Marital Sorting and Housing Prices in China: Evidence from Online Dating

Y. Joy Chen and Sitian Liu*

May 22, 2019

Abstract

We estimate the effects of China's surging house prices on individuals' marital preferences and equilibrium assortative matching patterns. Using data from China's largest dating website, we estimate mate preferences based on users' decision to reply to a first-time message from a contact. We find that (1) site users, in particular women, have strong preferences for home-ownership, and increases in housing prices are associated with higher reply rate by women; (2) compared to non-homeowners, homeowners have stronger preferences for home-ownership; (3) home-ownership increases users' competitiveness and this effect becomes more pronounced for men when as housing prices increase; (4) there is weak evidence on the impact of housing prices on equilibrium assortative matching patterns.

*Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305. We thank Pascaline Dupas, Saumitra Jha and Jonathan Levin for their helpful comments. We are grateful to Mr. Yang Li and Mr. Jinlong Wu from Jiayuan for their generous assistance with data sampling and collection.

1 Introduction

This paper studies how the surge in housing prices affects mate preferences and assortative matching patterns in China's marriage market. Anecdotal evidence suggests economic factors, in particular home-ownership, are becoming increasingly important in women's assessment of men's marriageability. According to a survey by Shanghai Daily in March 2010, 80 percent of mothers with young daughters were reluctant to let their daughters marry men who do not own an apartment. Another survey conducted by the New York Times shows that more than 70 percent of single women would only consider men who own an apartment as a prospective husband (Jacobs 2011).

That housing prices could affect mate preferences and matching patterns in China's marriage market is highly plausible. Since 2007, China's unrelenting real estate boom has "[frozen] working-class buyers out of the market" (Jacobs 2011). Meanwhile, the belief that men should provide a home for their bride is deeply rooted in the Chinese culture. Those factors suggest that single women may be more likely to search for a prospective husband who owns an apartment. Moreover, home-ownership serves as a status good in the marriage market since wealth improves one's attractiveness and home-ownership is a more visible form of wealth than others (Wei et al. 2017). As a result, soaring housing prices in China may lead to women placing increasingly greater weight on home-ownership when searching for a prospective husband.

To investigate this question, we make use of two datasets. The first dataset comes from *Jiayuan.com*, the largest dating website in China, and the second dataset contains monthly housing prices data in major Chinese cities after 2010. The *Jiayuan* dataset contains rich information on users' attributes and messaging activity, which enables us to estimate individuals' marital preferences following the methodology of Hitsch et al. (2010a). Our empirical strategy employs users' decision to reply to a first-time message from a contact as a proxy for interest, and examines how this decision varies with observable characteristics of the user as well as the contact.

First, we find that users value home-ownership—compared to non-homeowners, reply rates to homeowners increase by 2.3% for men and 3.3% for women. Moreover, rising house prices produce different effects on preferences for home-ownership across men and women. While a one log-point increase in housing prices in the contact's city of residence drives up women's probability of reply by 0.5%, no similar effects are observed for men.

Then, we decompose those preferences by users' home-ownership status. We find that male and female homeowners are 1.6% and 0.6% more likely to reply to a fellow homeowner respectively, and an increase in housing prices in the contact's city of residence do not produce heterogeneous effects on these preferences for homeowners and non-homeowners. We further examine the effects of surging housing prices on the competitiveness of homeowners in the online dating market—measured by the number of contact-initiated messages received per day, and the income and homeownership of contacts who initiate conversations—and verify that above findings are indeed robust. This suggests that higher housing prices provide male homeowners additional competitive advantage in the marriage market, and this is driven by stronger preferences for home-ownership across all women.

Finally, we investigate whether these effects on preferences translate to differences in equilibrium matching patterns. We define a user to be matched with a contact if the number of message exchanges among them exceeds the 95% percentile for the user. We do not find evidence of assortative matching along home-ownership. Moreover, increases in housing prices in the male partner's city of residence are associated with stronger assortative matching along education, and weaker assortative matching along income. No such effect is observed for increases in housing prices in the female partner's city of residence.

Our study is closely related to the literature on preferences and assortative matching in the marriage market. Becker (1973)'s theory of marriage argues that assortative matching is necessary for the marriage market to arrive at an equilibrium, and findings from ensuing studies reveal that marital matching is assortative along various anthropometric, ethnic and socioeconomic factors (Kalmijn 1998; Blackwell and Lichter 2004; Browning et al. 2014). Assortative matching patterns are driven by both horizontal and vertical preferences. Under horizontal preferences, individuals prefer to be matched with others who are similar to themselves along certain traits. This is consistent with findings that people prefer partners who are similar in age (Buss 1989; Kenrick and Keefe 1992; Buunk et al. 2004; Hitsch et al. 2010b). Under vertical preferences, individuals evaluate others based on the absolute values of certain traits. This is consistent with the observation that people prefer more wealthy partners over less wealthy ones (Belot and Francesconi 2013; Hitsch et al. 2010b).

A second strand of literature studies the underlying search and matching mechanisms of the marriage market. Choo and Siow (2006) and Wong (2003) develop structural models to study equilibrium marital matching patterns. Chiappori et al. (2012) propose a framework that is consistent with several matching processes to study the rates of substitution between multiple spousal characteristics. However, tractability issues required that restrictive assumptions be imposed on matching processes and individual preferences. To overcome modeling limitations, an emerging empirical literature pursues data and approaches to estimate individuals' revealed preference. For instance, Hitsch et al. (2010a,b), Lee (2015) and Huber and Malhotra (2017) estimate marital preferences using data from online dating websites. Kurzban and Weeden (2005), Fisman et al. (2006) and Belot and Francesconi (2013) use data from speed-dating activities to examine gender differences in marital preferences and how meeting opportunities affect mate choices.

Our study contributes to the literature that examines the effect of political and cultural preferences on marital matching. Huber and Malhotra (2017) demonstrate that political attitudes play an important role in dating, and Banerjee et al. (2013) show that preferences for in-caste marriage have little effect on equilibrium matching. We contribute to the literature by providing a unique, micro-scale examination of the marriage market in conjunction with the real estate market. To the best of our knowledge, only a few papers in this literature investigate similar topics. Klein (2017) demonstrates that positive house price shocks decrease the probability of divorce in the United States. Wei et al. (2017) argues that competition for potential spouses motivates individuals to pursue more expensive houses as a status good. They show that an increase in the gender ratio of the pre-marital age cohorts accounts for 30-48% of the rise in real urban housing prices in China from 2003 to 2009.

We perform the opposite exercise to Wei et al. (2017) by examining the spillover effects of real estate market conditions on individuals' marital preferences. Soaring housing prices are likely to bring about stronger substitution from other desirable characteristics towards home-ownership, leading to an increase in the attractiveness of homeowners in the marriage market. This change is nontrivial because homeowners are not exclusively the wealthy, and stronger preferences for home-ownership is not simply a reflection of stronger preferences for income and wealth.[1] Hence, soaring house prices improve the attractiveness of wealthy as well as otherwise mediocre homeowners, and in the meantime, reduce the relative value of desirable characteristics such as education and income.

This paper is structured as follows. Section 2 describes the data and sampling methods. Section 3 outlines the empirical framework. Section 4 presents and discusses estimation results. Section 5 concludes.

2 Data Description and Institutional Details

In this section, we provide a brief overview of the Chinese housing market and *Jiayuan.com*, the online dating website from which user data were collected. We then describe the datasets used in this paper.

2.1 Housing Market and Data

Background. Prior to the mid-1980s, China had no housing market. State-owned enterprises employed a majority of the urban working population, and provided apartments for workers and their families, who then paid heavily subsidized rents. Workers enjoyed lifetime usage of those apartments, but were not permitted to sell them.

Housing reforms began in the mid-1980s, allowing "part of the state-owned housing to be sold to employees at a subsidized price... and workers who left their state jobs to remain in the same housing if they had purchased it" (Iyer et al. 2019). Housing prices were determined by the local governments based on characteristics of the apartments (e.g. size and number of rooms), and characteristics of the buying worker (e.g. age and job tenure). In 1998, the 23rd Decree issued by the State Council required work units to "integrate any implicit housing benefits into employees' salary and households had to buy or rent their residential housing units in the private housing market" (Wu et al. 2012).

[1] For example, men from urban families could have purchased homes when they were much more affordable before the housing market boom. Employees of state-owned enterprises were allowed to purchase homes at a government-subsidized price as part of their employee benefits. It is also not uncommon in China for parents to purchase an apartment and register it under their child's name.

Following the reforms, housing construction and subsequently housing prices began to grow rapidly. The ratio of newly built private housing to the total annual flow supply increased from about 33% to 72% from 1993 to 2007 (Wu et al. 2012). Housing prices has been rising since 1991, leading to a property bubble between 2005 and 2011. From 2002 to 2010, housing prices of 35 major cities in China maintained an average annual growth rate of 12.68%; in large cities such as Beijing and Shanghai, housing prices had an annual growth rate of 15.75% and 14.93% respectively, compared with an national average of 12.68% (Yang and Chen 2014).

Data. Our data on housing prices come from the China Index Academy (CIA), a property market research organization in China. CIA provides monthly housing prices of 100 major cities in China since June 2010.

We categorize Chinese cities into different tiers according to an industry research report produced by Jones Lang La Salle (JLL),[2] a professional services and investment management company in real estate. This classification is based on cities' economic and real estate activities, and has seven categories: ultra tier 1, tier 1, tier 1.5, tier 2, tier 3-growth, tier 3-emerging and tier 3-early adopter.[3] Furthermore, we define three broader categories based on the JLL classification. We define large metropolitan areas to be all cities in the ultra tier 1 and tier 1 categories, medium metropolitan areas to be all cities in the tier 1.5 and tier 2 categories, and small metropolitan areas to be all other cities.

Below, we display in Table 1 the average yearly housing prices based on both the JLL classification of the 60 cities and our broader categories. We can see that there exists considerable variation in house prices across cities of different tiers, and that tier 1 cities experienced the biggest increases in house prices from 2012 to 2015.

2.2 Online Dating and Data Collection

Mechanics of Jiayuan. *Jiayuan.com* is the largest online dating website in China. By June 2014, Jiayuan had more than 110 million registered users and 5.4 million monthly active users.

Jiayuan offers users with free registration. Upon sign-up, users are required to fill out information about their gender, date of birth, marital status, age, education level, monthly income, and to submit a brief self-introduction. Then, users can choose to improve their personal profiles that are viewable to other users. In addition to the information provided during sign-up, users can indicate their geographical and hukou location (i.e. location of permanent residence), whether they

[2] "China60: From Fast Growth to Smart Growth". URL: China60.

