DISCUSSION PAPER SERIES

DP16035

What Do Parents Want? Parental Spousal Preferences in China

Eva Raiber, Weiwei Ren, Jeanne Bovet, Paul Seabright and Charlotte Wang

DEVELOPMENT ECONOMICS

CEPR

What Do Parents Want? Parental Spousal Preferences in China

Eva Raiber, Weiwei Ren, Jeanne Bovet, Paul Seabright and Charlotte Wang

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Development Economics

What Do Parents Want? Parental Spousal Preferences in China

Abstract

In many societies, parents are involved in selecting a spouse for their child, and integrate this with decisions about migration and educational investment. What type of spouse do parents want for their children? We estimate parents' spousal preferences based on survey choices between random profiles. Preference data are elicited from parents or other relatives who actively search for a spouse on behalf of their adult child in Kunming, China. Economic variables (income and real estate ownership) are important for the choice of sons-in-law, but not daughters-in-law. Education is valued on both sides. We simulate marriage outcomes based on preferences for age and education and compare them with marriage patterns in the general population. Homogamy by education can be explained by parental preferences, but not by age: parents prefer younger wives, yet most couples are the same age. Additionally collected preference data from students can explain age distributions. Survey data from parents suggest that while they prefer younger wives, they also accept wives of the same age. Overall, marriage markets have a likely positive influence on education investments for both boys and girls.

Eva Raiber - eva.raiber@univ-amu.fr
Aix-Marseille School of Economics

Weiwei Ren - waterysun@163.com
Yunnan Normal University, China

Jeanne Bovet - jeanne.bovet@gmail.com
Northumbria University, Newcastle-upon-Tyne, UK.

Paul Seabright - paul.seabright@tse-fr.eu
Toulouse School of Economics and CEPR

Charlotte Wang - charlotte.wang@ipag.fr
IPAG Business School, Paris, France

What Do Parents Want? Parental Spousal Preferences in China[*]

Eva Raiber[1], Weiwei Ren[2], Jeanne Bovet[3], Paul Seabright[4,5], and Charlotte Wang[6]

[1]*Aix Marseille School of Economics, Aix-Marseille University, France* , [2]*Yunnan Normal University, China,* [3]*Northumbria University, Newcastle upon Tyne, UK,* [4]*Toulouse School of Economics, University of Toulouse, France,* [5]*Institute for Advanced Study in Toulouse, France,* [6]*IPAG Business School, France*

April 12, 2021

Abstract

In many societies, parents are involved in selecting a spouse for their child, and integrate this with decisions about migration and educational investment. What type of spouse do parents want for their children? We estimate parents' spousal preferences based on survey choices between random profiles. Preference data are elicited from parents or other relatives who actively search for a spouse on behalf of their adult child in Kunming, China. Economic variables (income and real estate ownership) are important for the choice of sons-in-law, but not daughters-in-law. Education is valued on both sides. We simulate marriage outcomes based on preferences for age and education and compare them with marriage patterns in the general population. Homogamy by education can be explained by parental preferences, but not by age: parents prefer younger wives, yet most couples are the same age. Additionally collected preference data from students can explain age distributions. Survey data from parents suggest that while they prefer younger wives, they also accept wives of the same age. Overall, marriage markets have a likely positive influence on education investments for both boys and girls.

Keywords: Marriage, Preference estimation, China, Parental matchmaking, Matching

JEL codes: J12, I26, D10

1

1 Introduction

What do people look for in a marriage partner? Partner choice has important consequences for most life dimensions such as socioeconomic status, fertility, and overall well-being. In many contemporary societies, and most historical ones, parents are involved in the crucial process of selecting a spouse for their child. Even in societies where parents have no explicit involvement, parental support for their adult children's choices can make an important difference to the quality and durability of a marriage (Reczek et al. [2010]). In Chinese society, where marriage is extremely important, 33% of couples that married between 1980 and 2014 were introduced to each other by relatives.[1]

While estimating partner preferences has gained momentum in the economic literature (e.g. Belot and Francesconi 2006; Fisman et al. 2006; Hitsch, Hortaçsu, and Ariely 2010; Kurzban and Weeden 2007), parental preferences have mostly been neglected due to the focus on Western countries.[2] In the Chinese context, the literature mostly covers stated individual preferences and data from online platforms (e.g. Ong and Wang 2015; Xia et al. 2014). In this paper, we investigate elicited parental preferences for spousal characteristics.

Parents in all countries are the main decision-makers when it comes to primary and secondary education, and they heavily influence tertiary education and other pre-marital investment such as acquiring real estate. These decisions may also take into account how the investments change the characteristics of their child's expected spouse (Chiappori et al. [2009]). Particularly in poor countries, these may depend on social factors such as parentally-influenced norms governing age and type of marriage (Maertens [2013]), and may have effects that persist across generations (Delprato et al. [2017]).

Parental spousal preferences of sons might therefore not only influence the type of wife their son marries but also the expected marriage market returns of daughters' pre-marital investments, and vice versa. It remains an important question whether parental influences, particularly on the choice of daughters-in-law, encourage or discourage educational investments in boy and especially in girls.

In China, recent data from the China Family Panel Study (CFPS) shows that marriages are highly assortative on age and education. Assortative matches are characterized as those in which spouses have the same educational level or similar age. Yet it is not possible, without making strong assumptions, to derive spousal preferences solely from these marriage outcomes. Both homogamic preferences - preferences for those of the same type - as well as increasing preferences - preferences for those of the "best type" - can explain assortative matches.[3] Assortative matches can also arise if there are no specific preferences but individuals with similar characteristics have a higher probability of meeting. These different options have different implications for the marriage market returns to education and the timing of marriage, and it is therefore valuable to obtain direct information on parental preferences.

---

[1] China Family Panel Study 2016, see notes for figure 6.

[2] Notable exceptions are Adams and Andrew [2019] and Banerjee et al. [2013] who focus on India.

[3] When educational levels are similar between men and women, increasing preferences lead to assortative matches. The highest type of man would then match with the highest type of woman, the second highest type with the second highest type and so forth. This holds under the assumption of non-transferable utility, which we make explicitly later in the paper, or alternatively under the assumption of a supermodular marriage surplus.

Parental spousal preferences are the most relevant when their child is of marriageable age and when they are involved in the search for their child's spouse (even if involvement means only to comment on the options). We interviewed men and women who were searching a spouse for their adult child or relative at a public park in Kunming, in Southern China. The phenomenon of parents and other relatives searching for potential suitors in public parks at so-called "marriage markets" is nowadays common in Chinese cities. Their existence underlines the continuing involvement of parents in marriage decisions.

Parents see themselves as agents for their children, complementing their child's search effort in half of the cases and often having their approval or even encouragement. They often spend substantial time on the search. Respondents were asked to evaluate a series of randomly created hypothetical profiles that we use to estimate spousal preferences. Profiles include information about income, education, age, ethnicity and real estate ownership. They mimic the information that parents usually exchange at the public park. Furthermore, we ask respondents what educational levels and age they deem acceptable.

We find that parents dislike profiles with an educational level that is lower than their daughter's or their tertiary-educated son's. Parents searching on behalf of tertiary-educated women have a clear preference for profiles with higher education. We find little evidence for a dislike of women who are more educated than the man. High income and real estate ownership are important for parents who search on behalf of their daughter, but not for those who search on behalf of their son. Parents searching on behalf of their daughter prefer profiles with a similar age. Parents searching on behalf of their son prefer younger women. These preference parameters can be used to provide estimates for marriage market returns to education and for the trade-off between marrying early or delaying marriage to continue with education. [4]

Our data set is of interest for several reasons in addition to having a unique set of parents as respondents. In contrast to data from online dating websites, we can credibly claim that parents are looking for a spouse for their adult child and not just for a short-term relationship. Physical presence at the park can be seen as a signal for a serious search effort. Most respondents have been at the park more than once, and questionnaire responses indicate that the search for a spouse has been discussed within the family. Thus, we are sure that respondents have thought about what they are looking for and are experienced in selecting potential candidates.

Also, because choice data are taken from hypothetical profiles which are randomly created, we do not have the issue of a selected pool to choose from and we can ensure that the characteristics presented to subjects are not directly correlated. Since we focus on economic variables and do not include physical attractiveness, the results can be easily compared to outcome data from other data sets. The trade-off between income and physical attractiveness is investigated in a companion paper with a separate profile task .[5]

---

[4]We also find that parents have a preference for male profiles with majority Han ethnicity, but not for female profiles. The sample is majorly urban and from the Han majority.

[5]Reference is omitted to ensure anonymity. The profile task used in companion paper took place after the hypothetical choice task used in this paper. The results are in line with the results here: respondents value income only when they are searching on behalf of a female subject. When they are searching on behalf of a male subject, income is insignificant.

Furthermore, we ask if parents get what they want: do observed marriage outcomes align with parents' preference? As we could not contact the respondents again because of our assurances of anonymity, we use a simulation approach and compare simulated marriages with actual marriages in the general population. Based on parents' estimated preferences for age and education, we simulate marriages with a standard Gale-Shapley algorithm (Gale and Shapley 1962). For the supply of spouses, we use individuals who recently married from the CFPS 2014 and 2016 which contains information on age and education for both spouses.

The baseline specification of the simulations predicts educational assortativity well. As the simulation ignores search friction in the marriage sorting process, only uses average preferences, and assumes non-transferable utility, the degree of assortativity on education is even higher in the simulation than in the actual outcomes. Yet, we can observe that parents' dislike of sons-in-law that are less educated than their daughter decreases the proportion of matches where this happens. In the observed distribution, matches in which the wife is slightly more educated than the husband are as common as matches in which the husband is slightly more educated.

For the age distribution, we observe a discrepancy between the simulation results and the actual outcomes: The simulations predict the most common case to be couples where the husband is 1 to 3 years older. In the real distribution, the most common case is the spouses having the same age. The difference between simulation based on parental age preferences and real outcome could be explained by changes in age preferences when unmarried men get older. Yet, allowing for age-specific preferences in the simulation does not improve the fit of the simulation. Weighing the general population to make it similar to the male population represented at the park does not improve the fit either.

The discrepancy could also reflect a difference in age preferences between parents and their children. We contrast parental preferences with preferences estimated based on a local student sample. Indeed, while parents prefer a younger wife, male students do not have the same preferences for a younger partner and simulation results map well onto the observed age patterns. Finally, survey data suggest that while parents prefer a younger wife for their son, they accept wives that have the same age as their son until their son is over 30. As most marriages for between individuals in their mid-20, most marriages are within parents' accepted limits, but reflect young men's age preferences more than the parents' in our sample.

1.1 Links with the literature

This paper adds to the growing literature that studies marriage preferences theoretically and empirically. The economic literature on marriage is based on the seminal work on marriage markets by Becker [1973] who models a market with two sides (men and women) and assumes transferable utility. Agents form matches and bargain about the distribution of the surplus that is generated by the match. The functional form of this marriage surplus is of particular interest.[6] It reflects preferences of both sides. Following Choo and Siow [2006], several papers have estimated the marriage surplus which can be identified under relatively weak assumptions. A summary of can be found in

---

[6]The marriage surplus is mathematically defined as the utility created when two people get married minus the utilities of them stay single.

Chiappori and Salanié [2016], and even more recently in Chiappori [2020].[7]

However, without assumptions on the structural form of the marriage surplus, the bargaining and searching process, one cannot identify men and women's preferences separately. Recent studies have thus relied on additional preference data, for example from online dating websites. These papers usually assume non-transferable utility where men and women get a determined fixed utility from the match as it simplifies interpretation.[8] We follow this approach of identifying preferences directly from additional data. The empirical strategy is most closely related to Hitsch, Hortaçsu, and Ariely [2010] who use data from an online dating website and follow the searching framework of Adachi [2003]. Similarly, Banerjee et al. [2013] use rankings of responses to marriage advertisements in an Indian newspaper. Other papers use data from speed dating (Fisman et al. 2006; Fisman et al. 2008) or online dating (Belot and Francesconi 2006; Kurzban and Weeden 2007; Xia et al. 2014; Ong and Wang 2015). Adams and Andrew [2019] use hypothetical marriage scenarios to elicit parents' spousal preferences for education and marriage age and beliefs about marriage prospects.

This paper also contributes to the topic of parental pre-marital investment, parental involvement in marriage decisions and their influence on the marriage outcomes. In the Chinese context, Huang, Jin, and Xu [2012] find that couples that were introduced to each other by their parents or another relative have a higher cumulative income, but lower marital harmony. Huang, Jin, and Xu [2016] add that couples that rely on their parents for finding a spouse have more children.

Parents are the main decision-makers when it comes to early human capital investment. Their expectations on the returns to education are crucial for the educational decisions they make about their children. Studies that estimate the returns to education usually look at individual returns to education in the labor market. Giles et al. [2019] use the disruption in educational access due to the Cultural Revolution as an instrument for education and find that a college degree versus high school degree increases hourly wages by around 37%. Li et al. [2012] compare earnings between twins and find that a college degree increases earnings by around 40%. The uninstrumented and not corrected returns are even higher as they include the effects of unobserved ability or family background. If a college degree also increases the chance to marry a highly educated spouse, this increases the overall returns to education.

Our preference data and the simulations indeed suggest that marriage returns to education are high. A profile with a woman with a tertiary degree is around 9 percentage points more likely to be selected by the parents of tertiary educated man than a woman with a high school degree. A profile with a man with a tertiary degree is around 20 percentage points more likely to be selected by the parents of tertiary educated woman than a man with a high school degree. In the simulations without search frictions, 78% of couples have the same educational level (52% in the actual population).

There is a broader literature on marriage and human capital investment in developing countries outside China to which this paper contributes. Attanasio and Kaufmann [2017] confirm that Mexican students take marriage prospects into account in college enrolment decisions. Maertens [2013] show the importance of marriage age for educational investments in India, and Delprato

---

[7]Notable papers in this field are Chiappori, Salanié, and Weiss [2015], Choo and Seitz [2013], Galichon and Salanié [2015], Logan, Hoff, and Newton [2008] and Wong [2003].

[8]An exception is Del Boca and Flinn [2014] who combine data on marriage outcomes and household production.

et al. [2017] demonstrate the magnitude and persistence of these effects for sub-Saharan Africa. The Chinese experience is highly relevant to other countries that are seeking to develop integrated approaches to schooling policies taking marriage market dimensions into account.

Finally, this paper complements the literature on the changes in marriage patterns. Hu and Qian [2016] find that educational homogamy has increased over time, while Mu and Xie [2014] find that homogamy in age, as measured by the average age difference between the spouses, increased until the 1990s, but since then has decreased slightly. Hu [2016] finds that the household registration status (hukou) plays an important part in determining marriages. The household registration status is inherited from the parents and is an important determinant of socioeconomic status. Yu and Xie [2015] find an increase in the importance of economic prospects in urban China. Our study confirms that economic variables, such as income and real estate, are important characteristics in the marriage market, yet only for men. Education and age are characteristics taken into account on both sides.

2 Context

2.1 Parental Involvement in Marriage Decisions in China

The way to find a spouse in Chinese society changed substantially in the past century. For centuries, arranged marriage had been dominant. Parents chose the spouse for their child, often with the help of a professional matchmaker (Huang, Jin, and Xu 2016; Xia and Zhou 2003). After the Chinese Communist Party came to power the government passed the Marriage Law, adopted in 1950, making arranged marriage illegal. It was in line with its effort to get rid of traditional Chinese classes (Engel 1984). Moreover, the government helped to abolish the traditional marriage system by encouraging women to join the labor force (Pimentel 2000; Xia and Zhou 2003). However, in rural areas, arranged marriage continued to be important and parents still influenced marriage outcomes (Xia and Zhou 2003). For instance, Riley [1994] uses data from a survey collected in 1986/1987 and find that the number of arranged marriages has decreased. Yet, parents continue to have influence over marriage decisions. Huang et al. [2012] find that 12% of all urban couples interviewed in 1991 were introduced to each other by their parents or other relatives.

