices both satisfy this property.

G.1 Perfect PAM

A more complicated issue arises with the Perfect PAM. This property basically states that a contingency table T_{PAM} under perfect assortativeness (meaning that $k_2 = 0$ or $k_3 = 0$) should achieve the maximum value of the index and no other table under imperfect assortativeness ($k_2 * k_3 > 0$) should lead to an index value than T_{PAM}.

They proposed two versions of PAM: the strong version means the stated property hold as long as one of $\{k_2, k_3\}$ is 0, while the weak version says the property holds when $k_2 = k_3 = 0$. The weak version is therefore a necessary condition for the strong version.

The log odds ratio satisfies both strong and weak PAM condition because the ratio is $+\infty$ for a contingency table with perfect assortativeness. The absolute difference measure, however, does not. Let's consider the case when $k_2 = k_3 = 0$, we have

$$\begin{aligned} AbsDiff &= \frac{k_1k_4 - k_2k_3}{K^2} \\ &= \frac{k_1(K - k_1)}{K^2} \end{aligned}$$

which achieves a maximal value of 0.25 when $k_1 = 0.5K$ (the educated and the uneducated are balanced) but goes to 0 when k_1 approaches 0 or 1. Therefore, we can easily construct a table whose index value is greater than the index of a perfect assortativeness table, as shown in Table G13.

[9]Symmetry means that if we "re-label" college as non-college and non-college as college, the resulted index should not change.

Table G13: Explore the Perfect PAM property of the Absolute Difference Index

Wife-Husband Type	T1: perfect assortativeness	T2
C-C(k_1)	0.02	0.2
C-NC(k_2)	0	0.05
NC-C(k_3)	0	0.05
NC-NC(k_4)	0.98	0.7
Absolute Diff.	0.0196	0.1375

T1 and T2 represent two matching contingency tables. T1 exhibits perfect assortativeness, while the matching in T2 is not perfect but the college ratio is much higher than that in T1.