[3] The full classification is as follows. Ultra tier 1: Beijing, Shanghai. Tier 1: Guangzhou, Shenzhen. Tier 1.5: Chengdu, Chongqing, Hangzhou, Nanjing, Shenyang, Suzhou, Tianjin, Wuhan, Xi'an. Tier 2: Changsha, Dalian, Jinan, Ningbo, Qingdao, Wuxi, Xiamen, Zhengzhou. Tier 3 growth: Changchun, Changzhou, Dongguan, Foshan, Fuzhou, Guiyang, Harbin, Hefei, Hohhot, Kunming, Nanchang, Nanning, Shijiazhuang, Wenzhou. Tier 3 emerging: Haikou, Huzhou, Jiaxing, Jinhua, Lanzhou, Luoyang, Nantong, Quanzhou, Shaoxing, Taiyuan, Tangshan, Urumqi, Weifang, Xuzhou, Yantai, Zhongshan, Zhuhai. Tier 3 early adopter: Huai'an, Jilin, Mianyang, Weihai, Wuhu, Xiangyang, Xining, Yancheng, Yangzhou, Yichang, Yinchuan, Zibo.

Table 1: Average Prices of Residential Housing (RMB Per Square Meter)

City Classification	2012	2013	2014	2015
Panel (a): Jones Lang LaSalle				
Ultra tier 1	25,390	28,677	32,428	32,571
Tier 1	18,922	22,206	24,351	24,697
Tier 1.5	9,634	10,145	10,370	9,923
Tier 2	9,857	10,483	10,909	10,701
Tier 3-growth	8,511	8,654	8,632	8,234
Tier 3-emerging	7,593	7,777	7,773	7,465
Tier 3-early adopter	5,744	5,945	5,985	5,813
Cities not included in JLL	6,490	6,739	6,782	6,434
Panel (b): Three Broad Categories				
Large metropolitan areas	22,156	25,442	28,390	28,634
Medium metropolitan areas	9,745	10,314	10,640	10,312
Small metropolitan areas	7,085	7,279	7,293	6,987

Notes: This table displays average prices of residential apartments, in RMB/m^2, for years 2012 to 2015. Panel (a) displays average prices within each JLL-classified tier. Panel (b) displays average prices within the three metropolitan areas defined based on the JLL classifications.

have children, religion, home-ownership and car-ownership. In particular, home-ownership consists of the following subcategories: owning a home with or without mortgage, planning to purchase a house, living with parents or relatives, renting an apartment alone or with friends, or living in an enterprise-owned apartment. Users can also provide more details on their financial status, lifestyle, employment, and hobbies and interests.

Although users can start sending out messages immediately after registration, better services require subscription fees. For instance, in order for an initial message to be readable by the recipient, either the sender or the recipient needs to buy a virtual stamp. Before a user clicks into an incoming message, the website displays to the her some basic information about the contact, including a photo, age, height, city of residence, income and house ownership.

To incentivize its members to report truthful information, Jiayuan maintains a credit level for each user, which is visible in the user's profile page. A user can improve her credit level by uploading verification documents such as college diplomas or government-issued ID cards. For the purpose of this paper, it is important to know that users provide truthful information and that they indicate preferences for a prospective husband or wife rather than a casual relationship. We believe this is likely to be true since the subscription fee requirement would likely deter less serious users. In addition, since we have information on users' credit level, we can ensure the reliability of information by limiting the sample to users with high credit levels.

Data. Our sample contains 10,069 users who became registered after April 2012 on Jiayuan.

Section A in the Appendix provides a detailed description of the data sampling procedure we used. Table 2 displays the composition of our sample of users. We split the users' location of residence into large, medium and small metropolitan areas to exploit variations in housing prices. Users are aged between 22 and 40 years old, and have complete information on their house-ownership status. We select up to 250 *contacts* for each user. Contacts are defined as either conversation initiators or first-time responders to messages sent by the selected users. We end up with 456,916 user-contact pairs for male users and 1,138,326 user-contact pairs for female users.

Table 2: Geographical Distribution of Users by Gender and Home-Ownership

User Type	Large Cities	Medium Cities	Small Cities
Panel (a): Homeowners			
Male	848	851	851
Female	716	849	849
Panel (b): Non-homeowners			
Male	851	849	850
Female	854	850	849

Notes: This table displays the number of users by gender, home-ownership and geographical location in our data sample. Panel (a) shows the geographical distribution of homeowners by gender, and panel (b) shows the geographical distribution of non-homeowners by gender.

In Table 5, we present a summary of the distribution of personal characteristics for all users and contacts in the sample, by gender and home-ownership. We provide distributions by brackets because the original information that Jiayuan asks from their users are organized in brackets.

Age. Since our data sampling procedure limits the age of target users to be between 22 and 40 years old, and online dating website users are young in general, more than 70% of the individuals in our sample are between 22 and 35 years old. The share of users who are above 30 years old is much higher for homeowners than for non-homeowners, and this difference is larger for women.

Marital Status and Children. The majority of users are single and do not have children. Compared to women, a higher fraction of men are divorced and/or have children, and this difference is more pronounced among non-homeowners.

Education. The percentage of users with a college degree or above is higher for homeowners than for non-homeowners, and women's overall education level is slightly higher than men. More specifically, for men, the percentage with bachelor's degrees or above is 42.67% among non-homeowners and the percentage is 62.52% among homeowners; for women, the percentage with bachelor's degrees or above is 46.49% among non-homeowners and 63.1% among homeowners.

Income. In general, the percentage of high-income earners (with monthly income greater than

10,000 RMB) is higher among men than women, and the percentage of both medium- (with monthly income between 5,000 and 10,000 RMB) and high-income earners is higher among homeowners than non-homeowners. More specifically, for men, the percentage of high-income earners is 12.71% among non-homeowners and 34.73% among homeowners; for women, the percentage of high-income earners is 5.9% among non-homeowners and 21.41% among homeowners.

Car Ownership. The percentage of car-owners is higher for men than women, and for homeowners than non-homeowners.

Lastly, we present summary statistics on user activity in Table 6. Figure 1 show the distribution of reply rate and average daily contacts. Panel A in Table 6 summarizes users' reply rate, which is defined by the number of contacts that a user replies to divided by her total number of contacts. The average reply rate is 53.5% for men and 38.5% for women. Panels B and C summarize the total number of contacts per user and the average number of contacts per month per user, which is calculated as the total number of contacts over the number of months elapsed. It is important to remember that the number of contacts is truncated at 250 per user because we only sampled up to 250 contacts for each user. The average number of contacts per month of a user may exceed 250 if the user gains 250 contacts within a short period of time.

Panels D and E summarize the total number of self-initiating contacts per user and the average number of self-initiating contacts per month per user. The means are much higher for females than for males. This is reasonable since on dating websites, men actively send out more messages than women, and it is natural that women should receive more messages than men. Panel F displays the average months elapsed for a user to gain 250 contacts. If a user's total number of contacts is less than 250, then this statistic shows the user's total active months.

3 Empirical Strategy

In this section, we first present and discuss the framework for estimating individual's marital preferences. We then present the regression specifications for examining individuals' competitiveness in the online dating market and the equilibrium matching patterns.

3.1 Mate Preference Estimation

Our dataset includes characteristics and messaging activity records of 10,069 site users, as well as information of up to 250 contacts of each user. A user's pool of contacts contains both conversation initiators and first-time responders. We define this set of contacts as a user's choice set of potential partners, and denote it by S_u.

Now, let $U(u, c)$ be user u's expected utility from a potential match with contact c, and let v_u

be user u's reservation utility, which is his utility from the outside option of not keeping in touch with c and is thus independent of c's characteristics. Suppose that u receives an initial introductory message or a first reply from a contact $c \in S_u$, he will decide to continue this conversion with c if and only if

$$U(u,c) \geq v_u. \tag{1}$$

This threshold rule follows from Hitsch et al. (2010a,b), who point out that it arises from two-sided search models in the marriage market, and captures the search process for a partner in the online dating market as well as the idea that online dating site users understand their own values, which is reflected by their reservation utility. They assume that Equation 1 applies when u browses c's profile and chooses to send an email,[4] and argue that the "the threshold rule may not hold if sending a first-contact e-mail is costly or if there is a psychological cost of being rejected" (Hitsch et al. 2010b). Since we examine the decision of continuing a conversation rather than starting a conversation, this psychological cost would not be a threat to the threshold rule in our context.

We assume that the latent utility that a user u gets from being with a potential mate c depends on two factors: a vector of his own characteristics x_u and a vector of the contact's characteristics x_c. More specifically, $x_u = (age_u$, $height_u$, $educ_u$, $mStatus_u$, $child_u$, $locale_u$, $industry_u$, $income_u$, $house_u)$ and similar for x_c, where the eight components represent age, height, level of education, marital status (single or divorced), whether or not the individual has a child, hukou status in the city of residence, industry of occupation, level of monthly income, and home-ownership status. Following (Hitsch et al. 2010a), our utility function takes the form in Equation 2, accounting for both horizontal and vertical preferences.

The first line displays a user's preference for age. We assume that this preference depends on the contact's age and the age difference between the user and the contact, where $\left\{\mathrm{G}_i^{\mathrm{age}}\right\}_{i=1}^{6}$ is a set of age-difference categories and the omitted category is $|age_u - age_c| \leq 1$. For instance, $\mathbb{1}\left\{age_u - age_c \in \mathrm{G}_1^{\mathrm{age}}\right\} = \mathbb{1}\left\{age_u - age_c \leq -10\right\}$, and its coefficient estimate represents the user's preference for a contact who is more than 10 years older than her relative to a contact who is similar to her in age. Similarly, we assume that preferences for height and BMI also depend on the contact's height and BMI, as well as height and BMI differences.