The economic reforms of the late 1970s substantially changed the life of Chinese people, as China became increasingly open to the rest of the world (Chang et al. 2011; Higgins et al. 2002). The economic reforms led to an increase in economic opportunities, inequality and mobility. Marriage again became a way to increase a family's social status and improve its financial situation (Fan and Huang 1998; Han, Li, and Zhao 2015), and Chinese parents continue to influence their children's marriage decision (Pimentel 2000). This process was accompanied by changes in marital preferences (Chang et al. 2011; Higgins and Sun 2007, Higgins et al. 2002).

Indeed, data from the China Family Panel Study 2010 shows that the share of first marriages that were arranged by parents dropped sharply in the late 1940s and early 1950s (illustrated in Fig. 5 in Appendix A). The share of those who were introduced to their spouse by a friend or relative increased at the same time, and continues to be the predominant way. The 2010 survey does not make a difference between being introduced by friends or relatives. The 2016 survey asks the same question for the current spouse to a subset of the respondents, and separates between those two options. In this subset, being introduced to each other by relatives was the predominant way to

meet the spouse until the mid 1990s, and continues to be among the most important ways to meet a spouse, together with being introduced by friends and meeting at work (illustrated in Fig. 6 in Appendix A).

2.2 Parental Search in Public Parks

Every Saturday, one corner of the Green Lake Park in Kunming, the capital and largest city of the province of Yunnan in South China, hosts a "marriage market".[9] These marriage platforms are a new phenomenon but already widely spread. The most famous marriage search platform in a public park is at the People's Park in Shanghai which started in 2004.[10]

In a dedicated area of the park in Kunming - accessible to the public - individuals search for a spouse for either themselves or for someone else. This marriage search platform was initiated by parents in around 2005 who used their weekends to chat to other parents with unmarried children. Over time, it developed into an established event. Parents and other participants talk to each other or post sheets of papers with basic information of their "unmarried subject" on the wall of the park, with their own contacts. They may check the information of others on the wall or address one of the marriage agencies present at the park.[11] Parents would then set up meetings where their unmarried subject could meet the potential spouse. They therefore do not arrange marriages but make a pre-selection of candidates.

2.3 Marriage Patterns in Contemporary China

We use the China Family Panel Study (CFPS) from 2014 and 2016 to describe current marriage patterns.[12] The CFPS is a nationally representative survey, which contains information on the age and educational level of both spouses. In 2016, 78.4% of the respondents between 20 and 50 years are married, 18.1% are single, 2.2% are divorced and 0.8% are widowed. For those between 40 and 50, 2% have never been married and 2.7% are divorced.

Recent marriages between 2012 and 2016 are assortative on the educational level (see Fig. 1 panel (a)). Educational levels are defined according to highest educational degree obtained: no degree (illiterate/semi-literate), primary school degree, middle school degree (also called junior high school), high school degree (also called senior high school), undergraduate degree, and graduate degree (master or Ph.D.). For nearly all educational levels, it is most common to marry someone who has the same educational level (the exception being individuals with a postgraduate degree, which is still rare). The second most likely case is that the spouse has an educational level that is one below or above. The Pearson correlation coefficient between the educational level of husband and wife is significant at 0.41 We can also see in panel (b) that the educational distribution of men and women are very similar for this period.

---

[9]With an estimated population of nearly 4 million (Cox 2018), Kunming is a middle size city in China.

[10]According to an article in Al Jazeera: https://www.aljazeera.com/indepth/inpictures/2013/04/201343113125739211.html.

[11]In Kunming, marriage agencies also advertised their services at the public park. However, our survey results suggest that their services are rarely taken up and their business models seem dubious.

[12]The CFPS is a large-scale, nationally representative panel survey project conducted by the Institute of Social Science Survey at Peking University.

The average age difference between husband and wife is 1.52 years. However, the peak of the age difference distribution is at the wife and the husband having the same age (see Fig. 1 panel (c)). In most couples the husband is less than five years older. In around 13% of the cases, the wife is one or two years older. The Pearson Correlation Coefficient is strongly significant at 0.88. Panels (d) illustrate the age distributions. The age requirement for marriage in China is 20 for women and 22 for men and we include only couples where both spouses are between 20 and 50 years old at the time of marriage.
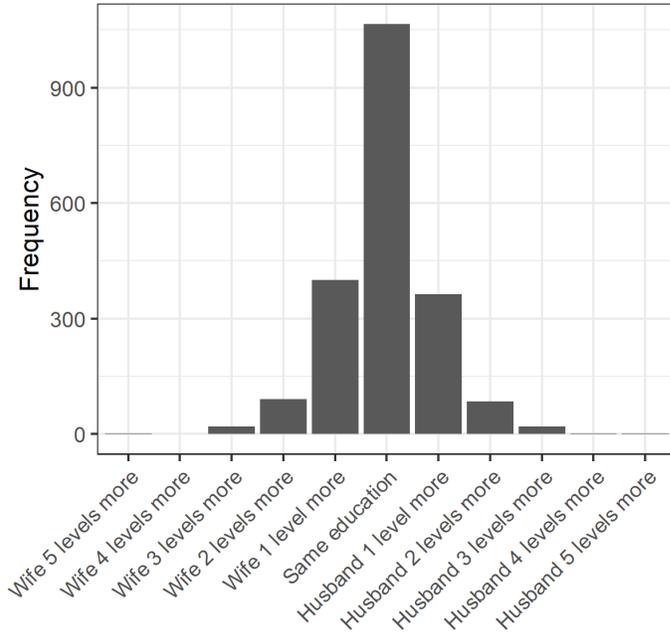
(a) Educational difference
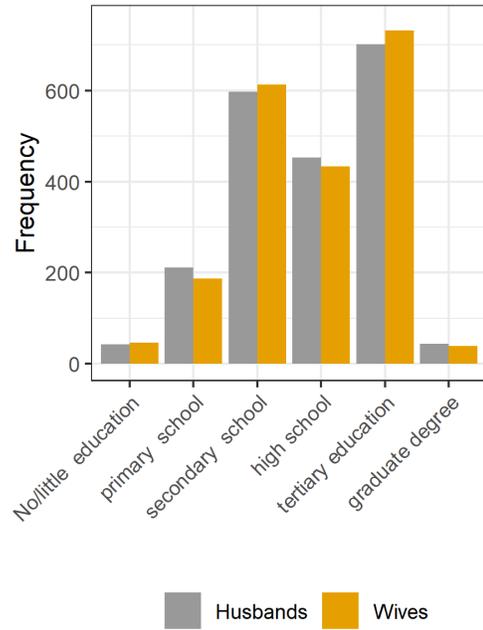
(b) Education distributions
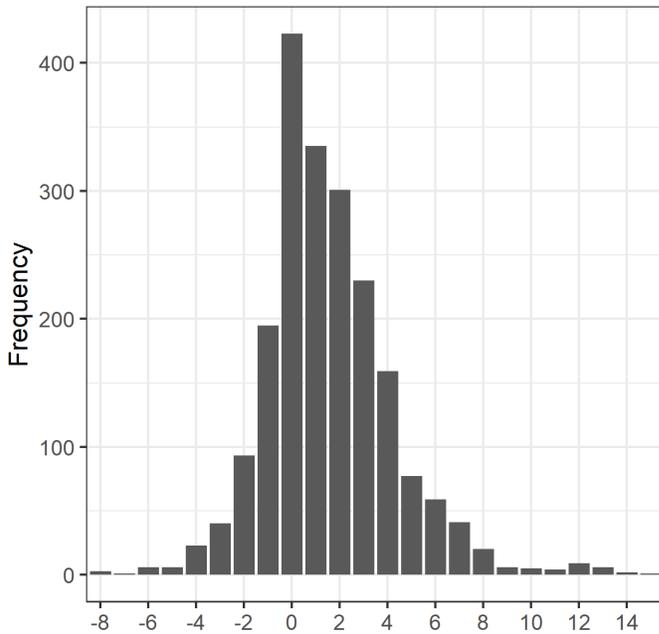
(c) Age difference (Husband - Wife)
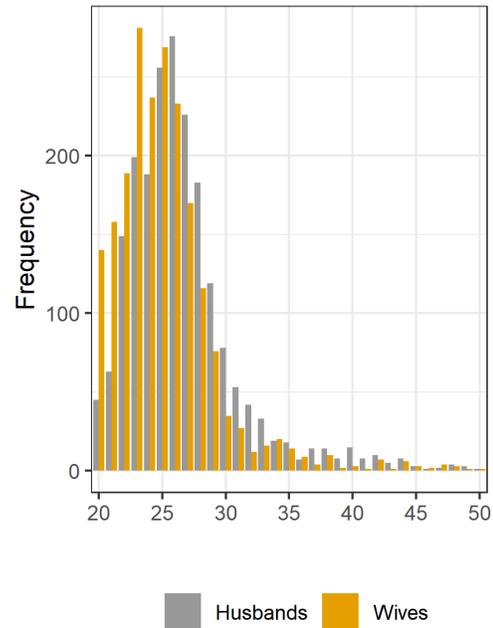
(d) Age distributions

Fig. 1. Frequency histograms for the educational and age difference between husband and wife. Couples who married between 2012 and 2016 between the age of 20 and 50. *Note*: Difference is calculated as husband minus wife. Age difference trimmed at -10 and +15. *Data Source:* CFPS 2014 and 2016.

3 Data

We ran a survey at the Green Lake Park every Saturday from late spring to early summer of 2016 named the Questionnaire for Search Activities for a Marital Partner in Yunnan (QSAMPY).[13] Participants were randomly approached by a student enumerator and received a small gift at the end of the survey.

Around 75% of respondents are searching for a spouse on behalf of someone else. Those that search for themselves have mostly been married before and are older than those who are represented by someone else. Those that are represented by their parents, an uncle or an aunt, have usually never been married. We are interested in parental preferences for first marriage and thus exclude those searching for themselves. We call the respondents searching for someone else "parents", though it includes other relatives if not explicitly specified. We call those on whose behalf the respondent is searching the "unmarried subject" or the "child", though it includes some nieces and nephews and combines those that are never married, divorces and separated - and all of them are above 20 years old. Our sample in total includes 412 observations.

3.1 Respondents: the "parents"

Respondents are predominantly female (64%) with an average age of 61. The majority is retired (86%) and married (91%).[14] Only 5% are at the park together with the unmarried subject. Around half of the respondents search on behalf of their daughter and one third on behalf of their son. The rest searches on behalf of their niece, their nephew or another relative. Summary statistics are displayed in table 3. The search platform does not attract many non-locals. Most QSAMPY respondents live and are registered in the city where the data was collected. As expected for an urban Chinese sample, most parents only have one child (73%) and 82% of parents live with the unmarried subject. Though the province of Yunnan is ethnically diverse, 94% of respondents state Han ethnicity. The share of ethnic minorities in our sample is thus lower than in the city where it was around 15% in 2007.[15]

The search efforts of the unmarried subject, the respondent and other family members are often complementary. Nearly half of the respondents indicate that the unmarried subject is also searching for a spouse. When the respondent is a parent, the spouse of the respondent is often also involved in the search process (36%). When the respondent is another relative, the parents are usually looking as well (67%). Of those that say that the unmarried subject is not searching, 'not having time/being too busy' is the most stated reason, followed by 'not wanting to' and 'being too shy'.

Only 23% of the interviewed sample is at the park for the first time. Respondents focus on the search at the park: only 4% use another platform (mostly online platforms). Most of the unmarried subjects (72%) know that they are represented at the park. Of those, the majority (68%) approve of this procedure. Overall, 30% of the unmarried subjects encourage the respondent to search on their behalf. These responses highlight that in the majority of cases the search for a spouse has

---

[13]We obtained approval the TSE-IAST Review Board for Ethical Standards in Research reference 2016-03-003, as well as permission from the Yunnan Normal University and the Police department.

[14]5% are widowed and 2.5% are separated or divorced.

[15]The original source for this information is not available anymore. However, it is quoted in several other websites, such as Wikipedia and https://www.gokunming.com/en/blog/item/397/kunming_residents_by_the_numbers.

been discussed within the family.

To investigate the respondent's motivation we ask how they themselves met their spouse and ask them to rate pre-selected motives. 36% of QSAMPY respondents were introduced to their spouse by friends, only 3% by their families and 17% by other relatives. Less than 1% indicates that their marriage was arranged (by their parents, relatives or a marriage agency). These numbers are comparable with those of the general population. Therefore, respondents are not searching because they themselves met their spouse in a similar way. Furthermore, parents rank altruistic motives as more important than more selfish ones (see table 4 in Appendix A). Altruistic motives include 'wanting their child to have someone who takes care of him or her' and 'having their child to have someone he/she feels affection' and more selfish motives include 'having a son or daughter-in-law take care of the respondent when the respondent is older' or 'having grandchildren'. Yet, the latter category is also often ranked as important. This is similar for both female and male unmarried subjects. These survey results suggest that parents see themselves as agents for their children. This does not exclude that parents put direct or indirect pressure on their children to find a spouse.

3.2 Unmarried Subjects: the "children"

The average age of unmarried subjects is 33 and 58% are women. The sample is very educated: 98% have at least finished professional high school and 60% have completed a university degree. 92% are employed and 5% are self-employed or entrepreneurs. 94% of unmarried subjects are of Han ethnicity, 2.8% of Bai ethnicity, and 2.2% of another ethnicity.

The never-married subjects are older and more educated than the average never-married person within the city. For this comparison, we use roster data from the representative Skills Towards Employability and Productivity (STEP) program, collected in 2012.[16] While the sex ratio in the general population is balanced, women are over-represented in the QSAMPY data. In the STEP data, only 1 unmarried person completed tertiary education, compared to 60% of unmarried subjects, and only 55% of unmarried STEP individuals finished high school, compared to 98% of unmarried subjects. In the STEP sample, only 62% state having worked at least one hour in the last seven days compared to 97% of unmarried subjects.[17] The Green Lake Park search platform seems to attract local, educated working individuals that are above the average marriage age.

3.3 The student sample

As a comparison group, we collected data from 283 students at the Yunnan Normal University. Students were randomly approached in front of the student canteens. Students are commonly used for surveys about marriage preferences because they are generally unmarried but of marriageable age. The results obtained from this sample can be compared with other studies. On the downside, students are not necessarily looking for a long-term partner at the time of the interview. They might not have very clear preferences or confound them with preferences they have for a current

---

[16]Collected by the World Bank within urban Kunming. The sampling method follows census paths. Table 5 in Appendix A shows the comparison.

[17]There are obvious limitations to the comparability between the QSAMPY data collected in 2016 and the STEP data from 2012. The marriage age between 2012 and 2016 presumably increased, as did educational attainments. However, the difference in the average age and educational levels are unlikely to just be driven by an increase by the average trends.

short-term partner.

The student sample is female-biased: 57% of the respondents are women. Yet, this reflects the university population where women are over-represented. The average age is 21. The student sample is more ethnically diverse with only 82% of the students are Han Chinese and 18% from ethnic minorities. Only 11% of the students are from the city of Kunming. The majority (65%) are from another place in Yunnan and 23% from another province. The students are, therefore, more rural (70% grew up in a rural area compared to 9% of unmarried subjects), ethnically diverse, more educated and substantially younger than the Green Lake Park unmarried subjects.