[4] Hitsch et al. (2010a,b) define a user's choice set to be all contacts whose profiles she has viewed. We define the choice set more restrictively because we do not have information on users' browsing history.

$$
\begin{aligned}
U(x_u, x_c; \theta) =& \alpha \cdot age_c + \sum_{i=1}^{6} \mathbb{1}\left\{age_u - age_c \in \mathrm{G}_i^{\text{age}}\right\} \\
&+ \beta \cdot height_c + \sum_{i=1}^{6} \mathbb{1}\left\{height_u - height_c \in \mathrm{G}_i^{\text{height}}\right\} \\
&+ \sigma \cdot BMI_c + \sum_{i=1}^{6} \mathbb{1}\left\{BMI_u - BMI_c \in \mathrm{G}_i^{\text{BMI}}\right\} \\
&+ \sum_{j,k \in \mathrm{G}^{\text{educ}}, j \neq k} \mathbb{1}\left\{educ_u = j \text{ and } educ_c = k\right\} \\
&+ \sum_{j,k \in \mathrm{G}^{\text{mStatus}}, j \neq k} \mathbb{1}\left\{mStatus_u = j \text{ and } mStatus_c = k\right\} \\
&+ \sum_{j,k \in \mathrm{G}^{\text{child}}, j \neq k} \mathbb{1}\left\{child_u = j \text{ and } child_c = k\right\} \\
&+ \sum_{j,k \in \mathrm{G}^{\text{locale}}, j \neq k} \mathbb{1}\left\{locale_u = j \text{ and } locale_c = k\right\} \\
&+ \sum_{i=1}^{18} industry_c^i + \lambda \cdot income_c + \pi \cdot house_c
\end{aligned} \tag{2}
$$

The fourth to seventh lines display a user's preferences for education level, marital status, whether or not the contact has a child, and whether the contact has hukou in the user's city of residence (i.e. whether the contact is a permanent resident of this city). We assume that the preferences depend on the characteristics of both the user and the contact. Here, $\mathrm{G}^{\text{educ}} = \{$*senior vocational school*, *vocational school*, *college*, *master* and *PhD/higher*$\}$, $\mathrm{G}^{\text{mStatus}} = \{$*single*, *divorced*$\}$, $\mathrm{G}^{\text{child}} = \{$*has children*, *no children*$\}$, $\mathrm{G}^{\text{locale}} = \{$*has hukou*, *no hukou*$\}$, and the omitted categories are those where the user and the contact have similar characteristics. For instance, the coefficient estimate for the term $\mathbb{1}\left\{educ_u = college \text{ and } educ_c = master\right\}$ represents the preference of a user with a college degree for a contact with a masters degree relative to a contact with a college degree, holding all else constant.

The first term in the last line accounts for a user's preferences towards a contact's occupational industry, which is divided into 18 groups, and the omitted group of occupation is student. The remaining two terms account for a user's preferences for income and home-ownership, which are assumed to be vertical—that is, everyone prefers a potential mate to have higher income and to own a home.

For simplicity of interpretation, we estimate preference parameters using a discrete choice model:

$$
reply_{uc} = U(x_u, x_c; \theta) + Z_{uc}\tau + \psi_u + \delta_t + \phi_{city} + \epsilon_{uc}, \tag{3}
$$

where $reply_{uc}$ is a dummy variable that is equal to 1 if u replies to c's message and zero otherwise.

Z_{uc} includes other variables that may affect a user's reply decision, including whether the contact and the user live in the same city, whether the contact and the user live in the same province but in different cities, whether the conversion is user-initiated, and the number of contacts that the user had during the past month. ψ_u is a user fixed effect that accounts for reservation utility v_u, δ_t is year fixed-effect, and ϕ_{city} is the contact's city fixed-effect.

We estimate Equation 3 and all of its extensions with weights to account for oversampling of homeowners in our data selection process. We calculate those weights using the actual proportions of males and females by home-ownership and metropolitan area in Jiayuan's full database, and their counterparts in our sample data.

3.2 Users' Competitiveness

To complement the analysis on the impact of contacts' home-ownership and housing prices on users' preferences, we also investigate the effect of users' home-ownership and housing prices on their competitiveness in the marriage market. This is because the estimation of Equation 3 is based on a restricted choice set of self-initiated contacts and first-time respondents, and it is important to see whether the same results hold when individuals choose contacts to send first-time messages.

First, we measure a user's competitiveness by the average number of contact-initiated messages received by user u per day, $FreqMsg_u$. This is the number of contacts who initiated a conversation with the user, divided by the number of days between the first and the last time the user received a message from a self-initiating contact. We use the following specification:

$$FreqMsg_u = \alpha + \beta_1 house_u + \beta_2 \log P_{t-1,u} + \beta_3 house_u \cdot \log P_{t-1,u} + X_u\gamma + \epsilon_u, \tag{4}$$

where $house_u$ is a dummy variable that is equal to 1 if user u owns a house, $P_{t-1,u}$ is one-month lagged housing price of u's city of residence, and X_u is a vector of the user's characteristics.

Secondly, we measure a user's competitiveness with D_{uc}, the "desirability" of the users' self-initiating contacts. We measure desirability by contacts' monthly income and home-ownership, since these two attributes are vertical. We propose the following specification:

$$D_{uc} = \alpha + \beta_1 house_u + \beta_2 \log P_{t-1,u} + \beta_3 house_u \cdot \log P_{t-1,u} + X_u\gamma_1 + X_c\gamma_2 + \delta_t + \phi_{city} + \epsilon_{uc}, \tag{5}$$

where X_u and X_c are vectors of characteristics of the user and the contact, respectively.

3.3 Assortative Matching Patterns

Recall that our estimation strategy of users' marital preferences is based on the initial reply decisions of users, but a reply does not necessarily lead to a match between a user and a contact. To investigate the assortative matching patterns, we have to define a "match" in this environment. Since we know neither the content of messages exchanged between users and contacts, nor their activities offline, we define a match between a user and a contact based on the number of message exchanges. More specifically, a user and a contact are *matched* if, between the pair, both the average number of message exchanges per day and the total number of rounds of messages exceed the 95^{th} percentile for the user. For robustness checks, we do the same analysis using the 97.5^{th} percentile.

We investigate assortative matching patterns in education level, income and home-ownership by estimating the correlations of those attributes in matched couples. Since many attributes of an individual are correlated, we make sure to control for other characteristics of matched couples. We estimate the following equation for observed matches:

$$\begin{aligned} char_g = \alpha + \beta_1 char_{-g} + \beta_2 \log P_{t-1,-g} + \beta_3 char_{-g} \cdot \log P_{t-1,-g} \\ + X_g \gamma_1 + X_{-g} \gamma_2 + \epsilon_{g,-g}, \end{aligned} \tag{6}$$

where $char \in \{educ,\ income,\ house\}$, $g \in \{f,\ m\}$, and $-g \in \{f,\ m\} \setminus \{g\}$, with f denoting the female and m denoting the male in a matched couple. The coefficient of interest is β_1 when housing prices are not included, and it represents whether matching outcomes reflect sorting in education, income, or home-ownership. Then we control for housing price terms and the coefficient β_3 captures changes in matching patterns in response to increases in housing prices.

4 Estimation Results and Discussion

4.1 Marital Preferences: Anthropometric and Socioeconomic Factors

First, we estimate users' marital preferences and examine gender differences. We do so by estimating Equation 3 for men and women separately. Then, we interact each term in Equation 3 with a dummy variable $female_u$, to test the hypothesis that women have stronger preferences for house-ownership than men do. Thus, we are interested in the coefficients on $house_c$ and $house_c \cdot female_u$.

Table 7 displays the full set of estimates from a linear probability model for men and women separately. Columns 1 and 2 show preference estimates for men and women without house price effects, and columns 3 and 4 show preference estimates when interaction terms with housing prices are added. Table 8 displays the estimates of equation (2) from a linear probability model, with the addition of gender interaction terms. Again, Columns 1 and 2 show preference estimates without housing prices, and Column 3 and 4 show the estimates with housing price interaction terms.

Age. We assume that preferences for age depend on the age of the contact as well as the age difference between the contact and the user. Results show that men generally favor women who are younger than them and disfavor women who are older than them. In contrast, women favor men who are similar in age and disfavor men who are too young or too old.

Height. We assume that preferences for height depend on the height of the contact as well as the height difference between the user and the contact. Unsurprisingly, men are interested in women who are shorter than them and dislike women who are taller than them. On the other hand, women have a strong preference for men who are taller than them.

Body Mass Index. We assume that BMI preferences depend on the BMI of the potential mate (contact) as well as the BMI difference between the user and the contact. As expected, men show more interest in women with a lower BMI than their own, while women exhibit a slight preference for men whose BMI are higher than their own.

Education. We allow education preferences to depend on both the education level of the user and the education level of the contact. Men display a strong preference towards women who have a similar education level, and a dislike for women with a higher education level. On the other hand, all women prefer men who have a higher education level than their own.

Income. Both men and women prefer a high-income contact over a low-income contact. In addition, women value a contact's income level more than men.

4.2 Marital Preferences: Home-Ownership

Now, we examine individuals' preferences for home-ownership and the effect of housing prices on preferences. We test whether users positively respond to rising housing prices. To do so, we include interaction terms of home-ownership and the logarithm of one-month lagged average housing prices in contact c's city of residence, and we are interested in the coefficient on $house_c \cdot \log P_{t-1,c}$.

Table 3 replicates parts of Tables 7 and 8 that are related to home-ownership and housing prices. Panel A display coefficient estimates for men and women separately, whereas Panel B display estimates for gender differences. Columns 1 and 2 show results without housing prices, and columns 3 and 4 show results with housing prices.

We can see that both genders strictly prefer a potential mate who owns a home, and women have stronger preferences for home-ownership than men do. More specifically, men are 2.3% more likely to reply to a contact who owns a home, up from an average reply rate of 28.6%; women are 3.3% more likely to reply to a contact who owns a home, up from an average rate of reply of 18.3%. In addition, regression on gender differences reveals that women are 0.8% more likely than men to reply to a contact who owns a home. All effects are statistically significant at the 5% level. The fact that women value partner's home-ownership status more strongly is consistent with China's cultural tradition that men's family should provide a home for the bride. It also suggests that home-ownership improves the competitiveness of men more than it does for women.

Next, we examine whether preferences for home-ownership changes with housing prices. We find that women respond positively to rising house prices: when the average housing prices of a homeowner's city of residence rise by 1 log point, women are 0.5% more likely to respond and this effect is statistically significant at the 5% level. The regression on gender differences reproduces the size of this effect, but larger standard errors make this result statistically insignificant. On the other hand, housing prices have a negative effect on the likelihood of reply from men, though this effect is tiny in size and statistically insignificant. It suggests that rising house prices make male homeowners more attractive in the marriage market, but not for female home-owners.