4 Preference Estimation: What do Parents Want?

4.1 Hypothetical profiles

During the interviews, respondents were shown four pairs of hypothetical profiles. The profiles displayed information on age, income, education, ethnicity and real estate ownership. They mimicked information that participants usually share with others at the park either written or orally. We asked respondents to state if the profile represents a person that they would want their unmarried subject to meet ("meet choice").[18] We therefore have 8 observations per respondent. The profiles were clearly described as hypothetical and the answers were not incentivized. The choice data, thus, do not provide revealed preferences. However, respondents are about to choose whom they want their unmarried subject to meet at the park. They have thought about which characteristics they prefer and are about to evaluate the same information. They do not have an incentive to deviate from the strategy they use for their actual choices. Therefore, stated preferences are presumably close to revealed preferences in this context.

The advantage of the hypothetical choice data is that characteristics are randomly created and not restricted due to platform entry or first sign of interest (such as in Banerjee et al. [2013] and Hitsch et al. [2010]). However, before the respondents saw the profiles, they had to choose a broad age category that they were interested in. The choice included 20 to 39 years old, 30 to 49 years old, 40 to 59 years old or over 50 years old. Each age bracket spanned 20 years or more. We wanted to avoid repeatedly showing participants profiles that were far off the desired age range. Yet, we wanted to observe their acceptable age limits. For students, the age category was either 16 - 29 or 20 - 39.

Figure 2 displays an example profile pair. The educational level was drawn from middle school degree, high school degree, university (bachelor) degree and graduate (masters) degree.[19] High school degree and university degree had a higher likelihood of being drawn. Again, this was designed to avoid participants' facing several unacceptable profiles. The monthly income indicated was between 2000 Yuan and 8000 Yuan. The ethnicity was either Han, Yi or Dai, with Han ethnicity having a higher likelihood to be on the profile. Real estate ownership was either "yes" or "no"

---

[18]Afterwards, they were also asked which profile of the two they preferred ("preference choice"). The results are omitted due to strong similarities.

[19]Education is officially mandatory until the completion of middle school (9 years of education). Afterwards, students can decide to continue schooling at different types of (senior) high school for three years which was not further specified in the profiles.

with the same likelihood and was not further specified.

Profile A			Profile B	
Age:	35		Age:	30
Education:	Bachelor Degree		Education:	High School
Monthly Income:	2000		Monthly Income:	4000
Ethnicity:	Han		Ethnicity:	Bai
Real Estate:	Yes		Real Estate:	Yes

Fig. 2. Example for the hypothetical profiles that were shown to respondents of the QSAMPY 2016. Translation, original profiles were in Mandarin Chinese.

4.2 Preference Estimation Framework

We assume that participants agree to a meeting if they expect the utility generated by the potential match to be higher than their "reservation utility" which is the utility of staying single and continuing the search. We define the utility a woman w gets from marrying a man m as $u_w(m)$ and the utility a man m gets from marrying a woman w as $u_m(w)$. We denote the reservation utility of women w v_w and the reservation utility of man m v_m. The probability of the respondents selecting a profile is defined as the probability that the utility derived from the match is higher than the reservation utility:

$$Prob(Meet = 1) = Prob(u_w(m) - v_w \geq 0) \tag{1}$$

This method is derived from the model of Adachi [2003] and also used in Hitsch et al. [2010]. We parameterize the utility function of woman w $u_w(m)$ as a function of man m's observed characteristics, how they interact with woman w's characteristics plus an error term that captures unobserved pairing-specific characteristics, with θ being the parameters to estimate:

$$u_w(m) = f(m, w'm, \theta) + \epsilon_{w,m} \tag{2}$$

Assuming $\epsilon_{w,m}$ is i.i.d. with the standard logistic distribution, we can derive the binomial logistic regression:

$$Prob(Meet = 1) = Prob(f(m, w'm, \theta) - v_w \geq 0) = \frac{exp(f(m, w'm, \theta) - v_w)}{1 + exp(f(m, w'm, \theta) - v_w)} \tag{3}$$

This equation can be estimated using a logit regression. We control for v_w or v_m as appropriate by including individual fixed effects.

4.3 Preference Estimation Results

We run a logit regression with the choice indicator as dependent variable. The indicator equals 1 if the respondent would want the unmarried subject to meet the person described in the profile. We include the characteristics of the profile: indicators for the different educational levels, the age

13

difference between the profile and the subject, the logarithm of income, an indicator for real estate ownership and for Han ethnicity. The reference category for education is "Profile: Junior High School". The results are displayed in table 1 separately for men and women. We also split the sample between unmarried subjects with and without a tertiary degree. We include individual fixed effects and cluster standard errors at the individual level. The coefficients display average marginal effects.

In table 2, the educational levels and income are replaced by the difference between the unmarried subject's educational level and income and the profile's educational level and income. Real estate is also interacted with own real estate ownership. This specification has fewer observations as respondents were reluctant to share certain information about the unmarried subject, particularly income. There is too little variation in the ethnicity of the unmarried subject, so that we do not include interaction terms with own ethnicity.

A small share of unmarried subjects has already been married (18%). The following results are robust to excluding them. There are too few observations to include interactions according to marital status.

4.3.1 Results for Women

We first look at spousal preferences when the unmarried subject is a woman. The respondents are thus looking for a son-in-law.

- **Education:** The likelihood of selecting the profile increases significantly with the educational level of the profile. This holds for both women with and without a university degree (Table 1 columns 1 to 3). On average, a profile with a graduate degree is 36 percentage points more likely to be selected than one with junior high school. The marginal effect of education is higher when the female subject has a tertiary degree. Respondents searching on behalf of a female subject with a tertiary degree have strictly increasing preferences for education: they dislike men with a lower educational level and like men with a higher educational level (see table 2). When the female unmarried subject does not have a tertiary degree, respondents dislike profiles with less education, but the coefficient for a higher educational level is no longer significant.

- **Income:** Respondents prefer male profiles with higher income. A profile with double the income has a 16 percentage points higher likelihood to be selected. The coefficient is higher for women who do not have a tertiary degree but not significantly so. Table 2 confirms that preferences are strictly increasing, though for subjects with a tertiary degree the coefficient for having more income than the subject is not significant.

- **Real Estate:** On average, owning real estate increases the likelihood of being chosen by around 7 percentage points. This holds for women with and without a university degree. The preference is driven by women who do not own real estate themselves: if the female subject does not have real estate, a profile with real estate has a 12 percentage point higher likelihood to be selected.

- **Age:** Respondents dislike a negative and a positive age difference between the unmarried subject and the profile. They prefer someone of a similar age. Table 1 column 7 includes

14

squared terms of the age variables and Fig. 3 panel (a) illustrates the results. Respondents dislike a negative age difference more than a positive age difference: the predicted likelihood of selecting a profiles drops sharply with a negative age difference but only slowly with a positive age difference. Therefore, preferences are single peaked and homogamic.

- **Ethnicity:** We find that respondents choosing on behalf of a female subject have a preference for Han ethnicity. On average, they are around 5 percentage points more likely to select a profile that states Han ethnicity. This preferences is driven by respondents searching on behalf of a subject without tertiary education. These are 10 percentage points more likely to select a profile that states Han ethnicity.

Table 1: Regressing profile selection on profile characteristics - Green Lake Park Sample.

	Dependent variable: Indicator of wanting to meet							
	Female unmarried subject				Male unmarried subject			
	all	w/o degree	with degree	all	all	w/o degree	with degree	all
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Profile: HS degree	0.121***	0.151**	0.114***	0.118***	0.135***	0.112*	0.214***	0.121***
	(0.029)	(0.069)	(0.030)	(0.029)	(0.046)	(0.066)	(0.050)	(0.046)
Profile: Undergraduate degree	0.309***	0.251***	0.335***	0.306***	0.158***	0.064	0.306***	0.144***
	(0.036)	(0.075)	(0.039)	(0.036)	(0.047)	(0.070)	(0.057)	(0.048)
Profile: Graduate degree	0.372***	0.251***	0.419***	0.366***	0.123**	-0.060	0.391***	0.100*
	(0.041)	(0.074)	(0.047)	(0.040)	(0.057)	(0.076)	(0.069)	(0.055)
Profile: Han ethnicity	0.052**	0.102**	0.028	0.051**	0.030	0.010	0.046	0.023
	(0.023)	(0.046)	(0.025)	(0.022)	(0.031)	(0.038)	(0.046)	(0.031)
Profile: Log(income)	0.165***	0.183***	0.148***	0.162***	-0.032	-0.112**	0.078	-0.036
	(0.028)	(0.053)	(0.033)	(0.028)	(0.041)	(0.053)	(0.058)	(0.041)
Age difference (+)	-0.023***	-0.015**	-0.026***	-0.022***	-0.049***	-0.053***	-0.046***	-0.051***
	(0.004)	(0.007)	(0.005)	(0.008)	(0.009)	(0.015)	(0.012)	(0.011)
Age difference (-)	-0.045***	-0.024**	-0.055***	-0.086***	0.010*	0.006	0.010	0.055***
	(0.007)	(0.011)	(0.009)	(0.015)	(0.005)	(0.007)	(0.007)	(0.014)
Profile: Owns real estate	0.071***	0.085*	0.066**	0.072***	0.022	0.024	0.035	0.019
	(0.025)	(0.045)	(0.030)	(0.025)	(0.029)	(0.039)	(0.042)	(0.028)
Age difference (+) squared				-0.000				0.001***
				(0.001)				(0.000)
Age difference (-) squared				0.005***				-0.004***
				(0.002)				(0.001)
Observations	1733	496	1237	1733	1180	652	528	1180

Note: Logit regression including individual fixed effects, coefficients indicate average marginal effects. Standard errors, clustered on the individuals level, are in parenthesis. Female unmarried subject = respondent is choosing on behalf of a woman; male unmarried subject = respondent is choosing on behalf of a man. Reference category for education: "Profile: Junior High School". "HS" is the abbreviation for high school. Age difference is calculated at the age of the profile minus the age of the unmarried subject. † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *Data source*: QSAMPY 2016

Table 2: Regressing profile selection on profile characteristics relative to unmarried subject's characteristics - Green Lake Park Sample.

	Dependent variable: Indicator of wanting to meet					
	Female unmarried subject			Male unmarried subject		
	all	w/o degree	with degree	all	w/o degree	with degree
	(1)	(2)	(3)	(4)	(5)	(6)
Educational difference (+)	0.050	0.016	0.096**	-0.044	-0.048	0.047
	(0.031)	(0.042)	(0.045)	(0.031)	(0.037)	(0.060)
Educational difference (-)	-0.175***	-0.156***	-0.173***	-0.116***	-0.042	-0.144***
	(0.020)	(0.049)	(0.021)	(0.028)	(0.048)	(0.030)
Profile: Han ethnicity	0.048**	0.091*	0.027	0.028	0.010	0.041
	(0.023)	(0.047)	(0.026)	(0.032)	(0.041)	(0.049)
Income difference (+)	0.139***	0.188**	0.096	-0.062	-0.116	0.068
	(0.053)	(0.090)	(0.068)	(0.073)	(0.092)	(0.117)
Income difference (-)	-0.188***	-0.167	-0.191***	-0.010	0.105	-0.089
	(0.046)	(0.107)	(0.051)	(0.061)	(0.092)	(0.078)
Age difference (+)	-0.023***	-0.015**	-0.027***	-0.051***	-0.052***	-0.049***
	(0.004)	(0.007)	(0.005)	(0.010)	(0.015)	(0.013)
Age difference (-)	-0.045***	-0.021*	-0.055***	0.008	0.004	0.010
	(0.007)	(0.011)	(0.010)	(0.005)	(0.007)	(0.007)
Profile: Owns real estate	0.124***	0.113	0.133**	0.058	0.076	0.057
	(0.045)	(0.076)	(0.057)	(0.055)	(0.074)	(0.089)
Subject: Owns real estate X Profile: Owns real estate	-0.096*	-0.051	-0.115*	-0.042	-0.074	-0.022
	(0.054)	(0.095)	(0.066)	(0.065)	(0.088)	(0.102)
Observations	1693	480	1213	1148	636	512

Note: Logit regression including individual fixed effects, coefficients indicate average marginal effects, standard errors clustered on the individual level are in parenthesis. Female unmarried subject = respondent is choosing on behalf of a woman; male unmarried subject = respondent is choosing on behalf of a man. The differences are calculated at the characteristic of the profile minus the characteristic of the unmarried subject. † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *Data source*: QSAMPY 2016
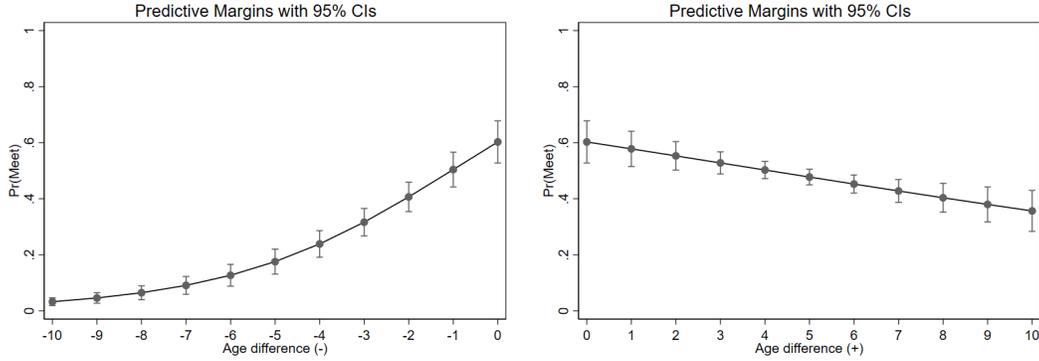
(a) Female unmarried subjects
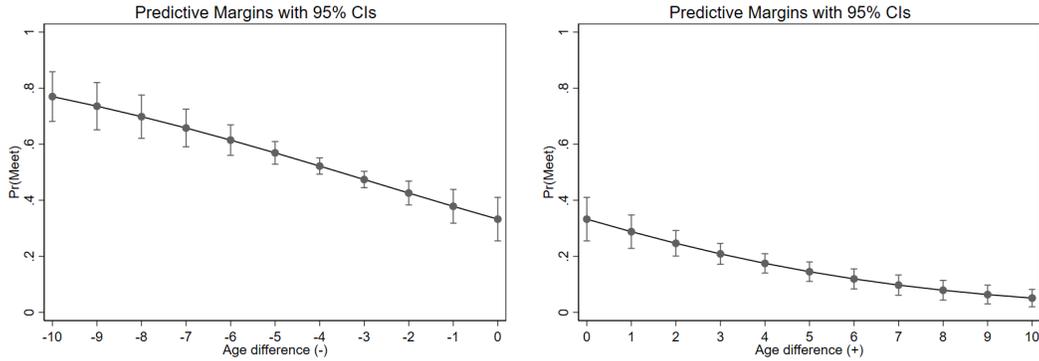
(b) Male unmarried subjects

Fig. 3. Predicted likelihood of choosing the profile according to the age difference (age of profile minus age of unmarried subject). Based on table 1 column 4 (panel a), column 8 (panel b).

4.3.2 Results for Men

We now look at spousal preferences when the unmarried subject is a man. Respondents are thus looking for a daughter-in-law.

- **Education:** Respondents on average prefer educated profiles. Yet, the coefficients are not significant for male subjects without a tertiary degree. For tertiary-educated subjects, a profile with an undergraduate degree has a 31 percentage point higher likelihood to be selected than one that only has a junior high school degree (table 1 column 7). We do not find evidence for a dislike of "too educated" profiles. The coefficient for being more educated than the subject is negative, but not significant (table 2 column 4). For men with a tertiary degree, respondents clearly dislike profiles with a lower educational degree. Yet, they do not have a preference for a higher degree. Preferences for educational levels seem to be homogamic.