Table 3: Effects of Housing Prices on Preferences for Home-Ownership

	(1)	(2)	(3)	(4)
Variables	Whether User Replies to Message			
Panel (a): Full Effects				
	Men	Women	Men	Women
House_c	0.023***	0.033***	0.037	-0.015
	(0.004)	(0.001)	(0.059)	(0.021)
$\text{House}_c \cdot \log P_{t-1,c}$			-0.001	0.005**
			(0.006)	(0.002)
Obs	192,723	692,364	170,909	588,529
R-squared	0.025	0.043	0.019	0.041
Dep Mean	0.286	0.183	0.286	0.183
Dep SD	0.452	0.387	0.452	0.387
Panel (b): Gender Differences				
	Men	Women: Diff	Men	Women: Diff
House_c	0.023***	0.010**	0.036	-0.053
	(0.004)	(0.004)	(0.058)	(0.062)
$\text{House}_c \cdot \log P_{t-1,c}$			-0.002	0.007
			(0.006)	(0.007)
Obs	885,087		759,438	
R-squared	0.034		0.032	
Dep Mean	0.212		0.212	
Dep SD	0.409		0.409	

Notes: *** $p < .01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors are clustered at user level. This table presents OLS regression results for the effect of home-ownership and housing prices on user's decision to reply. Panel (a) displays estimation results for male and female users separately. Panel (b) displays estimation results with coefficients on male users and on female-male differences. An observation is a user-contact pair. Dependent variable is a dummy that is equal to 1 if the user replies, and 0 otherwise. House_c is equal to 1 if contact c is a homeowner, and 0 otherwise. $P_{t-1,c}$ is 1-month lagged average housing price of contact c's city of residence.

We decompose the housing price effect by subpopulations. In particular, we investigate whether preferences for home-ownership and the effect of housing prices differ among homeowners and non-homeowners. To do so, we add interaction terms of contacts' home-ownership status and housing prices with user's home-ownership status, $house_u$, to Equation 3, and re-estimate it using a linear probability model.

Table 3 displays relevant estimates for men and women separately. Compared to male non-homeowners, male homeowners are 1.6% more likely to reply to a fellow homeowner; compared to female non-homeowners, female homeowners are 0.6% more likely to reply to a fellow homeowner. Both effects are statistically significant at the 5% level. This suggests that homeowners have stronger preferences for homeownership than non-homeowners. As for housing price effects, when the average housing prices of a homeowner's city of residence rise by 1 log point, a male homeowner is 1% less likely to reply to this contact compared to a male non-house-owner, and a female homeowner is 0.1% more likely to reply to this contact. Both effects are statistically insignificant.

Table 4: Effect of Housing Prices on Preferences for Home-Ownership: Homeowners vs. Non-Homeowners

Variables	(1) Whether User Replies to Message	(2)	(3)	(4)
	Men	Women	Men	Women
$\text{house}_c \cdot \text{house}_u$	0.016**	0.006***	0.115	0.002
	(0.008)	(0.002)	(0.120)	(0.036)
$\log P_{t-1,c} \cdot \text{house}_c$			0.005	0.006**
			(0.011)	(0.003)
$\log P_{t-1,c} \cdot \text{house}_u$			-0.000	-0.004
			(0.007)	(0.003)
$\log P_{t-1,c} \cdot \text{house}_c \cdot \text{house}_u$			-0.010	0.001
			(0.012)	(0.004)
Obs	192,723	692,364	170,909	588,529
R-squared	0.021	0.035	0.015	0.033
Dep Mean	0.286	0.183	0.286	0.183
Dep SD	0.452	0.387	0.452	0.387

Notes: *** $p < .01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors are clustered at user level. This table presents OLS regression results for the heterogeneous effects of housing prices on the decision to reply by homeowners and non-homeowners. An observation is a user-contact pair. Dependent variable is a dummy that is equal to 1 if the user replies, and 0 otherwise. House_c is equal to 1 if contact c is a homeowner, and 0 otherwise. House_u is equal to 1 if user u is a homeowner, and 0 otherwise. $P_{t-1,c}$ is 1-month lagged average housing price of contact c's city of residence. Columns (1) and (3) display results for male users, and columns (2) and (4) display results for female users.

Furthermore, note that the coefficient on $\text{House}_c \cdot \log P_{t-1,c}$ is positive and statistically significant, and has a very similar magnitude to the price effects for female users in Table 3. Thus, these results seem to suggest that the response of women's preferences for home-ownership to rising housing prices is not driven by the non-homeowners, but is uniform across all female users. Women seem to be valuing home-ownership more, as the relative difficulty of purchasing a home increases.

4.3 Competitiveness of Homeowners

To verify that results on preferences hold more generally, we now examine whether home-ownership improves a user's competitiveness in general. We estimate Equations 4 and 5 for men and women separately, first looking at home-ownership by itself, and then controlling for housing price terms. Relevant estimates are shown in Table 9 for men and women separately.

Panel A uses the average number of contact-initiated messages that a user receives per day as a proxy for her competitiveness. That is, it assumes that a more competitive user should attract more attention. For male users, having a house increases the average number of messages per day by 0.171; for female users, the increase is 1.342. However, the effect for males is not statistically significant. Turning to the effects of housing prices of the user's city of residence, we see that for male homeowners, a 1 log-point rise in housing price is associated with an increase in the number of contact-initiated messages per day by 0.835; for female homeowners, the effect is 0.578. This result is statistically significant at the 1% level for men, but insignificant for women. Consistent with findings in Section 4.2, we see that rising housing prices make male homeowners more attractive.

Panel B measures a user's competitiveness by the monthly income brackets[5] of conversation-initiators. The average monthly income bracket of contacts is 23 for male users and 29 for female users. For male users, home-ownership increases the contact's income bracket by 0.039; for female users, the increase is 0.09. Both results are statistically insignificant and very small given the measurement of income brackets. As for housing price effects, for male homeowners, a 1 log-point increase in the housing price of his city of residence is associated with an increase in his contact's income bracket by 0.414; for a female homeowner, this increase is 0.118. The house price effect is significant at the 1% level for males but insignificant for females. This suggests that rising house prices improves male house-owners' appeal to higher-income females, but the same effect does not exist for female house-owners.

Lastly, Panel C measures the competitiveness of a user by the home-ownership status of conversation-initiators. We observe that the average percentage of contacts who have a house is 12.9% for male users and 52.2% for female users. For men, home-ownership increases the probability of receiving contact from a fellow homeowner by 0.4%, and this effect is statistically insignificant. For women, home-ownership increases the probability of receiving contact from a fellow homeowner by 0.8%, and this effect is statistically significant at the 1% level. For both male and female home-owners, increases in housing prices produce no effect on the probability of their contacts being a homeowner. This result is consistent with our findings in Section 4.2 that house prices do not induce heterogeneity in preference for home-ownership across homeowners and non-homeowners.

[5] Bracket 10: Less than 2,000 RMB; 20: 2,000-5,000 RMB; 30: 5,000-10,000 RMB; 40: 10,000-20,000 RMB; 50: 20,000-50,000 RMB; 60: More than 50,000 RMB.

4.4 Assortative Matching Patterns

Lastly, we examine whether changes in preferences induced by housing prices lead to actual changes in equilibrium matching patterns. We estimate Equation 6 using OLS, and coefficient estimates for $\beta_1 - \beta_3$ are shown in Table 10.

Panels (a), (b) and (c) display assortative matching results along the dimensions of education level, income and home-ownership, respectively. We find a weak degree of assortative matching along education[6] and income, with a correlation of 0.143 and 0.081 respectively. We do not find evidence of assortative matching along home-ownership.

A one log-point rise in housing prices in the male partner's city of residence is associated with higher correlation in the education levels between the matched couple, and lower correlation in their income levels. This effect is statistically significant under the 97.5-percentile definition of matches, but not significant under the 95-percentile definition. Increases in housing prices in the female partner's city of residence produce statistically no effect on equilibrium matching patterns. Moreover, housing prices seem to induce a decrease in the correlation in home-ownership status among matched couples, but the effect is statistically insignificant.

5 Concluding Remarks

Our paper analyzes the effects of China's booming real estate market on its marriage market, using data from China's largest dating website. We first estimate the effects of housing prices on individual's marital preferences, with a particular focus on preference for home-ownership and heterogeneity in preference for home-ownership across genders. We find that individuals exhibits a preference for home-ownership, and women display stronger preferences than men. In addition, rising house prices produce an asymmetric effect on preferences for home-ownership. That is, higher prices elicit stronger preferences for home-ownership from women, but not from men. Breaking down those preferences by users' home-ownership status, we find that, homeowners display stronger preferences for home-ownership compared to non-homeowners, and they are unresponsive to changes in housing prices. Taken together, our results suggest that rising house prices give male homeowners an advantage in the marriage market, and this result is not driven by desperation from female non-homeowners in the face of higher housing prices.

To verify the robustness of these results, we estimate the effects of housing prices on homeowners' competitiveness in the marriage market. We use three proxies to measure a user's competitiveness: (1) the average number of contact-initiated messages she receives per day, (2) the monthly income bracket and (3) the home-ownerships status of the contacts who actively message her. Our results suggest that higher housing prices increase the competitiveness of male homeowners under the first

[6] Education level is measured in brackets. Bracket 1: senior vocational school; 2: vocational school; 3: college; 4: master; 5: PhD.

two measures, and have no effects for women homeowners.

Lastly, using the number of message exchanges to define matches, we find no evidence of positive assortative matching along the home-ownership dimension. In addition, higher housing prices in the male partner's city of residence are associated with stronger sorting along education and weaker sorting along income, while no such effects are observed for housing prices in the female partner's city of residence.

References

Banerjee, Abhijit, Esther Duflo, Maitreesh Ghatak, and Jeanne Lafortune. 2013. "Marry for What? Caste and Mate Selection in Modern India." *American Economic Journal: Microeconomics*, 5(2): 33–72.

Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy*, 81(4): 813–846.

Belot, Michele, and Marco Francesconi. 2013. "Dating Preferences and Meeting Opportunities in Mate Choice Decision." *Journal of Human Resources*, 48(2): 474–508.

Blackwell, Debra L., and Daniel T. Lichter. 2004. "Homogamy among Dating, Cohabiting, and Married Couple." *The Sociological Quarterly*, 45(4): 719–737.

Browning, Martin, Pierre-Andre Chiappori, and Yoram Weiss. 2014. *Economics of the Family*. Cambridge: Cambridge University Press.

Buss, David M. 1989. "Sex Differences in Human Mate Preferences: Evolutionary Hypotheses Tested in 37 Cultures." *Behavioral and Brain Sciences*, 12(1): 1–49.

Buunk, Bram P., Pieternel Dijkstra, Douglas T. Kenrick, and Astrid Warntjes. 2004. "Age Preference for Mates as Relate to Gender, Own Age, and Involvement Level." *Evolution and Human Behavior*, 22(4): 241–250.