- **Income:** On average, there is no preference for or against the female profile's income (table 1, column 4). Yet, respondents choosing on behalf of subjects without a tertiary degree dislike a high income (column 5). For those choosing on behalf of men with a tertiary degree the coefficient is positive but not significant. For relative income (table 2), none of the coefficients are significant. Yet the signs are opposite between subjects with and without a

18

tertiary education. For men without tertiary education, respondents might be less likely to select the profile when the woman earns more than the man.

- **Real Estate:** We do not find a preference for real estate ownership. The coefficient for real estate for those who do not own any themselves is positive, though not statistically significant (see table 2 column 4).

- **Age:** Respondents dislike a positive age difference (where the woman is older than the unmarried subject). Introducing squared terms (table 2, column 8), we find that respondents prefer younger women but this preferences decreases with the age gap. Fig. 3 panel (b) illustrates the dislike of older women on the right panel and a preference for younger women on the left panel. At higher age differences we have less power and do not know if preferences are decreasing at some point (the mathematical optimum lies at 15 years difference). It is not clear if preferences are single peaked for a high negative age differences or strictly decreasing in age.

- **Ethnicity:** For male subject, we do not find a preference for Han ethnicity. While the coefficient is positive, it is not significantly different from zero on average, nor for male subject with or without tertiary education.

4.4   The Value of Age, Education and Income in the Marriage Market

We find that respondents searching on behalf of a women on average have strictly increasing educational preference. Respondents searching on behalf of men on average dislike profiles with less education. We only find little evidence for a dislike of "over-educated" women. The coefficient for a female profile indicating a graduate degree compared to middle school is negative when parents select on behalf of their non-tertiary educated son, but it is not significant. Yet, when comparing a female profile that states "high school degree" and a female profile that states "graduate degree" the latter is significantly less likely to be selected by parents with a non-tertiary educated son.[20] However, their likelihood to select a female profile with high school is not significantly higher than one with a bachelor degree.

However, parents still have a preference for the woman being younger. On average, a woman with a high school degree would have to be over 10 years younger to have the same selection likelihood as a women with a university degree that has the same age as the unmarried man. For a man with tertiary education, a female profile with a university degree that is three years older than the man has the same likelihood to be selected as a woman with a high school degree and the same age as the man. For men without tertiary education, older female profiles are always less likely to be selected, independent of their educational level.

Respondents generally prefer high income male profiles with real estate, particularly so when their daughter does not have real estate herself. Having real estate increases the chances to be selected on average by 7 percentage points. Doubling the income increases a man's selection likelihood by 16 percentage points. High school increases the likelihood to be selected by 11 percentage points compared to middle school. Obtaining a university degree, compared to high school, increases the male profile's selection likelihood by around 18 percentage points and is thus slightly more important that a doubling of the income (at the average income). On average, a man with university degree has a higher selection likelihood than a man with a high school, even when he is

---

[20]Coefficient: $-.172$, $p < 0.01$.

up to 7 years older.

Being of Han ethnicity increases a male profile's selection likelihood by 5 percentage points on average, a value only slightly lower than owning real estate. The increase is similar to a 50% increase in income.

Interestingly, for tertiary-educated daughters, respondents have a preference for more educated men but no significant preferences for an income that is higher than hers. It only seems important that the man earns at least as much as the female subject. On the other side of the market, respondents searching on behalf of men without a university degree dislike female profiles with high income and this dislike seems to be toward those who earn more than the male unmarried subject. Yet, for men with tertiary education, we do not find this preference.

In fact, tertiary-educated women with high income have good chances among those that they themselves prefer: tertiary-educated men. A female profile with a tertiary degree is around 9 percentage points more likely to be selected by the parents of tertiary educated man than a profile with a high school degree. Reversely, a male profile with a tertiary degree is around 20 percentage points more likely to be selected by the parents of tertiary educated woman than a male profile with a high school degree. A match between a woman with a graduate degree and high income and a man without tertiary education and lower income is disliked on both sides.

5 Match Simulations: What do Parents Get?

The previous section focused on what parents *prefer.*. Yet, we would also like to know how these preferences are connected to the actual outcome. Do parents get what they want? Unfortunately, we could not contact the respondents again because of our assurances about anonymity. We therefore use a simulation approach based on the estimated preferences. In this section we ask if the preferences can explain the marriage patterns in the general population. The estimated preferences determine the spouse demand functions and the characteristics of the recently married general population determine the supply. We then compare simulated marriages with actual marriages and discuss when they overlap and when not.

5.1 Simulation Specifications

The CFPS 2014 and 2016 include information on the educational levels and the age of both spouses for marriages that were formed between 2012 and 2016. Unfortunately, they do not include information on ethnicity or pre-marital income or real estate ownership for both spouses. We use these observation as supply of husbands and wives. For each woman w and man m, we predict the likelihood of choosing each man m or women w based on Eq. 3 and the estimated parameters of θ. We denote this likelihood the "selection likelihood". In the profile task, this is equivalent to the respondent selecting the profile p on behalf of the unmarried subject i. Denote Y_{ip} the indicator for saying "yes" to profile p for subject i. We use θ from three specifications:[21]

---

[21]Table 6 in the appendix displays the estimates for θ for the three specifications for men and women.

- Specification 1 (Baseline): Includes the positive and negative age and educational difference (profile minus unmarried subject) as well as age squared terms:

$$Y_{ip} = \theta_1 \Delta Age_{ip}(+) + \theta_2 \Delta Age_{ip}(-) + \theta_3 \Delta Age_{ip}^2(+) + \theta_4 \Delta Age_{ip}^2(-) +$$
$$\theta_5 \Delta Education_{ip}(+) + \theta_6 \Delta Education_{ip}(-) + \epsilon_{ip} \tag{4}$$

- Specification 2: Includes the educational differences interacted with an indicator if the subject has a university degree and the age differences:

$$Y_{ip} = \theta_1 \Delta Age_{ip}(+) + \theta_2 \Delta Age_{ip}(-) + \theta_3 \Delta Age_{ip}^2(+) + \theta_4 \Delta Age_{ip}^2(-) +$$
$$\tilde{\theta}_5 \Delta Education_{ip}(+)' \, uni \, degree_i + \tilde{\theta}_6 \Delta Education_{ip}(-)' \, uni \, degree_i + \epsilon_{ip} \tag{5}$$

- Specification 3: Includes the age differences interacted with an indicator if the subject is over 30 and the educational differences:

$$Y_{ip} = \tilde{\theta}_1 \Delta Age_{ip}(+)' \, over \, 30_i + \tilde{\theta}_2 \Delta Age_{ip}(-)' \, over \, 30_i + \tilde{\theta}_3 \Delta Age_{ip}^2(+)' \, over \, 30_i$$
$$+ \tilde{\theta}_4 \Delta Age_{ip}^2(-)' \, over \, 30_i + \theta_5 \Delta Education_{ip}(+) + \theta_6 \Delta Education_{ip}(-) + \epsilon_{ip} \tag{6}$$

The higher the predicted selection likelihood of a woman w for men m, the higher the utility woman w derives from being matched with men m (see Eq. 1). The predicted selection likelihood thus gives us a cardinal ranking: The man with the highest selection likelihood is the woman's first choice, the man with the second highest selection likelihood is the woman's second choice and so forth.

The marriage market simulation corresponds to the man-proposing Gale-Shapley algorithm (Gale and Shapley 1962): Men propose to the woman that they rank the highest based on the predicted selection likelihood. A woman with one proposal stays with the proposer. A woman who receives more than one proposal selects the man that she ranks highest among them and rejects the others. In the second round, the rejected men propose to the woman that they attribute the second highest rank. Women then repeat their selection among the proposals. This is repeated until all individuals are matched.[22]

The Gale-Shapley mechanisms forces all women and men to match. Yet, there could be matches that would not form because one side would rather prefer to stay single. To address this issue, we calculate the lowest selection likelihood that is observed in the real outcome data for each specification and each educational level. For each women w and each man m, we predict the selection likelihood of their actual spouse, based on θ. We determine the lowest 1-percentile of actual selection likelihoods for women and men in each educational category respectively as the minimum thresholds.[23] We then include this sex and education-specific threshold as outside option in the simulation. In every proposal round, men can propose to their outside option if it has a higher rank than the next women they would otherwise propose to. Simultaneously, women reject all offers that are lower than their outside option. At the end, we can compute how many men and women stayed with their outside option.

---

[22]In order to verify the robustness of the resulting distribution, we also run the algorithm with women making the proposals and men rejecting and accepting proposals (woman-proposing Gale-Shapley mechanism). The results are very similar.

[23]We choose to use the lowest 1-percentile instead of the lowest value to make our results robust to outliers.

5.2 Simulation Results

5.2.1 Educational distribution

The frequency histograms of the actual educational difference between husband and wife and the simulated educational difference are displayed in Fig. 4 panel (a). Table 7 in Appendix A summarizes the main characteristics of the actual distribution and the three predicted distributions. We observe that the baseline specification 1 predicts a high degree of educational assortativity. Indeed, simulated matches are more assortative on education than is observed in the actual distribution. The share of simulated matches with the same educational level is 78% compared to 52% in the real distribution. This higher degree of assortativity can be explained by the lack of search frictions in the simulation. Search frictions in the marriage market can include search costs or geographical limits who one can meet. Also, the preference estimate is the average preference, and this might neglect important preference heterogeneity. Finally, assuming non-transferable utility might also lead to a higher level of assortativity in the simulations than in the actual matches that form under transferable utility.

Specification 2 and specification 3 based on parents' preferences predict less assortative matches, have a higher coefficient of variation in the educational differences between spouses and a lower correlation between the spouses' educational levels. These specifications have group-specific estimates: specification 2 has separate estimates for education for individuals with a high university degree and without. In specification 3, we have group-specific estimates on age for those where the unmarried subject is below 30 and above. These group-specific estimates, though possibly closer to the true parameters, are also noisier. The noisily measured estimates can explain the low correlation coefficients and the high coefficients of variation.

Finally, we observe that in all three simulations, there are more matches in which the husband has one more educational level than the wife than matches in which the wife has one more educational level than the husband. In the actual distribution, these two cases have a very similar frequency. In our sample, parents prefer a daughter-in-law with approximately the same educational level as their son, and sons-in-law that are at least as educated as their daughter. Matches in which they both have the same level are thus an equilibrium outcome, as specification 1 clearly highlights. Matches in which the husband has one more educational level are also a frequent outcome in the simulation while matches in which the wife has one more educational level are less so.

5.2.2 Age distribution

Figure 4 panel (b) illustrates the comparison in the distributions for the age differences between husband and wife. The baseline specification shifts the peak of the distribution to the right by one year - where husbands are one year older than their wives. In the actual distribution, the modal point is at the husband and the wife having the same age. In specification 2, most matches have an age difference between 0 and 4, yet the mode is at the husband being 2 years older. In specification 3, the mode is even at the husband being 3 years older. Again, specifications 2 and 3 based on parents preferences predict distributions with a higher coefficient of variation and lower correlations between the age of spouses.

In the modal simulations there is an age gap of between one and three years, driven by parents preferring younger daughters-in-law. The actual outcome, where most common case is spouses having the same age but the average age gap is husbands being 1.5 years older, suggests weaker preferences of this type. This could be because the unmarried subjects in the QSAMPY sample are on average older than the median marriage age in the general population. Parents with an older son might prefer a higher age gap for fertility reasons. Indeed, some studies show that male individual age preferences change with age (Kenrick and Keefe 1992; Walter et al. 2020). Yet, accounting for different preferences for unmarried subjects above and below 30 only increases the age gap in the simulation rather then decreasing it (specification 3).

5.2.3 Allowing for unmatched individuals

Fig 8 in Appendix A compares the simulated distributions with and without the outside option to stay unmarried for education and Fig. 9 in Appendix A for age. Figure 14 in Appendix A displays the rates of the unmarried population by education and sex. In specification 1, less than 1% end up unmatched. Therefore, the two distributions are almost identical. Specification 1 thus seems to have a good fit: it predicts an unmarried rate similar to the one for the real distribution (1%).

In specification 2, 7% are unmatched (mostly men and women with low levels of education). There are fewer couples in which the wife has three educational levels more than the husband, and where the wife is older than the husband. In specification 3, 14% are unmatched: Men with lower levels or education have the highest proportion of unmarried, as well as highly educated women (undergraduate and graduate degree). Again, there are fewer couples in which the wife is older and barely any couples in which the wife is more educated. The results for specification 2 and 3 highlight that parents prefer matches in which the husband has the same level as the wife or is more educated, and parents prefer a daughter-in-law who is younger than their son.

5.2.4 Weighting the general population

Men and women who married in the years 2012-2016 are clearly different from the men and women that are represented at the public park in the QSAMPY sample. In particular, QSAMPY subjects are mostly urban, older and more educated. In order to improve the comparison between the two samples we first only use couples in the CFPS that indicate that they live in an urban area (57% of households).[24] Fig. 10 in Appendix A illustrates the results in the second row. Though urban couples have a higher propensity to have the same age and the same education, the differences between all households and only urban households are barely visible. The simulated matches are also very similar to the one using all households.

To make the actual population more similar to the QSAMPY sample according to age and education we use propensity score weighting after we run the simulation. The weights are calculated based on a logit regression that has as dependent variable an indicator that equals 1 if the observation is from QSAMPY sample and 0 if it is from the general population. Education and age are

---

[24]Unfortunately, we do not know the location before marriage and the household registration status for the spouse, only the respondent.