Chiappori, Pierre-Andre, Sonia Oreffice, and Climent Quintana-Domeque. 2012. "Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market." *Journal of Political Economy*, 120(4): 659–695.

Choo, Eugene, and Aloysius Siow. 2006. "Who Marries Whom and Why." *Journal of Political Economy*, 114(1): 175–201.

Fisman, Raymond, Sheena S. Iyengar, Emir Kamenica, and Itamar Simonson. 2006. "Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment." *Journal of Political Economy*, 121(2): 673–697.

Hitsch, Gunter J., Ali Hortacsu, and Dan Ariely. 2010a. "Matching and Sorting in Online Dating." *American Economic Review*, 100(1): 130–163.

Hitsch, Gunter J., Ali Hortacsu, and Dan Ariely. 2010b. "What Makes You Click?—Mate Preferences in Online Dating." *Quantitative Marketing and Economics*, 8(4): 393–427.

Huber, Gregory A., and Neil Malhotra. 2017. "Political Homophily in Social Relationships: Evidence from Online Dating Behavior." *Journal of Politics*, 79(1): 269–283.

Iyer, Lakshmi, Xin Meng, Nancy Qian, and Xiaoxue Zhao. 2019. "Economic Transition and Private-Sector Labor: Evidence from urban China." *Forthcoming in Journal of Comparative Economics.*

Jacobs, Andrew. 2011. "For Many Chinese Men, No Deed Means No Dates." *The New York Times*, URL: https://www.nytimes.com/2011/04/15/world/asia/15bachelors.html?_r=0.

Kalmijn, Matthijs. 1998. "Intermarriage and Homogamy: Causes, Patterns, and Trends." *Annual Review of Sociology*, 81 395–421.

Kenrick, Douglas T., and Richard C. Keefe. 1992. "Age Preferences in Mates Reflect Sex Differences in Human Reproductive Strategies." *Behavioral and Brain Sciences*, 15(1): 75–133.

Klein, Jennifer. 2017. "House Price Shocks and Individual Divorce Risk in the United States." *Journal of Family and Economic Issues*, 38(4): 628–649.

Kurzban, Robert, and Jason Weeden. 2005. "HurryDate: Mate Preferences in Action." *Evolution and Human Behavior*, 26(3): 227–244.

Lee, Soohyung. 2015. "Effect of Online Dating on Assortative Mating: Evidence from South Korea." *Journal of Applied Econometrics*, 31(6): 1120–1139.

Wei, Shang-Jin, Xiaobo Zhang, and Yin Liu. 2017. "Home Ownership as Status Competition: Some Theory and Evidence." *Journal of Development Economics*, 127(C): 169–186.

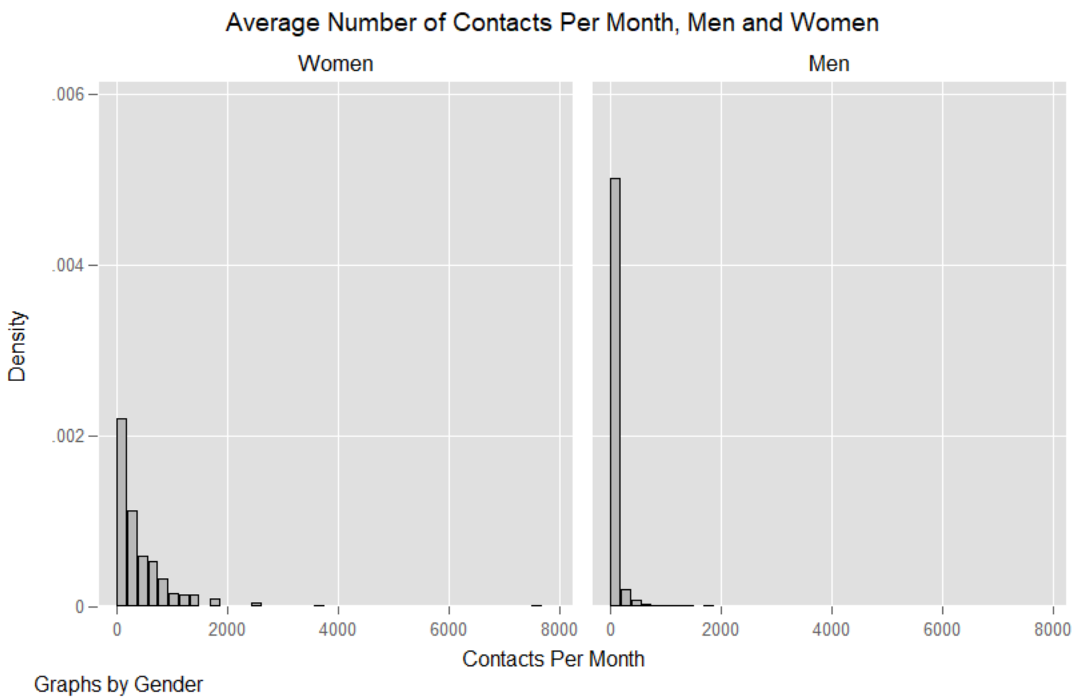
Wong, Linda Y. 2003. "Structural Estimation of Marriage Models." *Journal of Labor Economics*, 21(3): 699–727.

Wu, Jing, Joseph Gyourko, and Yongheng Deng. 2012. "Evaluating Conditions in Major Chinese Housing Markets." *Regional Science and Urban Economics*, 42(3): 531–543.

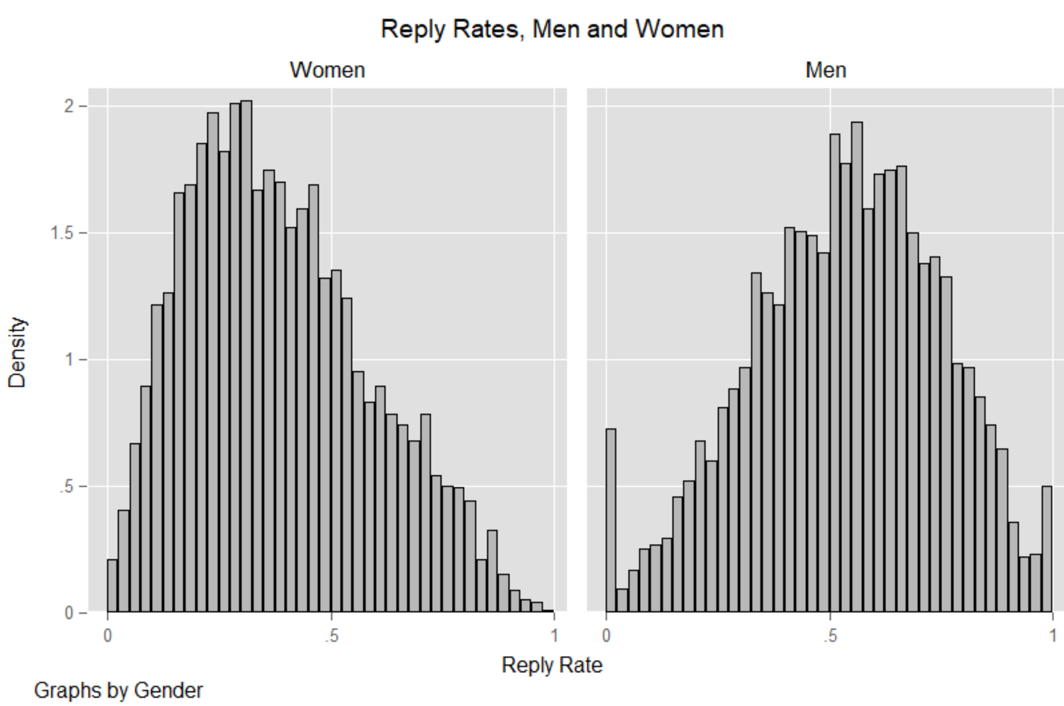
Yang, Zan, and Jie Chen. 2014. *Housing Affordability and Housing Policy in Urban China.* Berlin Heidelberg: Springer-Verlag.

Figure 1: Distribution of Users' Average Monthly Contacts and Reply Rate

(a) Average Monthly Contacts By Gender



(b) Reply Rates By Gender



Note: This figure presents the distribution of two variables. Panel (a) shows the distribution of the average number of contacts per month, for women and men respectively. Panel (b) shows the distribution of the average reply rate to incoming messages, for women and men respectively.

Table 5: Distribution of Personal Characteristics in the User Sample

	No Home		Owns Home	
Variables	Women	Men	Women	Men
Age				
22-	17.38	8.26	6.81	2.71
22-25	46.94	32.36	26.95	17.88
26-30	28.51	32.5	36.48	30.05
31-35	5.89	14.35	20.7	21.18
36-40	1.03	7.61	6.85	14.64
41-50	0.21	4.65	1.84	12.58
51-60	0.03	0.24	0.31	0.88
61+	0.02	0.03	0.06	0.08
Height				
155cm-	6.22	0.37	3.46	0.1
156cm-160cm	34.25	1.26	28.38	0.49
161cm-165cm	38.38	6.61	42.37	3.56
166cm-170cm	17.27	31.36	21.71	22.48
171cm-175cm	3.03	33.88	3.3	36.97
176cm-180cm	0.75	20.3	0.7	28.21
181cm-190cm	0.04	6.03	0.04	8.08
191cm+	0.06	0.19	0.04	0.13
BMI				
18-	16.06	1.69	16.60	0.66
18-20	28.93	9.31	30.12	4.75
20-22	11.9	23.22	13.61	19.5
22-24	3.38	19.62	3.95	25.9
24-26	0.97	9.97	1.02	15.57
26-28	0.33	3.36	0.28	4.61
28-32	0.28	1.39	0.19	1.52
32+	38.15	31.44	34.22	27.49
Marital Status				
Single	90.78	82.07	68.87	65.09
Divorced	8.94	17.2	29.47	32.19
Widow	0.27	0.73	1.66	2.72
Has Children				
No	93.14	86.15	76.36	74.64
Yes	6.86	13.85	23.64	25.36
Education Level				

Secondary Vocational School	20.62	29.48	11.43	13.92
Vocational School	32.89	27.86	25.47	23.56
College	42.36	38.79	56.32	54.45
Master	3.31	3.39	5.93	7.14
PhD or Higher	0.82	0.49	0.85	0.93
Monthly Income (RMB)				
2,000-	12.02	6.77	3.85	1.73
2,000-5,000	57.19	47.82	35.39	27.32
5,000-10,000	24.9	32.70	39.36	36.22
10,000-20,000	4.43	9.52	14.67	19.83
20,000-50,000	1.47	3.19	6.74	14.9
Car Ownership				
No	88.35	82.27	54.94	43.79
Yes	11.65	17.73	45.06	56.21
Metropolitan Areas				
Large	31.66	24.68	29.84	22.18
Medium	29.56	25.2	35.78	31.62
Small	38.79	50.12	34.37	46.2
Obs.	212,824	267,755	33,506	263,232

Notes: This table presents distributions of anthropometric and socioeconomic characteristics of users, by gender and home-ownership. The first two columns display statistics for non-homeowners, and the last two columns display statistics for homeowners. Odd-numbered columns show results for female users, and even-numbered columns show results for male users.