(a) Educational distributions   **(b)** Age distributions
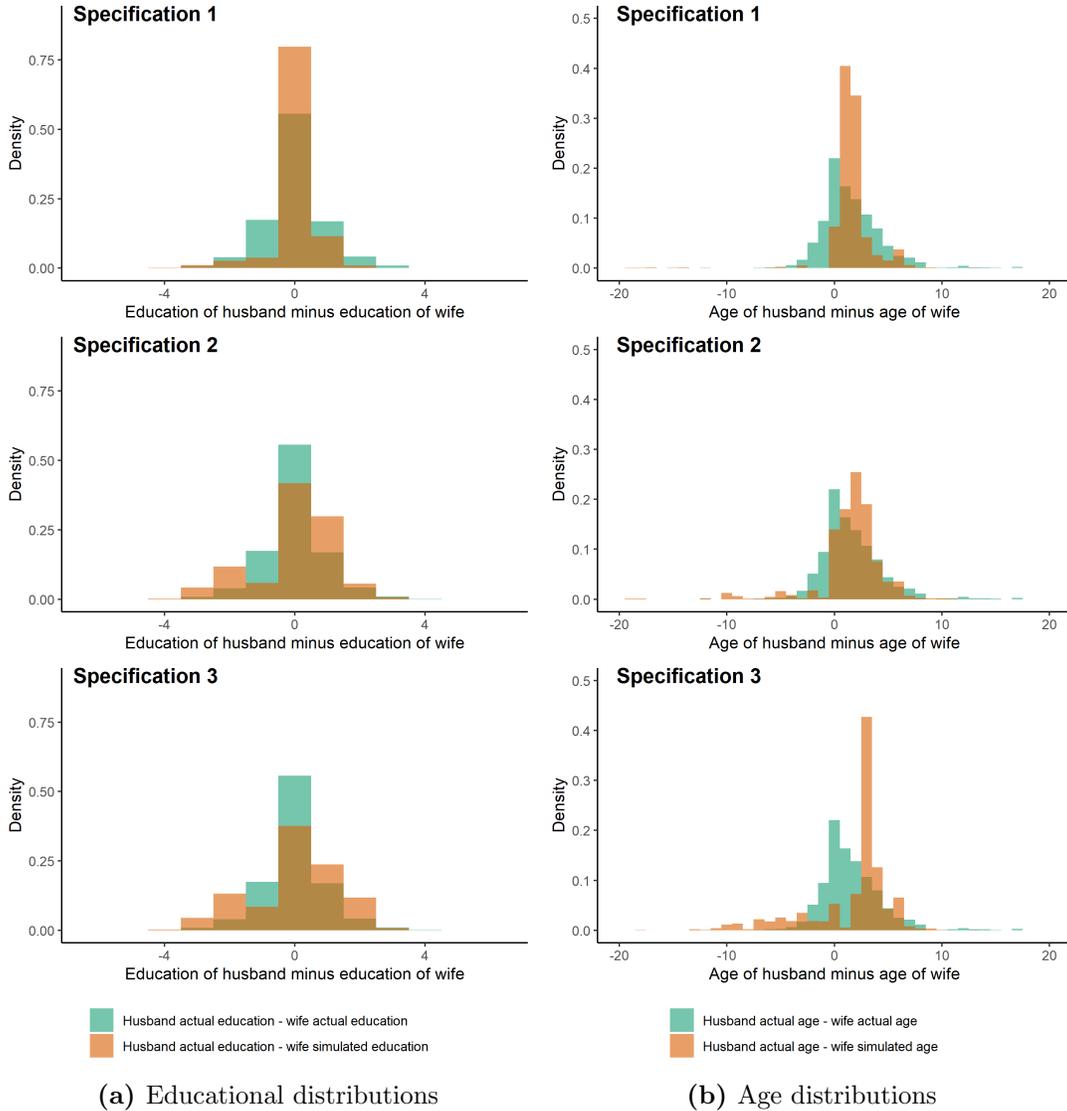
Fig. 4. Simulated educational distribution based on parental and students' preferences, in comparison with the real distributions. *Note*: Comparing actual educational differences between husband and wife and educational differences in simulated matches. Real distribution from spouses who married between 2012 and 2016. Preference parameters taken from table 6. Left-hand-side panels are without outside option of staying single and right-hand-side panels with outside option. *Data Source:* CFPS 2012 and 2016.

used as regressors.[25] Fig. 10 in Appendix A illustrates the results weighted for the characteristics of women represented by their parents in the third row and weighted for the characteristics of men represented by their parents in the fourth row.

Men who are similar to those in the QSAMPY sample are more likely to marry a wife who is younger and less educated than them, compared to other men in the overall population. The simulation does not predict this for education, where they are predicted to marry someone with the same level of education. For age, the weighting shifts the peak of the simulated distribution even further to the right where husbands are much older than their wives. Therefore, while it seems that older men have a higher likelihood of marrying younger wives, this is much less pronounced in the general population than is predicted by our preference estimates.

Women who are similar to those in the QSAMPY sample are more likely to marry a husband who is less educated than them, compared with other women in the overall population. This is predicted by the simulations as well, though more strikingly. They have the highest propensity to marry someone of their own age, the same as other women in the overall population. The simulations, however, do not capture this, and predict more couples with an older husband than the wife.

6 Discussion: What do young people want and what do parents accept?

Parental preferences seem to predict assortative matching according to educational levels quite well. Yet, parents' preferences predict slightly fewer matches where the wife has one educational level for than the husband in the specification allowing for education- or age-specific preferences (and the outside option of staying unmarried). Furthermore, parental preferences would predict there to be more couples where the husband is one to three years older than there actually are. In this section, we contrast parental preferences with preferences of a student sample that was collected in the same city, but is substantially different in its age distribution and more diverse in upbringing and ethnicity.

6.1 Students' preferences

Nearly all parents and relatives at the Green Lake Park came without the unmarried person. To have an idea about the preferences of young individuals who face the same profiles as the parents, we ran the survey at a local university. We would like to know where the preferences overlap and where they do not, and to see if the differences can explain the gap between simulated and observed marriage outcomes. Students' preferences are described in detail in Appendix B. Table 8 in Appendix B, displays the preference estimates.

Female students display the same increasing preferences for men's education, as do parents with tertiary-educated daughters. A male profile with an undergraduate degree is nearly 20 percentage points more likely to be selected than a male profile with a high school degree. They also have a preferences for real estate and high income. Female profiles with an undergraduate degree have the highest likelihood of being selected by male students, the same as by parents on average. Yet male

---

[25]Education in levels as a linear and squared term and a spline-based smooth function for age.

students also have a weak significant preference for a profile that has more education then them (see table 6, column (8)). This could also stem from the fact that they have not yet finished their education.

Using students' preferences for the simulation (specification 1), they also predict a high degree of assortativity on educational levels - identical to parents in specification 1 (see Fig. 16 panel (a)). They also predict the same share of matches where the man is more educated then when the woman is more educated, as is observed in the actual outcomes.

Female students display the same preferences for a partner with the same age, as do parents for their daughters. Male students have a dislike for older women, as parents do, but they do not have a significant preference for a younger partner. The age preferences are illustrated in panels (c) and (d) in Fig. 3. Male students are seemingly indifferent between a partner that has the same age, or is up to three years younger, but then the predicted likelihood to select a profile drops. Indeed, the likelihood curves for male and female students are nearly mirror-inverted, which is not the case for parents searching on behalf of much older men.

This difference in preferences leads to the peak of the age distribution to be one the husband and the wife having exactly the same age when using students' preferences for the simulation. The peak in the student's simulated distribution is therefore in line with the observed marriage distribution.

6.2   Accepted Age and Educational Levels

We also asked the parents what age levels they would accept for their child-in-law. Figure 15 panel (a) illustrates the lower and the upper age levels that parents indicated they would accept, according to the age of the unmarried subject. It also includes information on actual marriages in the general population. Indeed, parents with sons in their mid-20s accept daughters-in-law that have the same age than their son. Only when the son is in his mid-30s, parents state that they would not accept daughters-in-law that have the same age. For sons-in-law, parents always accept men who have approximately the same age as their daughter.

To investigate if these stated acceptance levels are indeed strict, we cross them with the profile selection data. We observe that 26% of the profiles shown to parents have an age *below* their lower stated accepted age level. Of those, 27% are still selected for a meeting. Parents searching on behalf of their son are more lenient (41%) than parents searching on behalf of their daughter (26%). 32% of the profiles indicate an age that is *above* the upper stated accepted age level. Of those, 23% are still selected for a meeting (21% for male subjects, 25% for female subjects). This suggests that while the stated acceptance levels bear some weight, they are not always enforced.

To summarize, we find that while parents prefer their son to marry a younger wife, they accept a wife with the same age, until their son is older. Since most people get married in their mid-20s, most marriages form between husband and wife who have no or a small age gap and are acceptable to the parents in our sample. Yet, that this is the actual outcome might better explained by taking into account the preferences of the younger, unmarried individuals - where men display no significant preferences for a younger wife. However, based on these data sources, we do not know if male individual preferences change with their age as well and converge towards the parental preferences

we find in the public park sample.

Panel (b) in Fig. 15 illustrates the same acceptance limits for education. It illustrates that parents have a dislike of sons-in-law with lower education than their daughters, and state that they would not accept such a match. Marriages between husband and wife that have the same education are accepted by both sides. At the same time, we find that the upper limit of parents searching on behalf of a man is almost never binding - they usually accept a daughter-in-law who is more educated than their son. The only exemption is that for sons that have a university undergraduate degree, some parents do not accept a women who has a graduate degree.

The acceptance limits also underline the finding that the parents in the public park sample want their son-in-law to have at least as much education as their daughter, which is not always the case in the observed marriage outcomes. Again, stated acceptance levels should not be interpreted as being always binding. We observe that of the 26% of profiles that have an educational level below the lower stated accepted educational level, 19% are still selected. Of those that have an educational level above the upper stated accepted educational level (only 8.5%), 30% are still selected.

7 Conclusion

This paper investigates parental preferences and their influence on marriage patterns, using data from China. Although the explicit involvement of parents in marriage decisions is more common in China than in Western countries, parental influence on decisions about marriage (as well as about education and location) are important in Western countries as well. The novelty of this paper is that it explicitly discusses parental preferences rather than seeking to infer them from outcomes. It does so by interviewing parents and other relatives that are currently actively searching for a spouse for their adult child. Parents see themselves as agents for their children, though they also indicate wanting to have grand-children and someone to take care of them when they are older.

Parents seem to have a preference for an age gap that students do not, and this preference predicts marriage patters that are different than the observed marriage patters. This could be due to age preferences changing with the age of the unmarried subject: parents mostly search on behalf of an unmarried subject that is older than the students. Yet, allowing for age-specific preferences and weighting the general population to make it more similar to the interviewed population helps to explain this discrepancy. It is also possible that there is a difference between generations. One possibility, up for future research, is that the younger generation is more in contact with other individuals of the opposite sex of the same age due to the increase in time spent on education. However, we also observe that parents accept daughters-in-law of the same age as their son when their son is in his 20s to mid 30s. As most marriages occur at this time, the observed marriages seem not necessarily preferred but acceptable to parents.

Parents want a daughter-in-law with the same education as their son (and usually accept one that has slightly more education). Parents of tertiary-educated daughters want a son-in-law who is at least as educated and might object against a match with a less educated man. It therefore seems reasonable to suggest that parental preferences are for now not an obstacles but rather an encouragement to investments in education on the part of girl children. Yet, while these preferences predict current matching patters well, and overlap substantially with students' preferences,

it raises the question of what will happen if women's educational attainments overtake those of men. This might matter not just because they could be constrained directly by preferences about the education levels of wives, but also indirectly by preferences on age levels, since more educated wives tend to be older when they enter the marriage market.

The Chinese population that recently married is one where women and men have very similar educational attainments, yet according to the UNESCO Institute for Statistics, since 2016, more women are enrolled in tertiary education than men.[26] For now, marriages between women and men where the woman is more educated are common. Yet, if the outside options for educated women improve, they may decide to delay marriage or stay unmarried.

References

Hiroyuki Adachi. A search model of two-sided matching under nontransferable utility. *Journal of Economic Theory*, 113(2):182–198, 2003.

Abi Adams and Alison Andrew. Preferences and beliefs in the marriage market for young brides. Technical report, CEPR Discussion Paper, 2019.

Orazio Attanasio and Katja Kaufmann. Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations. *Journal of Economic Behavior and Organization*, 140:35–55, 2017. doi: 10.1016/j.jebo.2017.05.002.

Abhijit Banerjee, Esther Duflo, Maitreesh Ghatak, and Jeanne Lafortune. Marry for what? caste and mate selection in modern india. *American Economic Journal: Microeconomics*, 5(2):33–72, 2013.

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

Michele Belot and Marco Francesconi. Can anyone be'the'one? evidence on mate selection from speed dating. 2006.

Lei Chang, Yan Wang, Todd K Shackelford, and David M Buss. Chinese mate preferences: Cultural evolution and continuity across a quarter of a century. *Personality and Individual Differences*, 50(5):678–683, 2011.

Pierre-André Chiappori and Bernard Salanié. The econometrics of matching models. *Journal of Economic Literature*, 54(3):832–61, 2016.

Pierre-André Chiappori, Bernard Salanié, and Yoram Weiss. Partner choice and the marital college premium: Analyzing marital patterns over several decades. 2015.

Pierre-André Chiappori. The theory and empirics of the marriage market. *Annual Review of Economics*, 12:547–578, 2020. doi: 10.1146/annurev-economics-012320-121610.

Pierre-André Chiappori, Murat Iyigun, and Yoram Weiss. Investment in schooling and the marriage market. *American Economic Review*, 99(5):1689–1713, 2009. doi: 10.1257/aer.99.5.1689.

---

[26]Accessed on the World Bank website https://data.worldbank.org/indicator/SE.TER.ENRR.FE?locations=CN in February 2019.

Eugene Choo and Shannon Seitz. The collective marriage matching model: Identification, estimation, and testing. *Structural Econometric Models*, pages 291–336, 2013.

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

Wendell Cox. Demographia world urban areas. *14th Annual Edition*, 2018.

Daniela Del Boca and Christopher J Flinn. Household behavior and the marriage market. *Journal of Economic Theory*, 150:515–550, 2014.

Marcos Delprato, Kwame Akyeampong, and Mairead Dunne. Intergenerational education effects of early marriage in sub-saharan africa. *World Development*, 91:173–192, 2017. doi: 10.1016/j. worlddev.2016.11.010.

John W Engel. Marriage in the people's republic of China: Analysis of a new law. *Journal of Marriage and the Family*, pages 955–961, 1984.

C Cindy Fan and Youqin Huang. Waves of rural brides: Female marriage migration in China. *Annals of the Association of American Geographers*, 88(2):227–251, 1998.

Raymond Fisman, Sheena S Iyengar, Emir Kamenica, and Itamar Simonson. Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics*, 121(2):673–697, 2006.

Raymond Fisman, Sheena S Iyengar, Emir Kamenica, and Itamar Simonson. Racial preferences in dating. *The Review of Economic Studies*, 75(1):117–132, 2008.

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

Alfred Galichon and Bernard Salanié. Cupid?s invisible hand: Social surplus and identification in matching models. 2015.

John Giles, Albert Park, and Meiyan Wang. The great proletarian cultural revolution, disruptions to education, and the returns to schooling in urban china. *Economic Development and Cultural Change*, 68(1):131–164, 2019.

Li Han, Tao Li, and Yaohui Zhao. How status inheritance rules affect marital sorting: Theory and evidence from urban China. *The Economic Journal*, 125(589):1850–1887, 2015.

Louise T Higgins and Chunhui Sun. Gender, social background and sexual attitudes among Chinese students. *Culture, Health & Sexuality*, 9(1):31–42, 2007.

Louise T Higgins, Mo Zheng, Yali Liu, and Chun Hui Sun. Attitudes to marriage and sexual behaviors: A survey of gender and culture differences in China and United Kingdom. *Sex Roles*, 46(3):75–89, 2002.

Günter J Hitsch, Ali Hortaçsu, and Dan Ariely. Matching and sorting in online dating. *The American Economic Review*, 100(1):130–163, 2010.

Anning Hu and Zhenchao Qian. Does higher education expansion promote educational homogamy? evidence from married couples of the post-80s generation in Shanghai, China. *Social science research*, 60:148–162, 2016.

Yang Hu. Marriage of matching doors: Marital sorting on parental background in China. *Demographic Research*, 35:557–580, 2016.

Fali Huang, Ginger Zhe Jin, and Lixin Colin Xu. Love and money by parental matchmaking: evidence from urban couples in china. *American Economic Review*, 102(3):555–60, 2012.

Fali Huang, Ginger Zhe Jin, and Lixin Colin Xu. Love, money, and parental goods: Does parental matchmaking matter? *Journal of Comparative Economics*, 2016.

Douglas T Kenrick and Richard C Keefe. Age preferences in mates reflect sex differences in human reproductive strategies. *Behavioral and Brain Sciences*, 15(1):75–91, 1992.

Robert Kurzban and Jason Weeden. Do advertised preferences predict the behavior of speed daters? *Personal Relationships*, 14(4):623–632, 2007.

Hongbin Li, Pak Wai Liu, and Junsen Zhang. Estimating returns to education using twins in urban china. *Journal of Development Economics*, 97(2):494–504, 2012.

John Allen Logan, Peter D Hoff, and Michael A Newton. Two-sided estimation of mate preferences for similarities in age, education, and religion. *Journal of the American Statistical Association*, 103(482):559–569, 2008.

Annemie Maertens. Social norms and aspirations: Age of marriage and education in rural india. *World Development*, 47:1–15, 2013. doi: 10.1016/j.worlddev.2013.01.027.

Zheng Mu and Yu Xie. Marital age homogamy in China: A reversal of trend in the reform era? *Social science research*, 44:141–157, 2014.

David Ong and Jue Wang. Income attraction: An online dating field experiment. *Journal of Economic Behavior & Organization*, 111:13–22, 2015.

Ellen Efron Pimentel. Just how do i love thee?: Marital relations in urban China. *Journal of Marriage and Family*, 62(1):32–47, 2000.

Corinne Reczek, Hiu Liu, and Debra Umberson. Just the two of us? how parents influence adult children's marital quality. *Demography*, 72(5):1205–1219, 2010. doi: 10.1111/j.1741-3737.2010. 00759.x.

Nancy E Riley. Interwoven lives: Parents, marriage, and Guanxi in China. *Journal of Marriage and the Family*, pages 791–803, 1994.

Kathryn V Walter, Daniel Conroy-Beam, David M Buss, Kelly Asao, Agnieszka Sorokowska, Piotr Sorokowski, Toivo Aavik, Grace Akello, Mohammad Madallh Alhabahba, Charlotte Alm, et al. Sex differences in mate preferences across 45 countries: A large-scale replication. *Psychological Science*, page 0956797620904154, 2020.

Linda Y Wong. Structural estimation of marriage models. *Journal of Labor Economics*, 21(3): 699–727, 2003.

Peng Xia, Kun Tu, Bruno Ribeiro, Hua Jiang, Xiaodong Wang, Cindy Chen, Benyuan Liu, and Don Towsley. Characterization of user online dating behavior and preference on a large online dating site. In *Social Network Analysis-Community Detection and Evolution*, pages 193–217. Springer, 2014.

Yan R Xia and Zhi G Zhou. The transition of courtship, mate selection, and marriage in China. *Mate selection across cultures*, pages 231–246, 2003.

Jia Yu and Yu Xie. Changes in the determinants of marriage entry in post-reform urban china. *Demography*, 52(6):1869–1892, 2015.

8 Appendix A: Additional Tables and Figures

8.1 Tables

Table 3: Summary statistics Green Lake Park Respondents

Variable	Mean	Std. Dev.	N
Respondent: Female	0.748		412
Respondent: Age	61.418	6.47	347
Respondent: Mother	0.641		412
Respondent: Father	0.211		412
Respondent: Married	0.915		410
Respondent: Widowed	0.049		410
Respondent: Divorced	0.022		412
Respondent: Retired	0.84		412
unmarried subject at park as well	0.046		412
Subject: Female	0.583		412
Subject: Age	33.187	5.592	412
Subject: Never Married	0.915		411
Subject: Working	0.968		411
Subject: At least high school	0.993		412
Subject: University degree	0.597		412

Data Source: QSAMPY 2016. Does not include respondents who look for a spouse for themselves.

Table 4: Share of respondent indicating reason for search as important or very important.

	All	Searching Husband		Searching Wife	
		Mother	Father	Mother	Father
unmarried subject having someone who helps him/her with household chores .	52%	41%	68%	62%	57%
unmarried subject having someone to take care of him/her in time of need (sickness/old-age).	86%	85%	79%	90 %	81%
unmarried subject having someone he/she feels affection for.	96%	95%	97 %	97 %	98%
Having someone to care for you (and your spouse) when you are older.	51%	51%	50%	56%	43%
Having grandchildren	63%	60%	57%	69%	68%
N	351	170	34	94	53

Data Source: QSAMPY 2016., only mothers and fathers, excluding uncles and aunts and other relatives

Table 5: Comparison between STEP and QSAMPY data

	QSAMPY (2016) unmarried Subjects		STEP (2012)			Difference	
	never married	no partner	never married	no partner	all	never married	no partner
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age (mean)	32.43	33.06	26.19	33.70	44.40	6.24***	-0.64
Female	59%	58%	44%	48%	51%	15pp***	10pp***
Completed min. (prof.) high school	98%	98%	55%	46%	25%	43pp***	52pp***
University degree	62%	60%	0%	0%	0%	62pp***	60pp***
Working[27]	97%	97%	65%	57%	62%	32pp***	40pp***
N	377	410	516	649	4970		

Note: Column (1) and (3) include information on those who have never married, column (2) and (4) who currently do not have a partner (those unmarried/divorced/widowed and single), column (5) includes the whole population (including married and with partner). Asterisks indicate significant difference at 1% (***). Comparisons are: QSAMPY never married to STEP never married; QSAMPY no partner - STEP no partner. Two sample t-test.
Data Source: QSAMPY 2016 and STEP 2012.

---

[27]STEP: Has been working at least 1 hour in last 7 days. QSAMPY: Employed or self-employed

Table 6: Estimations to simulate marriage outcomes

|	Female Subject			Male Subject			Female Students	Male Students
	Spe. 1 (1)	Spe. 2 (2)	Spe. 3 (3)	Spe. 1 (4)	Spe. 2 (5)	Spe. 3 (6)	Spe.1 (7)	Spe 1 (8)
Age Difference (+)	−0.133***	−0.134***	0.036	−0.163***	−0.164***	−0.119	−0.0001	−0.099***
	(0.021)	(0.021)	(0.073)	(0.039)	(0.039)	(0.135)	(0.046)	(0.033)
Age Difference (+) Sqrd	0.003***	0.003***	−0.007	0.004**	0.004**	−0.006	−0.004	0.002
	(0.001)	(0.001)	(0.005)	(0.002)	(0.002)	(0.013)	(0.003)	(0.002)
Age Difference (-)	−0.393***	−0.393***	−0.101	0.189***	0.194***	0.419	−0.263	−0.143*
	(0.055)	(0.055)	(0.236)	(0.049)	(0.049)	(0.288)	(0.176)	(0.084)
Age Difference (-) Sqrd	0.024***	0.024***	−0.027	−0.015***	−0.015***	−0.079	−0.012	−0.001
	(0.005)	(0.005)	(0.044)	(0.004)	(0.004)	(0.053)	(0.032)	(0.012)
Education Diff (+)	0.102	0.201	0.093	−0.009	0.058	−0.009	0.255**	0.142*
	(0.102)	(0.213)	(0.102)	(0.110)	(0.153)	(0.111)	(0.107)	(0.080)
Education Diff (-)	−0.585***	−0.311	−0.582***	−0.099	0.351*	−0.103	−0.448***	−0.524***
	(0.077)	(0.288)	(0.077)	(0.090)	(0.205)	(0.091)	(0.100)	(0.074)
Indicator: Educated		0.408*			0.114			
		(0.235)			(0.202)			
Education Diff(+) X Ind.: Educated		−0.054			−0.133			
		(0.245)			(0.238)			
Education Diff(-) X Ind.: Educated		−0.314			−0.539**			
		(0.299)			(0.230)			
Indicator: Over 30			0.360			−0.785**		
			(0.258)			(0.320)		
Age Difference(+) X Ind.: Over 30			−0.198**			−0.023		
			(0.078)			(0.146)		
Age Diff. (+) Sqrd X Ind.: Over 30			0.010**			0.009		
			(0.005)			(0.013)		
Age Difference(-) X Ind.: Over 30			−0.298			−0.154		
			(0.244)			(0.293)		
Age Diff. (-) Sqrd X Ind.: Over 30			0.051			0.061		
			(0.045)			(0.054)		
Observations	1,836	1,836	1,836	1,259	1,259	1,259	1,264	2,238

Note: Logit regression including individual fixed effects, coefficients indicate average marginal effects. Standard errors (unclustered) are in parenthesis. Female unmarried subject = respondent is choosing on behalf of a woman; male unmarried subject = respondent is choosing on behalf of a man. Differences are calculated as the profile minus the unmarried subject. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$. *Data Source:* QSAMPY 2016. QSAMPY 2016

Table 7: Comparing real marriage distribution with predicted distributions

	Real distribution	Spe. 1 Parents	Spe. 2 Parents	Spe.3 Parents	Spe. 1 Students
Education					
Average difference btw. spouses	-0.012	-0.012	-0.012	-0.012	-0.012
Correlation between spouses' levels	0.450	0.386	-0.006	-0.056	0.451
Coefficient of Variation (Difference btw spouses)	46.557	30.490	64.048	67.603	29.791
Most common case (mode)	Same level	Same level	Same level	Same level	Same level
Share same level	0.520	0.780	0.387	0.379	0.713
Kolmogorov–Smirnov Statistic	0.156	0.119	0.119	0.099	
Age					
Average difference btw. spouses	1.523	1.523	1.523	1.523	1.523
Correlation between spouses' levels	0.789	0.878	0.655	0.492	0.659
Coefficient of Variation (Difference btw spouses)	1.829	1.400	2.331	2.824	2.317
Most common case (mode)	Same age	Husband 1 year older	Husband 2 years older	Husband 3 years older	Same age
Share same age +/- 2 years	0.657	0.829	0.564	0.149	0.782
Kolmogorov–Smirnov Statistic		0.236	0.224	0.390	0.333
Share of unmarried	0.01	0.008	0.07	0.14	0.025

Note: Calculations based on the simulations without outside option, except the share of unmarried population (last row). Correlation coefficients is the Pearson's Correlation Coefficient. The share of unmarried in the real distribution is fixed to 1%.

8.2 Figures
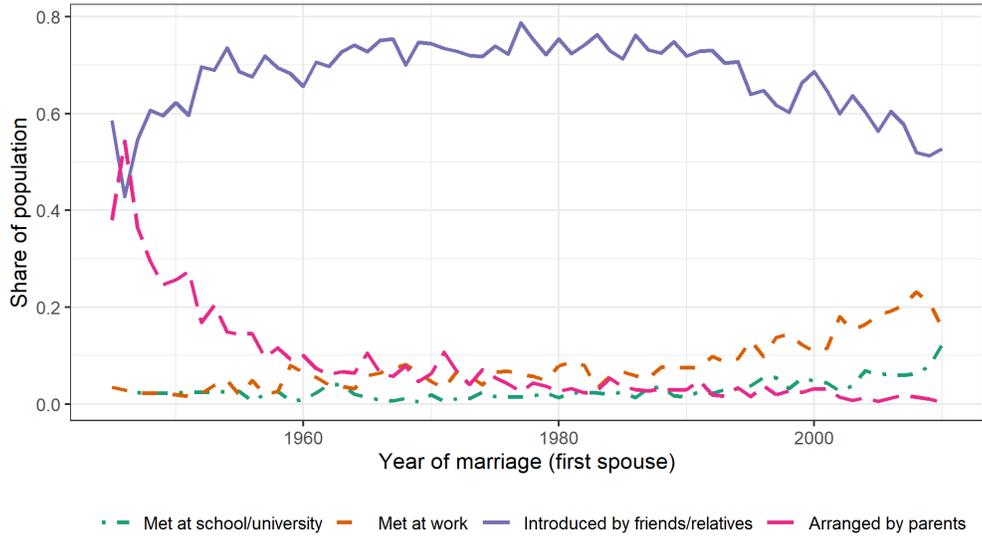
Fig. 5. Answer to the question how the respondent met their first spouse. *Data Source:* CFPS 2010, marriages between 1945 and 2010. Omits categories with low incident rates ("other") such as meeting online.
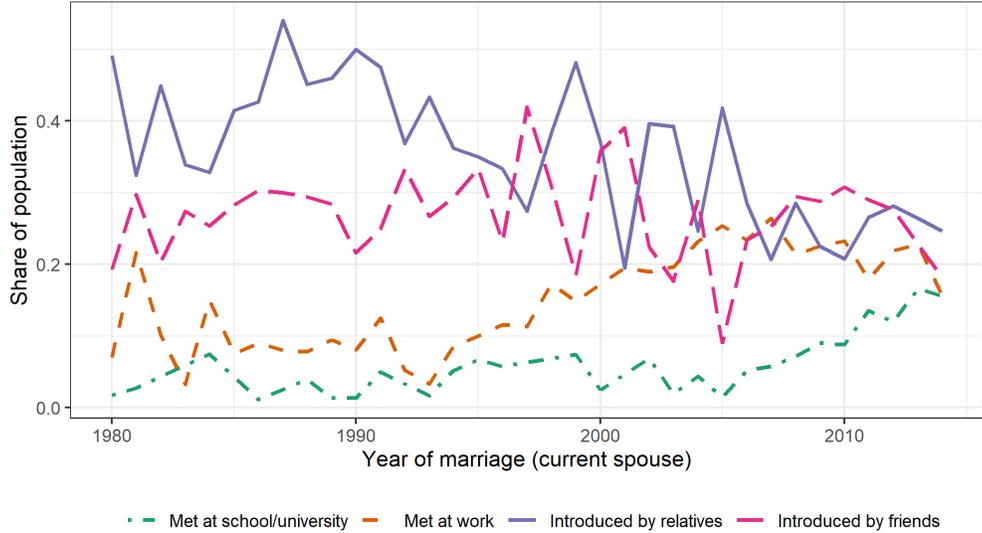

Fig. 6. Answer to the question how the respondent met their current spouse of a subset of the respondents. *Data Source:* CFPS 2016, marriages between 1980 and 2014. Omits categories with low incident rates ("other") such as meeting online.
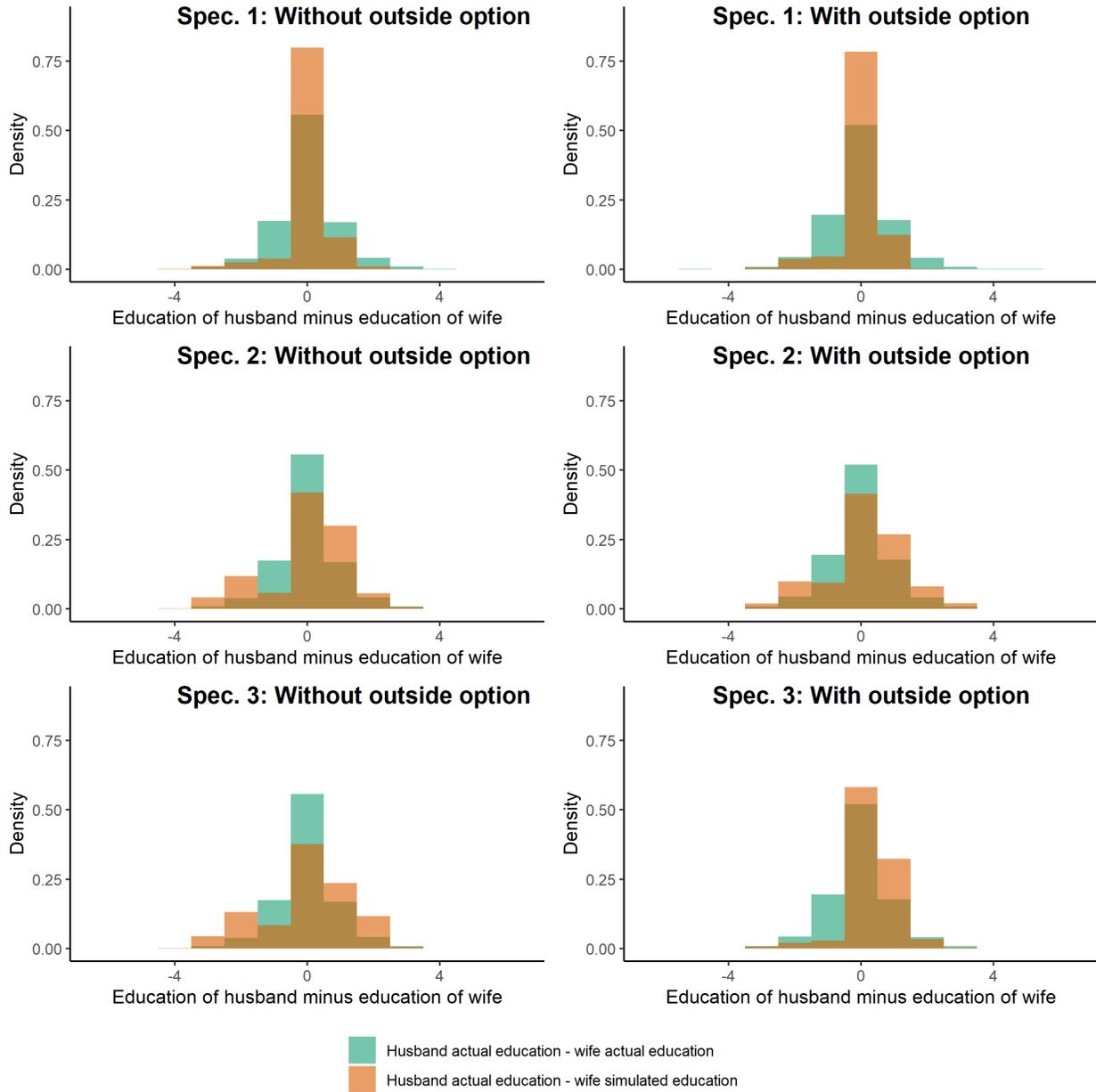
Fig. 7. Educational distributions