Table 6: Summary Statistics on User Activity By Gender

	Obs.	Mean	SD	Min	Max
Panel (a): Reply rate					
Men	5,076	0.58	0.22	0	1
Women	4,969	0.42	0.22	0	1
Panel (b): Total number of contacts					
Men	5,076	90.01	82.67	1	250
Women	4,969	229.09	47.57	3	250
Panel (c): Average number of contacts per month					
Men	5,076	56.55	106.68	0.09	1,875
Women	4,969	423.05	484.95	1.28	7,500
Panel (d): Total number of self-initiating contacts					
Men	5,041	72.95	68.59	1	248
Women	4,969	209.3707	48.49	2	250
Panel (e): Average number of self-initiating contacts per month					
Men	5,041	47.96	92.84	0.08	1,710
Women	4,969	391.88	456.42	0.62	7,260
Panel (f): Months elapsed					
Men	5,076	7.5	7.75	0.03	41.3
Women	4,969	8.74	8.04	0.03	41.2

Notes: This table presents summary statistics on user activity by gender. Panel (a) shows results for a user's average reply rate. Panel (b) shows results for the total number contacts that a user has. Panel (c) shows results for the average number of contacts that a user has each month. Panel (d) shows results for the total number of contacts who initiate conversations with a user. Panel (e) shows results for the average number of contact who initiate conversations with a user. Panel (f) shows results for the number of months elapsed before the end of a user's activity.

Table 7: Estimates of Preference Parameters

Variables	Whether User Replies to Message			
	Men	Women	Men	Women
Age				
Contact's age	-0.001	-0.002***	-0.002	-0.002***
	(0.001)	(0.000)	(0.001)	(0.000)
Contact 10+ years older	-0.010	-0.036***	-0.003	-0.043***
	(0.028)	(0.005)	(0.031)	(0.006)
Contact 5-10 years older	-0.037***	-0.014***	-0.028*	-0.016***
	(0.013)	(0.003)	(0.015)	(0.003)
Contact 2-5 years older	-0.032***	0.003*	-0.029***	0.002
	(0.007)	(0.002)	(0.007)	(0.002)
Contact 2-5 years younger	0.016***	-0.018***	0.018***	-0.017***
	(0.005)	(0.002)	(0.005)	(0.002)
Contact 5-10 years younger	0.015**	-0.036***	0.016**	-0.035***
	(0.007)	(0.004)	(0.008)	(0.004)
Contact 10+ years younger	-0.006	-0.063***	0.004	-0.062***
	(0.013)	(0.011)	(0.013)	(0.011)
Height				
Contact's height	0.003***	0.001***	0.004***	0.001***
	(0.001)	(0.000)	(0.001)	(0.000)
Contact 15cm+ taller	0.036	0.039***	0.028	0.040***
	(0.048)	(0.005)	(0.052)	(0.005)
Contact 10-15cm taller	-0.064	0.033***	-0.073	0.035***
	(0.049)	(0.003)	(0.046)	(0.004)
Contact 5-10cm taller	-0.091***	0.019***	-0.116***	0.021***
	(0.020)	(0.002)	(0.024)	(0.003)
Contact 5-10cm shorter	0.050***	0.008	0.051***	0.006
	(0.006)	(0.008)	(0.007)	(0.009)
Contact 10-15cm shorter	0.062***	0.026*	0.063***	0.029
	(0.008)	(0.015)	(0.008)	(0.018)
Contact 15cm+ shorter	0.047***	0.143***	0.050***	0.140***
	(0.010)	(0.017)	(0.011)	(0.019)
BMI				
Contact's BMI	-0.006***	-0.000	-0.006***	0.000
	(0.001)	(0.000)	(0.001)	(0.000)
Contact 6+ higher	0.053***	0.002	0.055***	0.002
	(0.016)	(0.003)	(0.018)	(0.003)
Contact 4-6 higher	0.005	0.006**	0.010	0.004*
	(0.014)	(0.002)	(0.016)	(0.003)
Contact 2-4 higher	-0.022**	0.007***	-0.019**	0.007***
	(0.009)	(0.002)	(0.010)	(0.002)
Contact 2-4 lower	0.010**	-0.005	0.011**	-0.005

	(0.004)	(0.004)	(0.005)	(0.004)
Contact 4-6 lower	0.011*	-0.004	0.013**	-0.001
	(0.006)	(0.006)	(0.006)	(0.007)
Contact 6+ lower	0.016*	-0.014	0.019**	-0.013
	(0.008)	(0.010)	(0.009)	(0.011)
Education				
User: senior vocational school				
Vocational school	-0.023**	0.010***	-0.016	0.008**
	(0.009)	(0.003)	(0.010)	(0.004)
College	-0.049***	0.016***	-0.049***	0.015***
	(0.009)	(0.003)	(0.010)	(0.003)
Master	-0.073***	0.022**	-0.066**	0.014
	(0.026)	(0.009)	(0.028)	(0.009)
PhD	-0.205***	0.005	-0.238***	-0.016
	(0.054)	(0.020)	(0.071)	(0.021)
User: vocational school				
Senior vocational school	-0.011	-0.024***	-0.012*	-0.026***
	(0.007)	(0.003)	(0.008)	(0.003)
College	0.006	0.011***	0.006	0.010***
	(0.006)	(0.002)	(0.006)	(0.003)
Master	-0.041**	0.032***	-0.043**	0.029***
	(0.019)	(0.006)	(0.021)	(0.006)
PhD	-0.183***	0.019	-0.240***	0.025
	(0.024)	(0.015)	(0.027)	(0.016)
User: college				
Senior vocational school	-0.040***	-0.044***	-0.045***	-0.047***
	(0.005)	(0.002)	(0.005)	(0.003)
Vocational school	-0.020***	-0.027***	-0.017***	-0.028***
	(0.004)	(0.002)	(0.004)	(0.002)
Master	0.001	0.027***	0.005	0.024***
	(0.008)	(0.004)	(0.008)	(0.004)
PhD	-0.121***	0.023**	-0.120***	0.027***
	(0.022)	(0.009)	(0.033)	(0.010)
User: master				
Senior vocational school	-0.124***	-0.107***	-0.117***	-0.113***
	(0.019)	(0.011)	(0.020)	(0.012)
Vocational school	-0.091***	-0.100***	-0.086***	-0.103***
	(0.017)	(0.011)	(0.018)	(0.011)
College	-0.048***	-0.049***	-0.045***	-0.048***
	(0.015)	(0.010)	(0.016)	(0.011)
PhD	-0.150***	0.093***	-0.093	0.099***
	(0.052)	(0.026)	(0.064)	(0.026)
User: PhD				
Senior vocational school	0.057	-0.198***	0.042	-0.218***

	(0.123)	(0.075)	(0.141)	(0.082)
Vocational school	0.018	-0.180***	-0.022	-0.187***
	(0.129)	(0.063)	(0.134)	(0.069)
College	0.097	-0.169***	0.058	-0.170***
	(0.116)	(0.064)	(0.126)	(0.063)
Master	0.160	-0.082	0.137	-0.085
	(0.109)	(0.051)	(0.125)	(0.053)
Marital Status				
User single,	-0.019***	-0.039***	-0.020***	-0.041***
contact divorced	(0.006)	(0.002)	(0.007)	(0.003)
User divorced,	-0.071***	-0.033***	-0.069***	-0.036***
contact single	(0.009)	(0.003)	(0.009)	(0.004)
Children				
User no children,	-0.030***	-0.013***	-0.033***	-0.016***
contact has children	(0.008)	(0.003)	(0.009)	(0.003)
User has children,	0.026**	0.000	0.009	0.001
contact no children	(0.012)	(0.003)	(0.012)	(0.004)
Residence Status				
User non-local; contact local	-0.007**	0.013***	-0.006	0.014***
	(0.004)	(0.002)	(0.004)	(0.002)
User local; contact non-local	0.001	-0.013***	-0.007*	-0.016***
	(0.004)	(0.002)	(0.004)	(0.002)
Industry				
Agriculture	-0.121***	0.012	-0.138***	0.016
	(0.022)	(0.008)	(0.027)	(0.010)
Mining	-0.113***	0.021***	-0.116***	0.021**
	(0.020)	(0.007)	(0.024)	(0.009)
Manufacturing	-0.079***	0.004	-0.076***	0.006
	(0.010)	(0.005)	(0.011)	(0.006)
Sales	-0.024***	0.000	-0.027***	0.002
	(0.008)	(0.006)	(0.009)	(0.006)
Transportation	-0.069***	0.013**	-0.065***	0.016**
	(0.012)	(0.006)	(0.014)	(0.007)
Tourism/Catering	-0.034***	0.002	-0.035***	0.006
	(0.012)	(0.006)	(0.013)	(0.007)
IT/Software	-0.039***	0.025***	-0.038***	0.027***
	(0.009)	(0.005)	(0.009)	(0.006)
Finance	-0.002	0.023***	-0.003	0.026***
	(0.009)	(0.006)	(0.010)	(0.007)
Real estate	-0.006	0.009	-0.011	0.011*
	(0.009)	(0.005)	(0.010)	(0.006)
Business services	-0.018**	0.016***	-0.021**	0.017***
	(0.008)	(0.005)	(0.008)	(0.006)