Fig. 8. Educational distributions with and without the outside option to stay unmarried. *Note*: Comparing actual educational differences between husband and wife and educational differences in simulated matches. Real distribution from spouses who married between 2012 and 2016. Preference parameters taken from table 6. Left-hand-side panels are without outside option of staying single and right-hand-side panels with outside option. *Data Source:* CFPS 2012 and 2016.
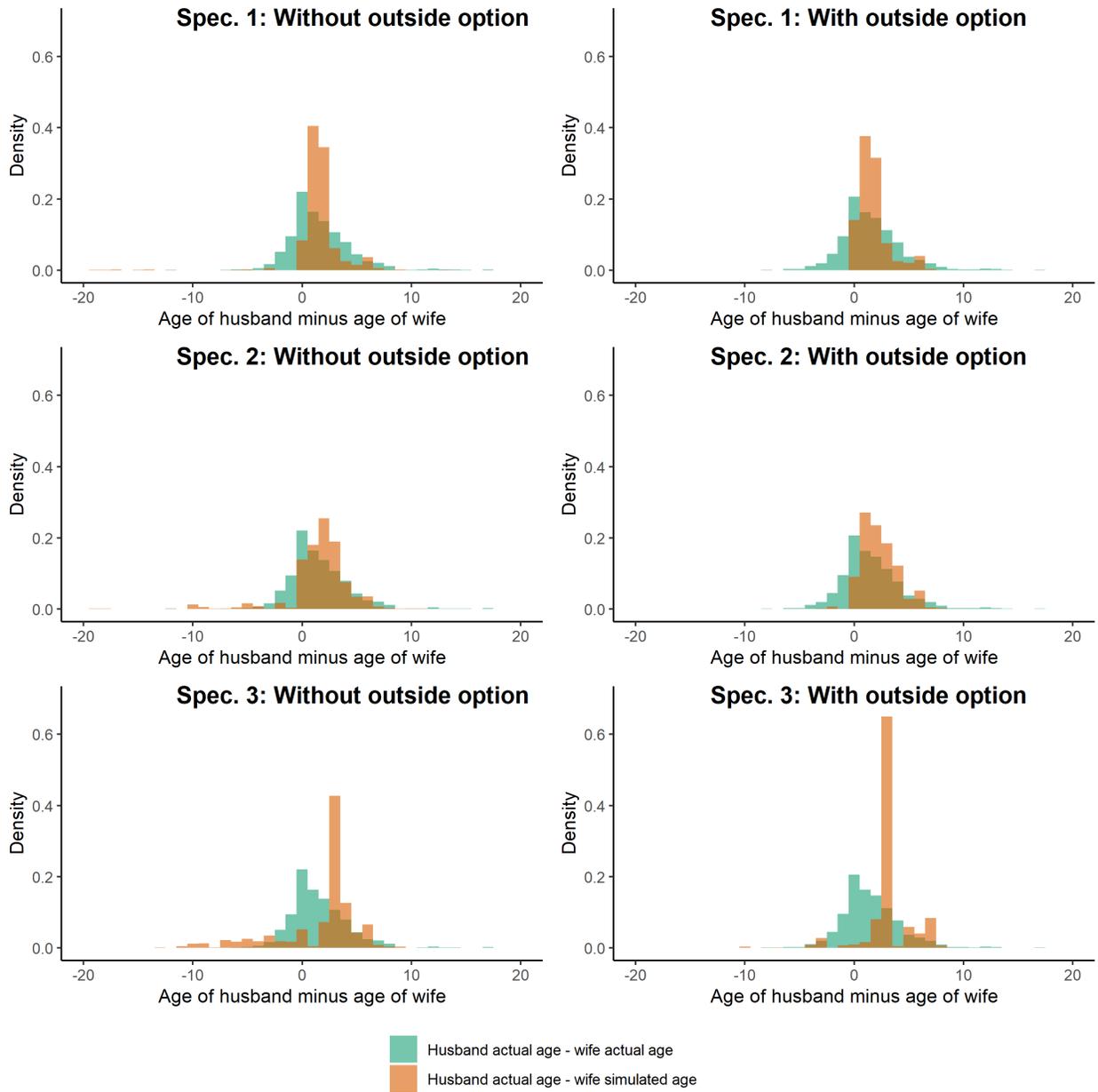
Fig. 9. Age distributions with and without the outside option to stay unmarried. *Note*: Comparing actual age differences between husband and wife and age in simulated matches. Real distribution from spouses who married between 2012 and 2016. Preference parameters taken from table 6. Left-hand-side panels are without outside option of staying single and right-hand-side panels with outside option. *Data Source:* CFPS 2012 and 2016.
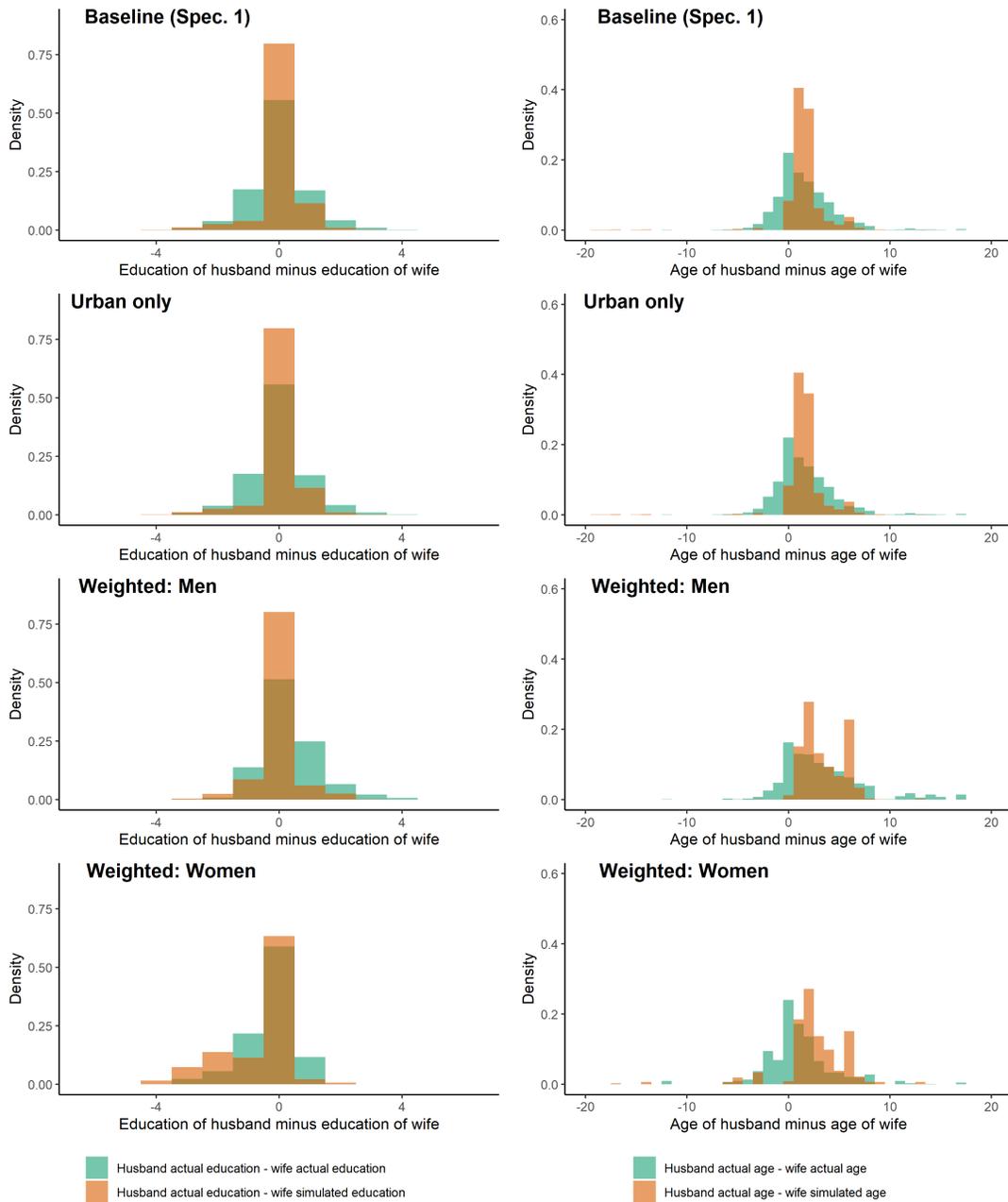
(a) Educational distributions

(b) Age distributions

Fig. 10. Simulated educational distribution based on parental preferences, in comparison with the real distributions. *Note*: Comparing actual educational differences between husband and wife and educational differences in simulated matches. Real distribution from spouses who married between 2012 and 2016. Preference parameters taken from table 6. First row: baseline specification 1. Second row: baseline specification 1 with only households that live in an urban area. Third row: Male observations weighted according to propensity scores. Forth row: Female observations weighted according to propensity scores. *Data Source:* CFPS 2012 and 2016.
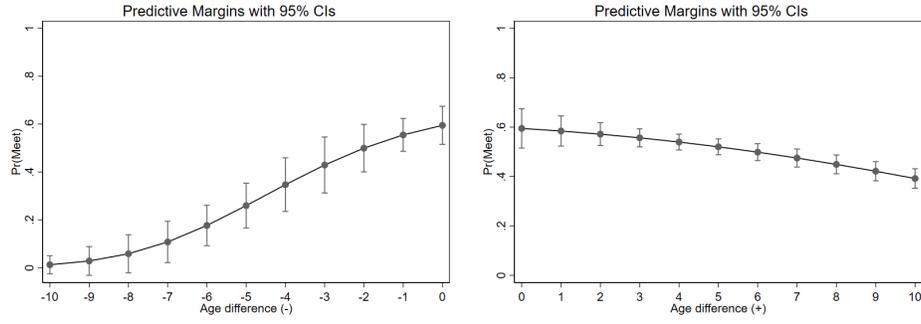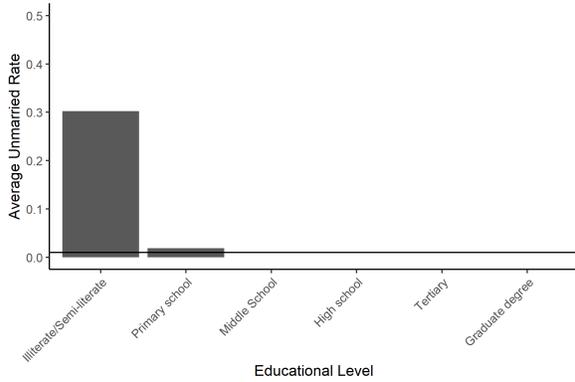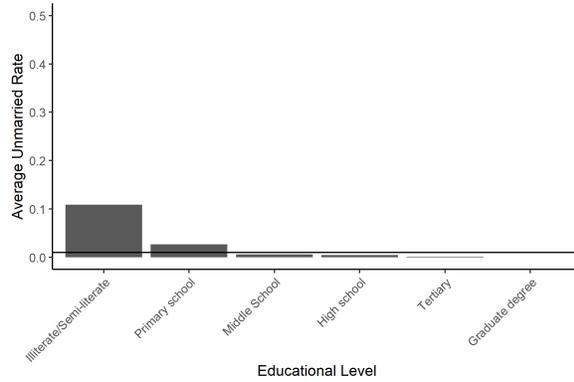
Fig. 11. Female students
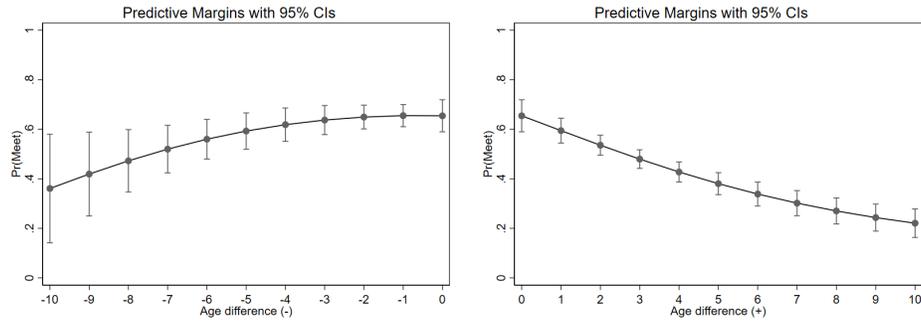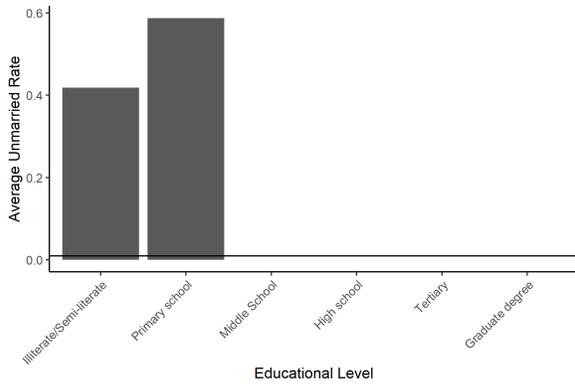

Fig. 12. Male students

Fig. 13. Predicted likelihood of choosing the profile according to the age difference (age of profile minus age of unmarried subject). Based on table 1 column 4 (panel a), column 8 (panel b), table 8 column 2 (panel c), column 5 (panel d).
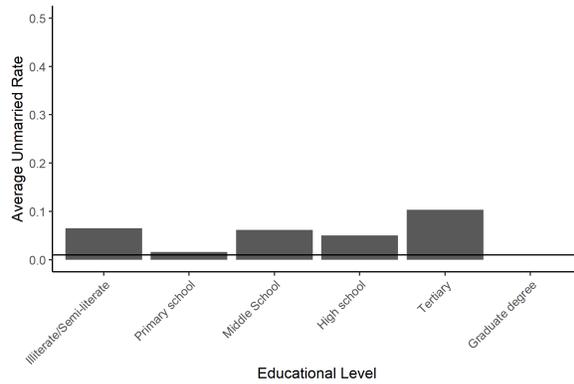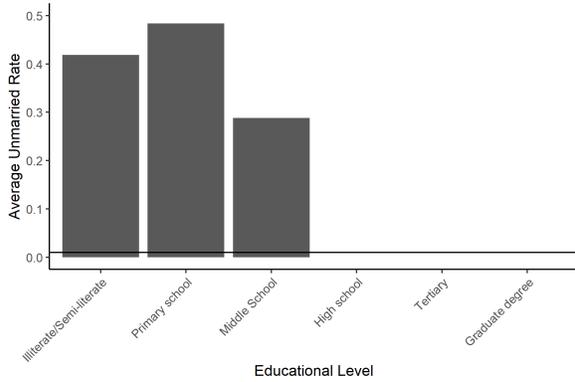
(a) Specification 1: Unmarried men
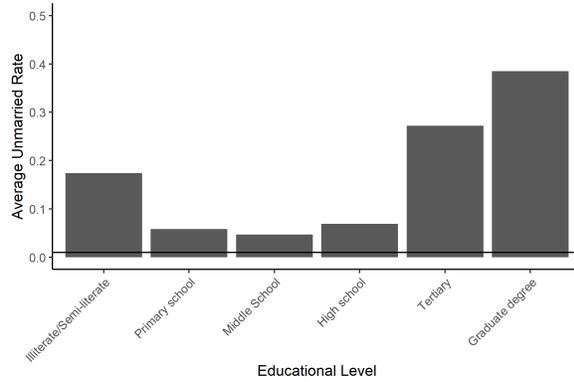
(b) Specification 1: Unmarried women

(c) Specification 2: Unmarried men

(d) Specification 2: Unmarried women
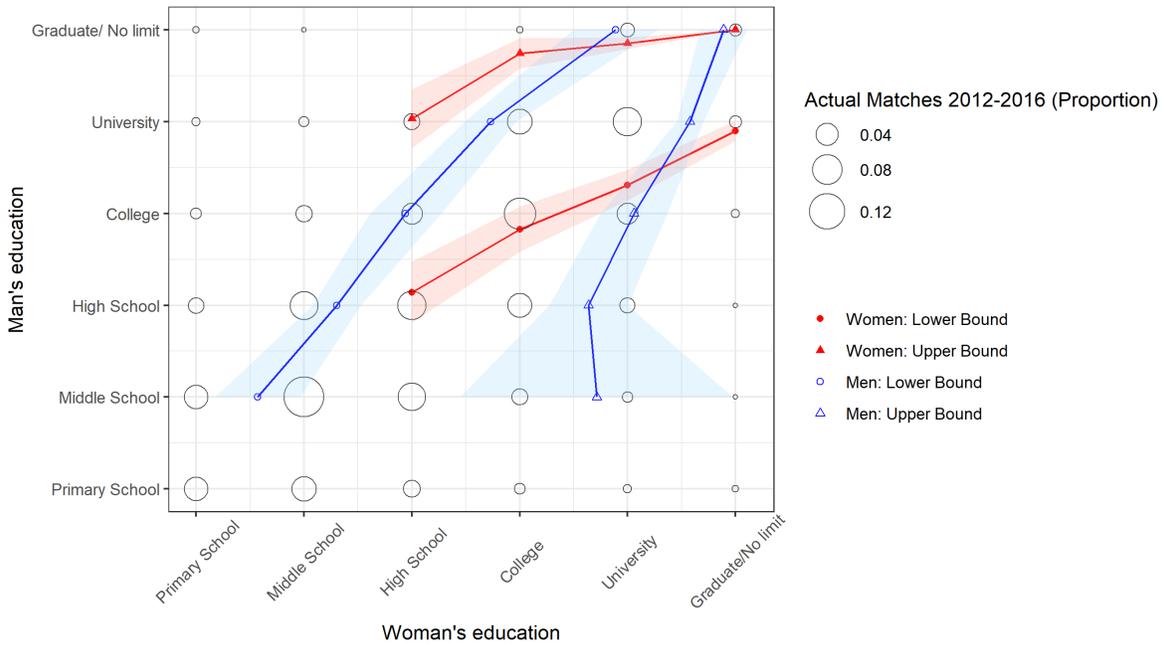
(e) Specification 3: Unmarried men

(f) Specification 3: Unmarried women

Fig. 14. Simulated sex-specific and education-specific unmarried rates according to the different specifications. The shares are compared to a rate of 1% which is fixed in the actual distribution by using the 1%-lowest selection likelihood for the minimum threshold.
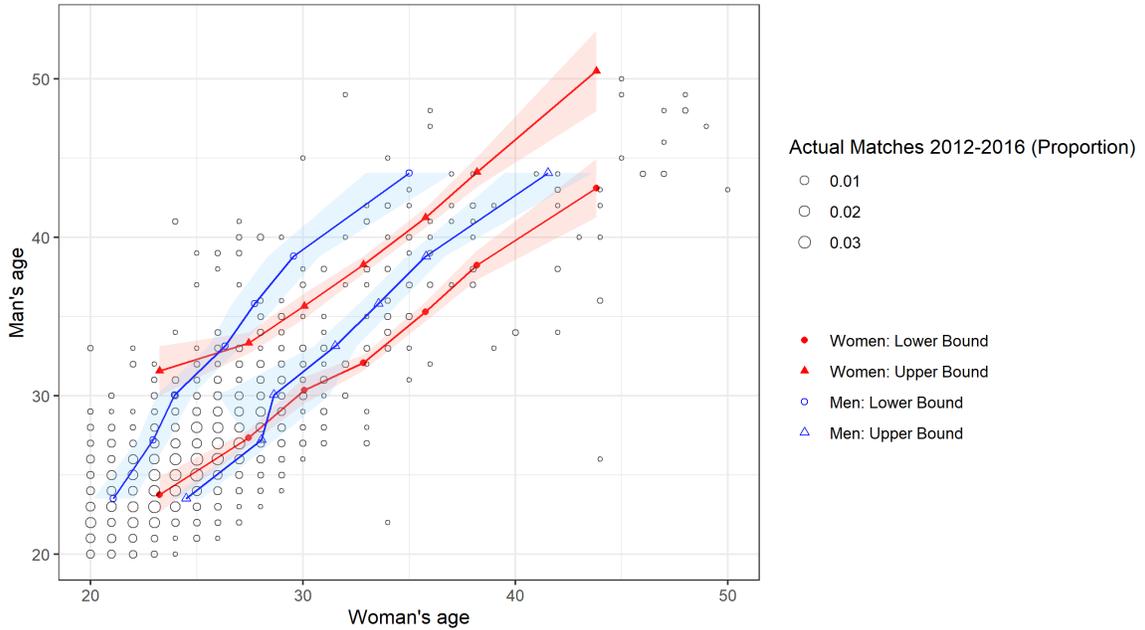
(a) Age: Acceptance Limits

(b) Education: Acceptance Limits

Fig. 15. Stated upper and lower acceptance limits for age and education according to sex and the education/age of the unmarried subject. *Notes:* Limits based on the QSAMPY 2016 questionnaires. Shaded areas indicate the 95% Confidence Intervals. Actual matches are based on the CFPS 2014 and 2016 and include those who married between 2012 and 2016 and where both spouses are between 20 and 50 years old. Panel (a): Age categories were constructed and illustrated using the average age per category. Only includes those who state a limit ("no limit"-option excluded). Panel (b): "no limit" was transformed into the lowest level (primary school) for the lower bound and the highest level (graduate degree) for the upper bound.

9 Appendix B: Students Preference Estimation Results

The results from the student sample are displayed in table 8. Since students have a low current income but a high future income, we do not look at the differences between the profile's income and current income.

- **Age:** Female students have a preference for meeting someone with a similar age (homogamic preferences). Yet the dislike of younger men is stronger than the dislike of older men. Male students have a dislike of meeting someone who is older. They seem to be indifferent between someone who has the same age and someone who is younger. The preferences are illustrated in Fig. 3 panels c and d.

- **Education:** Both male and female students have a strong preference for someone who is educated. The likelihood of the profile being selected is the highest with a graduate degree when a woman is choosing (strictly increasing preferences) and an undergraduate degree if a man is choosing (homogamic preferences).

- **Income:** Both male and female students have a preference for higher income. Interestingly, the coefficient for male students is nearly as high as the coefficient for female students.

- **Real estate:** Female students have a preference for someone who has real estate while male students do not.

- **Ethnicity:** Both male and female students have a strong preference for Han Chinese. Female students are 14 percentage points more likely to select a profile that states Han ethnicity. Male students are 11 percentage points more likely to do so. We investigate what individuals characteristics drive these ethnic preferences in table 9. The student sample is diverse in their upbringing and in ethnicity: 30% grew up in a rural area and 18% are part of an ethnic minority. We interact the indicator for the profile stating Han ethnicity with an indicator for the subject having grown up in a rural area (column 1 and 3) and with an indicator for the subject being part of an ethnic minority (column 2 and 4). We find that for female students, the preference for Han ethnicity is stronger for those who grew up only in a rural area (column 1). However, for male students, the interaction between the profile's ethnicity and upbringing is insignificant (column 3). Furthermore, the preference for Han ethnicity seems to be driven by those who are Han themselves. The coefficient for Han ethnicity is important and statistically significant for those who are Han themselves (columns 2 and 4). The interaction with the indicator for ethnic minority is negative, yet it is not statistically significant.

Table 8: Regressing profile selection on profile characteristics - Student Sample

	Dependent variable: Indicator for selecting 'meet' option					
	Female subject			Men subject		
	(1)	(2)	(3)	(4)	(5)	(6)
Profile: HS degree	0.177***	0.178***		0.117***	0.112**	
	(0.042)	(0.042)		(0.044)	(0.045)	
Profile: Undergraduate degree	0.371***	0.371***		0.374***	0.369***	
	(0.045)	(0.045)		(0.048)	(0.048)	
Profile: Graduate degree	0.429***	0.433***		0.286***	0.290***	
	(0.050)	(0.050)		(0.058)	(0.059)	
Profile: Han ethnicity	0.140***	0.137***	0.140***	0.105***	0.107***	0.096***
	(0.027)	(0.027)	(0.027)	(0.032)	(0.032)	(0.033)
Profile: Log(income)	0.215***	0.215***	0.217***	0.152***	0.152***	0.146***
	(0.032)	(0.032)	(0.032)	(0.033)	(0.034)	(0.034)
Age difference (+)	-0.026***	-0.010	-0.026***	-0.046***	-0.063***	-0.044***
	(0.004)	(0.012)	(0.004)	(0.007)	(0.013)	(0.007)
Age difference (-)	-0.076***	-0.034	-0.076***	-0.012	0.010	-0.012
	(0.013)	(0.045)	(0.013)	(0.009)	(0.026)	(0.009)
Profile: Owns real estate	0.067**	0.065**	0.063**	0.034	0.034	0.030
	(0.030)	(0.031)	(0.030)	(0.039)	(0.039)	(0.039)
Age difference (+) squared		-0.001			0.002	
		(0.001)			(0.001)	
Age difference (-) squared		-0.007			-0.004	
		(0.007)			(0.003)	
Educational difference (+)			0.105***			0.009
			(0.030)			(0.038)
Educational difference (-)			-0.199***			-0.203***
			(0.026)			(0.025)
Observations	1168	1168	1168	934	934	934

Note: Logit regression including individual fixed effects, coefficients indicate average marginal effects, standard errors are in parenthesis. Reference category for education: "Profile: Junior High School". "HS" is the abbreviation for high school. The differences are calculated at the characteristic of the profile minus the characteristic of the unmarried subject. † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *Data source*: QSAMPY 2016

Table 9: Regressing profile selection on profile characteristics - Student Sample - Investigating ethnic preferences

	Dependent variable: Indicator for selecting 'meet' option			
	Female subject		Male subject	
	(1)	(2)	(3)	(4)
Profile: HS degree	0.177***	0.179***	0.117***	0.116***
	(0.043)	(0.042)	(0.044)	(0.044)
Profile: Undergraduate degree	0.369***	0.372***	0.374***	0.373***
	(0.046)	(0.045)	(0.047)	(0.048)
Profile: Graduate degree	0.427***	0.431***	0.285***	0.283***
	(0.050)	(0.050)	(0.058)	(0.058)
Profile: Log(income)	0.212***	0.213***	0.150***	0.152***
	(0.032)	(0.033)	(0.033)	(0.033)
Age difference (+)	-0.027***	-0.026***	-0.046***	-0.046***
	(0.004)	(0.004)	(0.007)	(0.007)
Age difference (-)	-0.076***	-0.076***	-0.011	-0.012
	(0.013)	(0.013)	(0.009)	(0.009)
Profile: Owns real estate	0.067**	0.065**	0.034	0.034
	(0.030)	(0.030)	(0.039)	(0.039)
Profile: Han ethnicity	0.051	0.153***	0.151**	0.119***
	(0.049)	(0.029)	(0.062)	(0.036)
Rural childhood X Profile: Han ethnicity	0.123**		-0.065	
	(0.058)		(0.073)	
Subject: Minority X Profile: Han ethnicity		-0.097		-0.071
		(0.069)		(0.074)
Observations	1168	1168	934	934

Note: Logit regression including individual fixed effects, coefficients indicate average marginal effects, standard errors are in parenthesis. Reference category for education: "Profile: Junior High School". "HS" is the abbreviation for high school. The differences are calculated at the characteristic of the profile minus the characteristic of the unmarried subject. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. *Data source*: QSAMPY 2016
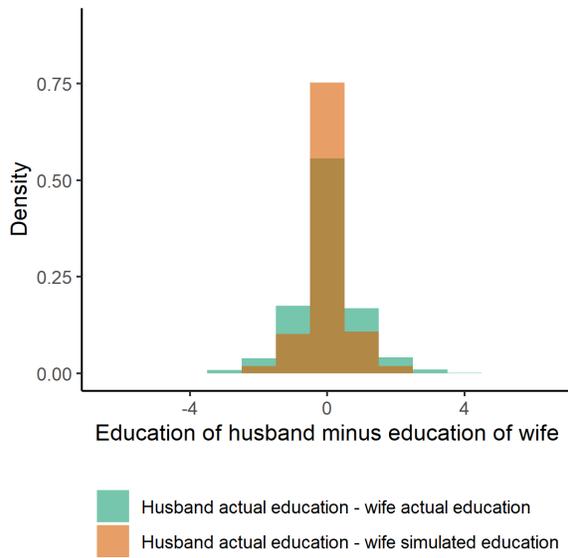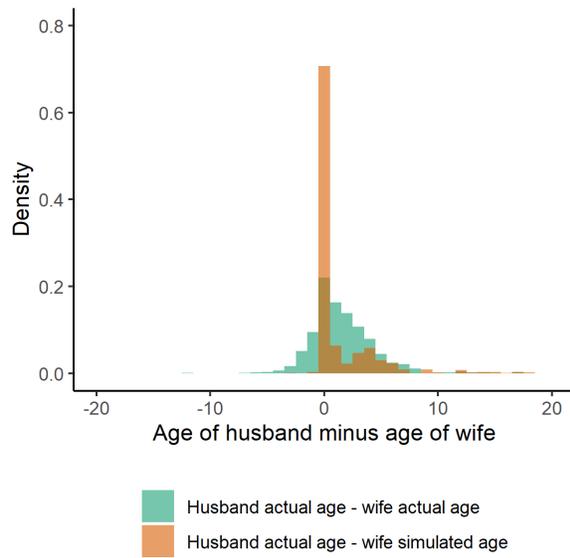
(a) Educational distributions   **(b)** Age distributions

Fig. 16. Simulated educational distribution based on students' preferences, in comparison with the real distributions. *Note*: Comparing actual educational differences between husband and wife and educational differences in simulated matches. Real distribution from spouses who married between 2012 and 2016. Preference parameters taken from table 6. Without the outside option to stay unmarried. Data Source:CFPS 2012 and 2016.