Research/Technical	-0.051***	0.004	-0.048***	0.006
	(0.011)	(0.006)	(0.012)	(0.007)
Community service	-0.043***	0.006	-0.043***	0.007
	(0.009)	(0.006)	(0.010)	(0.007)
Education	-0.010	0.013**	-0.012	0.016**
	(0.008)	(0.006)	(0.009)	(0.007)
Sanitation	-0.017*	0.020***	-0.019*	0.024***
	(0.009)	(0.007)	(0.010)	(0.008)
Entertainment/ Sport	-0.015*	0.017***	-0.016	0.019***
	(0.009)	(0.006)	(0.010)	(0.006)
Social management/Security	-0.022*	0.038***	-0.010	0.042***
	(0.012)	(0.006)	(0.013)	(0.007)
Other	-0.013*	0.006	-0.016*	0.007
	(0.008)	(0.005)	(0.009)	(0.006)
Monthly income	0.001***	0.003***	0.001***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
House	0.023***	0.033***	0.037	-0.015
	(0.004)	(0.001)	(0.059)	(0.021)
House $\cdot$ log$P_{t-1,c}$			-0.001	0.005**
			(0.006)	(0.002)
User-initiated	-0.001	0.061***	-0.008	0.058***
	(0.006)	(0.006)	(0.006)	(0.006)
User's contacts last month	-0.001***	-0.000***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Same city	0.066***	0.039***	0.058***	0.038***
	(0.005)	(0.002)	(0.005)	(0.002)
Same province	0.061***	-0.014***	0.042***	-0.017***
	(0.005)	(0.002)	(0.006)	(0.003)
Constant	0.025	-0.174***	-0.092	-0.090*
	(0.103)	(0.054)	(0.109)	(0.052)
Observations	192,723	692,364	170,909	588,529
R-squared	0.025	0.043	0.019	0.041
Number of users	5,015	4,967	4,981	4,967
Year F.E.	Yes	Yes	Yes	Yes
City F.E.	Yes	Yes	Yes	Yes
Dep Mean	0.286	0.183	0.286	0.183
Dep SD	0.452	0.387	0.452	0.387

Notes: *** $p < .01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors are clustered at user level. This table presents OLS regression results for users' preference estimates. An observation is a user-contact pair. Dependent variable is a dummy that is equal to 1 if the user replies, and 0 otherwise. Odd-numbered columns display results for male users, and even-numbered columns display results for female users.

Table 8: Estimates of Preference Parameters: Gender Difference

	No Price Effects		Price Effects	
	Men	Women: Dfiff	Men	Women: Diff
Age				
Contact's age	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Contact 10+ years older	-0.027	-0.010	-0.012	-0.032
	(0.028)	(0.029)	(0.031)	(0.031)
Contact 5-10 years older	-0.045***	0.031**	-0.031**	0.014
	(0.013)	(0.014)	(0.015)	(0.015)
Contact 2-5 years older	-0.035***	0.038***	-0.030***	0.032***
	(0.007)	(0.007)	(0.007)	(0.007)
Contact 2-5 years younger	0.018***	-0.035***	0.018***	-0.035***
	(0.005)	(0.005)	(0.005)	(0.005)
Contact 5-10 years younger	0.017**	-0.053***	0.016**	-0.052***
	(0.007)	(0.008)	(0.008)	(0.009)
Contact 10+ years younger	-0.006	-0.058***	0.004	-0.067***
	(0.013)	(0.016)	(0.013)	(0.018)
Height				
Contact's height	0.003***	-0.001**	0.004***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Contact 15cm+ taller	0.042	-0.003	0.033	0.007
	(0.047)	(0.047)	(0.051)	(0.051)
Contact 10-15cm taller	-0.071	0.104**	-0.070	0.105**
	(0.050)	(0.050)	(0.046)	(0.046)
Contact 5-10cm taller	-0.100***	0.120***	-0.116***	0.137***
	(0.021)	(0.021)	(0.025)	(0.025)
Contact 5-10cm shorter	0.051***	-0.043***	0.050***	-0.044***
	(0.006)	(0.010)	(0.007)	(0.011)
Contact 10-15cm shorter	0.063***	-0.036**	0.063***	-0.033*
	(0.008)	(0.017)	(0.008)	(0.020)
Contact 15cm+ shorter	0.048***	0.104***	0.050***	0.092***
	(0.010)	(0.020)	(0.011)	(0.022)
BMI				
Contact's BMI	-0.005***	0.005***	-0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
Contact 6+ higher	0.050***	-0.047***	0.053***	-0.052***
	(0.016)	(0.017)	(0.018)	(0.018)
Contact 4-6 higher	0.006	-0.000	0.009	-0.005
	(0.014)	(0.014)	(0.016)	(0.017)
Contact 2-4 higher	-0.021**	0.028***	-0.020**	0.027***
	(0.009)	(0.009)	(0.010)	(0.010)

Contact 2-4 lower	0.010**	-0.014**	0.011**	-0.016**
	(0.004)	(0.006)	(0.005)	(0.006)
Contact 4-6 lower	0.010*	-0.015*	0.012*	-0.012
	(0.006)	(0.008)	(0.006)	(0.009)
Contact 6+ lower	0.015*	-0.029**	0.018**	-0.031**
	(0.008)	(0.013)	(0.009)	(0.014)
Education				
User: senior vocational school				
Vocational school	-0.023***	0.033***	-0.017*	0.025**
	(0.009)	(0.010)	(0.010)	(0.011)
College	-0.049***	0.065***	-0.049***	0.065***
	(0.009)	(0.010)	(0.010)	(0.010)
Master	-0.070***	0.092***	-0.064**	0.078***
	(0.026)	(0.027)	(0.028)	(0.029)
PhD	-0.218***	0.223***	-0.238***	0.223***
	(0.053)	(0.056)	(0.071)	(0.074)
User: vocational school				
Senior vocational school	-0.009	-0.015*	-0.012	-0.015*
	(0.007)	(0.007)	(0.008)	(0.008)
College	0.008	0.004	0.006	0.004
	(0.006)	(0.006)	(0.006)	(0.007)
Master	-0.038**	0.071***	-0.041*	0.070***
	(0.019)	(0.020)	(0.021)	(0.022)
PhD	-0.196***	0.217***	-0.239***	0.265***
	(0.023)	(0.027)	(0.026)	(0.031)
User: college				
Senior vocational school	-0.040***	-0.005	-0.045***	-0.003
	(0.005)	(0.005)	(0.005)	(0.006)
Vocational school	-0.020***	-0.008*	-0.017***	-0.012**
	(0.004)	(0.004)	(0.004)	(0.005)
Master	0.000	0.028***	0.004	0.021**
	(0.008)	(0.009)	(0.008)	(0.010)
PhD	-0.138***	0.161***	-0.126***	0.153***
	(0.022)	(0.024)	(0.033)	(0.034)
User: master				
Senior vocational school	-0.122***	0.013	-0.114***	-0.001
	(0.019)	(0.022)	(0.020)	(0.023)
Vocational school	-0.089***	-0.013	-0.083***	-0.022
	(0.017)	(0.020)	(0.017)	(0.021)
College	-0.047***	-0.004	-0.044***	-0.006
	(0.015)	(0.018)	(0.015)	(0.019)
PhD	-0.150***	0.244***	-0.096	0.196***
	(0.052)	(0.058)	(0.064)	(0.070)
User: PhD				

Senior vocational school	0.020	-0.221	0.042	-0.262
	(0.146)	(0.164)	(0.138)	(0.160)
Vocational school	-0.014	-0.169	-0.019	-0.170
	(0.148)	(0.161)	(0.130)	(0.147)
College	0.065	-0.235	0.060	-0.231*
	(0.133)	(0.147)	(0.122)	(0.137)
Master	0.126	-0.207	0.136	-0.220
	(0.129)	(0.138)	(0.124)	(0.134)
Marital Status				
User single,	-0.016**	-0.022***	-0.018***	-0.023***
contact divorced	(0.006)	(0.007)	(0.007)	(0.007)
User divorced,	-0.071***	0.038***	-0.067***	0.030***
contact single	(0.009)	(0.009)	(0.009)	(0.010)
Children				
User no children,	-0.031***	0.017**	-0.034***	0.018*
contact has children	(0.008)	(0.009)	(0.009)	(0.009)
User has children,	0.032***	-0.032***	0.011	-0.011
contact no children	(0.012)	(0.012)	(0.012)	(0.013)
Residence Status				
User non-local; contact local	-0.008**	0.020***	-0.006	0.020***
	(0.004)	(0.004)	(0.004)	(0.004)
User local; contact non-local	0.000	-0.011***	-0.008*	-0.008*
	(0.004)	(0.004)	(0.004)	(0.004)
Industry				
Agriculture	-0.127***	0.140***	-0.135***	0.151***
	(0.022)	(0.023)	(0.027)	(0.029)
Mining	-0.120***	0.141***	-0.115***	0.137***
	(0.020)	(0.021)	(0.024)	(0.026)
Manufacturing	-0.079***	0.083***	-0.074***	0.080***
	(0.010)	(0.012)	(0.011)	(0.013)
Sales	-0.023***	0.022**	-0.026***	0.028**
	(0.008)	(0.010)	(0.009)	(0.011)
Transportation	-0.073***	0.085***	-0.066***	0.081***
	(0.013)	(0.014)	(0.014)	(0.016)
Tourism/Catering	-0.033***	0.035***	-0.033**	0.039***
	(0.012)	(0.013)	(0.013)	(0.015)
Information/Software	-0.039***	0.064***	-0.036***	0.064***
	(0.009)	(0.010)	(0.009)	(0.011)
Finance	-0.002	0.024**	-0.003	0.029**
	(0.010)	(0.011)	(0.010)	(0.012)
Real estate	-0.005	0.013	-0.010	0.022*
	(0.009)	(0.011)	(0.010)	(0.012)
Business services	-0.017**	0.033***	-0.020**	0.037***

	(0.008)	(0.009)	(0.008)	(0.010)
Research/Technical	-0.053***	0.057***	-0.048***	0.054***
	(0.011)	(0.013)	(0.012)	(0.014)
Community service	-0.042***	0.047***	-0.042***	0.049***
	(0.009)	(0.011)	(0.010)	(0.012)
Education	-0.008	0.021**	-0.011	0.027**
	(0.008)	(0.010)	(0.009)	(0.011)
Sanitation	-0.015*	0.035***	-0.018*	0.043***
	(0.009)	(0.012)	(0.010)	(0.013)
Entertainment/ Sport	-0.015*	0.032***	-0.015	0.034***
	(0.009)	(0.011)	(0.010)	(0.012)
Social management/ Security	-0.024**	0.062***	-0.009	0.052***
	(0.012)	(0.013)	(0.013)	(0.015)
Other	-0.012	0.018*	-0.015*	0.022**
	(0.008)	(0.010)	(0.009)	(0.011)
Monthly income	0.001***	0.002***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
House	0.023***	0.010**	0.036	-0.053
	(0.004)	(0.004)	(0.058)	(0.062)
House $\cdot$ $\log P_{t-1,c}$			-0.002	0.007
			(0.006)	(0.007)
User-initiated	0.038***		0.032***	
	(0.004)		(0.004)	
User's contacts last month	-0.000***		-0.000***	
	(0.000)		(0.000)	
Same city	0.046***		0.043***	
	(0.002)		(0.002)	
Same province	0.046***		-0.004	
	(0.002)		(0.003)	
Constant	0.222		0.103	
	(148.117)		(0.272)	
Observations	885,087		759,438	
R-squared	0.034		0.032	
Number of users	9,982		9,948	
Year F.E.	Yes		Yes	
City F.E.	Yes		Yes	
Dep Mean	0.212		0.212	
Dep SD	0.409		0.409	

Notes: *** $p < .01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors are clustered at user level. This table presents OLS regression results for gender differences in users' preference estimates. An observation is a user-contact pair. Dependent variable is a dummy that is equal to 1 if the user replies, and 0 otherwise. Odd-numbered columns display results for male users, and even-numbered columns display results for female-male differences.

Table 9: Effect of Housing Prices on Competitiveness of Homeowners

	(1)	(2)	(3)	(4)
	Men	Women	Men	Women
Panel (a): Average number of contact-initiated messages per day				
house_u	0.171	1.342*	7.552***	-4.038
	(0.120)	(0.710)	(1.938)	(9.820)
$\text{house}_u \cdot \log P_{t-1,u}$			0.835***	0.578
			(0.216)	(1.070)
Obs	2,947	2,867	2,946	2,867
R-squared	0.167	0.077	0.173	0.077
Dep Mean	1.599	13.06	1.599	13.06
Dep SD	3.095	15.21	3.095	15.21
Panel (b): Monthly income of contacts				
house_u	0.039	0.09	-3.861***	-1.01
	(0.086)	(0.072)	(1.261)	(0.828)
$\text{house}_u \cdot \log P_{t-1,u}$			0.414***	0.118
			(0.135)	(0.089)
Obs	107,028	351,284	106,502	351,142
R-squared	0.240	0.273	0.240	0.273
Dep Mean	23.467	29.2577	23.467	29.257
Dep SD	8.223	10.340	8.223	10.340
Panel (c): Home-ownership of contacts				
house_u	0.004	0.008***	0.083	0.016
	(0.004)	(0.003)	(0.055)	(0.030)
$\text{house}_u \cdot \log P_{t-1,u}$			-0.008	-0.001
			(0.006)	(0.003)
Obs	107,028	351,284	106,502	351,142
R-squared	0.162	0.212	0.162	0.212
Dep Mean	0.129	0.522	0.129	0.522
Dep SD	0.335	0.500	0.335	0.500

Notes: *** $p < .01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors are clustered at user level. This table presents OLS regression results for the effects of housing prices on the competitiveness of homeowners in the online dating market. In panel (a), an observation is a user; in panels (b) and (c), an observation is a user-contact pair. In panel (a), dependent variable is the average number of daily messages initiated by contacts. In panel (b), dependent variable is the income bracket of the contact. In panel (c), dependent variable is the home-ownership status of the contact. House_u is equal to 1 if user u is a homeowner, and 0 otherwise. $P_{t-1,c}$ is 1-month lagged average housing price of contact c's city of residence. Columns (1) and (3) display results for male users, and columns (2) and (4) display results for female users.

Table 10: Assortative Matching Patterns for Matched Couples under the 95- and 97.5-Percentile Definitions

	Matches: 95% percentile			Matches: 97.5% percentile		
Panel (a): Sorting Along Education						
	educ_m (1)	educ_m (2)	educ_f (3)	educ_m (4)	educ_m (5)	educ_f (6)
educ_f	0.143***	0.116		0.154***	-0.038	
	(0.012)	(0.170)		(0.017)	(0.239)	
$\text{educ}_f \cdot \log P_{t-1,f}$		0.003			0.021	
		(0.018)			(0.025)	
educ_m			-0.068			-0.359
			(0.187)			(0.272)
$\text{educ}_m \cdot \log P_{t-1,m}$			0.024			0.057**
			(0.020)			(0.029)
Obs	8,865	8,865	8,865	4,365	4,365	4,365
R-squared	0.200	0.200	0.160	0.200	0.200	0.164
Panel (b): Sorting Along Income						
	income_m (1)	income_m (2)	income_f (3)	income_m (4)	income_m (5)	income_f (6)
income_f	0.081***	0.400*		0.072***	0.592**	
	(0.017)	(0.211)		(0.024)	(0.291)	
$\text{income}_f \cdot \log P_{t-1,f}$		-0.034			-0.055*	
		(0.022)			(0.030)	
income_m			0.152			0.399**
			(0.127)			(0.185)
$\text{income}_m \cdot \log P_{t-1,m}$			-0.011			-0.038*
			(0.013)			(0.020)
Obs	8,865	8,865	8,865	4,365	4,365	4,365
R-squared	0.286	0.286	0.226	0.288	0.289	0.229
Panel (c): Sorting Along Home-ownership						
	house_m (1)	house_m (2)	house_f (3)	house_m (4)	house_m (5)	house_f (6)
house_f	0.006	0.187		-0.009	0.039	
	(0.014)	(0.168)		(0.019)	(0.240)	
$\text{house}_f \cdot \log P_{t-1,f}$		-0.019			-0.005	
		(0.018)			(0.025)	
house_m			0.097			0.126
			(0.095)			(0.139)
$\text{house}_m \cdot \log P_{t-1,m}$			-0.010			-0.014
			(0.010)			(0.015)
Obs	8,865	8,865	8,865	4,365	4,365	4,365
R-squared	0.215	0.215	0.175	0.219	0.219	0.170

Notes: *** $p < .01$, ** $p < 0.05$, * $p < 0.1$. Robust standard errors are clustered at user level. An observation is a user-match pair. This table presents results on assortative matching patterns. Columns (1), (2) and (3) adopt the 95-percentile definition of a match, and column (4), (5) and (6) adopt the 97.5%-percentile definition of a match. In panel (a), the dependent variable is the education level bracket of the match. In panel (b), the dependent variable is the income bracket of the match. In panel (c), the dependent variable is the home-ownership status of the match.

Appendix A Data Sampling Procedure

Faced with limitations on sample size and access to large-scale data, we make use of the following sampling procedure.

First, we select a total of 10,069 users who became registered on *Jiayuan.com* after April 2012. We randomly sample 3,269 of them from large metropolitan areas, 3,400 from medium metropolitan areas and 3,400 from small metropolitan areas.[7] The purpose for sampling users independently from large, medium and small metropolitan areas is to exploit variations in housing prices between more developed and less developed cities. Within each location group, we try to achieve an equal proportion of male and female users and homeowners and non-homeowners. Table 2 provides a summary of the geographical distribution of users by gender and home-ownership.

We impose two restrictions on users' online profiles while performing the sampling procedure: they must be aged between 22 and 40, and have indicated their home-ownership status in their profiles. We use this age range because the legal minimum age for marriage is 20 for women and 22 for men in China. We also want to use this restriction to reduce the likelihood of users seeking re-marriage because their preferences may differ markedly from the preferences of those seeking for first marriage. For example, factors such as physical attractiveness and socio-economic status are likely to be less important than, say, a potential spouse's personality and physical health conditions. Lastly, since some users choose not to answer questions related to home-ownership, the condition on home-ownership status would eliminate potential wasteful draws.

In addition to the above restrictions, we also impose conditions on users activity. Prior to obtaining the data used in this paper, we obtained from Jiayuan a much smaller set of user data with full records of their messaging activities. We analyzed this data for users' behavioral patterns, and propose two restrictions based on our findings: (1) the number of monthly messages that users have sent out since registration must be greater than the 3th percentile of the number of monthly messages sent out by all users, and (2) users must be a non-spammer. Since we estimate utility based on the sending and replying of messages, the first restriction would eliminate users who have few messages and whose utilities are difficult to estimate. We focus on non-spammers because they are more likely to be looking for serious and long-term partners rather than casual relationships. For instance, in our sample data, 20 out of 130 men contacted between 1,000 and 7,000 women since registration. It is thus unlikely that these users had taken the time to properly view women's profiles and to consider them as potential marriage partners.

The definition of a spammer is based on two key indicators: x_{it}, the number of conversations that user i initiates with new contacts in month t, and $\bar{x}_i$, the average number of conversations that user i initiates with new contacts per month during his period of activity—from the first time he logs onto Jiayuan until the last time. The statistics of any user i are then compared against the distribution of x_{jt} and y_j of all users j in the small sample. We define user i to be a spammer if either x_{it} or $\bar{x}_i$ exceeds the 97^{th} percentile in their respective distributions.

[7]The reason for selecting a smaller number of users in large metropolitan areas compared to small and medium metropolitan areas is that there are not enough women who own an apartment in large cities.

The following reasoning explains our choice of indicators in the definition of a spammer. Based on the small set of user activity data that we obtained, we see that people who send an incredibly large number of messages in the sample generally exhibit one of two behavioral patterns. First, they tend to be less active for the most of the time, when they send less than 10 messages per day, but could send up to 400 messages per day during some hyperactive days. The restriction on x_{it} is used to eliminate those users. Second, they tend to send a sizable number of messages consistently—for example, 50 messages per day for 30 consecutive days. The restriction on y_i is used to eliminate those users.

After sampling out 10,069 users who satisfy the conditions on both online profiles and user activity, we collect information on the contacts of these sample users. In the preceding empirical strategy section, we have made clear that utility estimation relies on a user's choice to respond to incoming messages. These messages are either sent to initiate conversations with that user, or as a response to that user's conversation-initiating message. Hence, the pool of contacts for which we collection information comprises of exactly these two types of people: conversation initiators and first-time responders. Due to statistical and administrative constraints, we impose a restriction on the maximum number of contacts for each user that we look at. For each user, starting from his registration, we select up to 275 contacts that are either conversation initiators or first-time responders, and collect the profile information of these contacts. We end up with 635,670 observations for male participants and 362,309 observations for female participants.

We argue that the contacts we obtain under this procedure are as good as random. If we assume that a user's behavior and preferences are sufficiently stable throughout her period of activity, then randomly sampling contacts would be the same as randomly choosing an interval of time from the user's period of activity, and it should not matter if this interval of time was taken from the beginning, in the middle or at the end of his active period. More importantly, we select a user's entire pool of contacts during a particular time interval, instead of randomly selecting a fixed number of contacts from a user's entire chat history. The reason is that a user's choice may depend not only on his preferences but also his opportunities, such as the number of contacts with whom he regularly exchanges messages, and the average characteristics of his options. Therefore, having a user's entire pool of options during a particular interval will shed light on the relative importance of individual characteristics in choosing a dating partner.