Anticipating China's unmarried males: stochastic microprojections of marriage market outcomes under demographic uncertainty*

Ethan J. Sharygin†

University of Pennsylvania

Population Studies Center

August 2011

Abstract Low fertility and high sex ratios in China have combined to create a large population of unmarried adult males. In order to compliment studes of consequences, this paper attempts to estimate the scale of non-marriage in light of China's demographic dynamics. An increasing share of the Chinese population is attaining secondary and post-secondary education. I show that marriage and fertility rates in China are responding in ways that will affect the patterns of male and female non-marriage. Projections from a novel agent-based model suggest that marriage delay for males and females will result in greater period shares non-married and will have possible tempo effects on the total fertility rate that should be anticipated.

Keywords: China; marriage market; stable marriage; hypergamy; sex ratios; missing girls; 'bare branches'; marriage squeeze.

JEL Classification: I18, J11, J12, J13, N35.

---

*The author is grateful to Warren Sanderson, Michel Guillot, Emily Hannum, Avraham Ebenstein, and Monica Das Gupta for advice and suggestions, and to and to participants in the summer population workshop at the Institute for Applied Systems Analysis for helpful comments.

†Address: 3718 Locust Walk, McNeil Bldg 239, Philadelphia, PA 19143. E-mail: *garba@pop.upenn.edu.*

Introduction

This paper seeks to address the consequences of a conflux of demographic dynamics in China, unique among countries for the longstanding low fertility and high sex ratios that have prevailed in the population since the late 1970s. Already, a number of recent papers have begun studying the anticipated social and economic fallout resulting from high adult sex ratios and male non-marriage. Previous studies of marriage market imbalances have focused on cases of male shortage, or have imposed unnecessarily static assumptions about the future population. As a result, a great deal of uncertainty remains as to the level of male non-marriage in China that may result from the serious adult sex ratio imbalance, and the responsiveness of the share never marrying to possible sources of demographic change. The objective of the present study is to provide a method of estimating shares never marrying upon which future discussions of consequences can be more certainly founded.

I begin with empirical observations about marriages in China from the 1982, 1990, and 2000 Censuses and proceed to parameterize the age and preference functions of the marriage market. I then conduct a population projection model applying demographic rates in an iterated, stochastic process. For each round of the projection model, I apply parameterizations of the age-specific nuptiality rates to determine entrants to the marriage market. I use a variant of the Gale-Shapley stable mariage algorithm to produce stable marriages for all entrants to the marriage market (Gale, 1962; **?**). I define a system of ranks for all pairwise combinations using parameterizations of empirical observations regarding marriage patterns from census microdata. The process is repeated each year, under differing assumptions about behavioral change regarding marriage norms and under uncertainty regarding the true sex ratio of births and level of fertility in the population. The resulting analytic dataset contains individuals by age, sex, education, province of residence, and ever-married status for years 2000-2050 under a variety of scenarios of future change and which take into account uncertainty in future vital rates.

Section 1 describes the nature of the demographic challenge facing China, including reference to the ongoing discussion of its causes and consequences and previous attempts to estimate the scale of male non-marriage. In Section 2, I investigate cohort and time trends in age-specific nuptiality rates. I find that hypergamy is the defining characteristic of the marriage market in China, and chart changes in a new index of hypergamy over time as well as other individual and cohort characteristics that predict marriage or nonmarriage and characteristics of spouse pairs. In Section 3, I discuss the application of the Gale-Shapley stable marriage algorithm to the population in China, including a discussion of the system to define preferences parametrically from the observed characteristics of marriage matches in previous Censuses. Section 4 describes the population projection model and the sources of present and future uncertainty in vital rates. In addition, hypothetical scenarios are introduced that capture the effects of both moderate and extreme possible changes in future marriage behavior. Section 5 presents findings for future population growth and rates of non-marriage in the Chinese population from the microsimulation model building upon the previous sections. Section 6 concludes with a discussion of the larger consequences of the findings.

Preliminary Results

Age-specific marriage rates in the China 2000 Census show that marriage years are remarkably compressed during ages 20-30. Few marriages occur below age 20 or above age 30 for females and males, with females exhibiting slightly longer left-hand tails and longer right-hand tails for males, representing a greater share of males marrying at older ages than females. Although Figure 1 is estimated across cohorts observed in the 2000 Census, nonetheless it is highly suggestive of continuation in the trend of very high shares ultimately marrying, with the exception being men of low education (who do continue to marry at ages above 50, albeit at a low rate). This group is already experiencing a decline in marriage rates compared to men of similar educational background as reflected in prior Censuses of 1982 and 1990.

[Figure 1 about here].

figure showing marriage attainment by cohort(age), sex, educ

The approach adopted in this marriage projection is to define parameters that determine marriage propensities and ranks of potential spouses. The first task is achieved by fitting the Coale-McNeil model schedule of nuptiality to the marriages observed during the 10 years between 1991-2000, to the population by age, sex, and education level. I compare the parametrized with the observed rates of first marriage. These propensities are calculated for every individual, and in the marriage model they are compared against a draw from a random uniform distribution to determine whether a never-married individual should be included in the marriage market.

[Figure 2 about here].

figure showing modeled and observed age-specific first marriage rates

At each given age and education level, I define a system of preferences in similar fashion by fitting the Coale-McNeil gamma function to the observed frequency distribution of marriages for each combination of age and education. Modeled rates aggregated by education for females age 19 and 29 are shown alongside the same figures for males (Figure 3). Since the distribution describes outcomes for those who married, the rates sum to unity.

[Figure 3 about here].

figure showing modeled and observed age-specific first marriage rates for two example female ages, categorized by the age of their spouses

According to these results, young males tend to marry women slightly younger than themselves, while older men marry women across a wide spread of ages. Young women tend to marry men significantly older, while women approaching 30 years old marry from a broader age pool of males. These marriage match propensities define a pool of preferences on the part of each individual on the marriage market, and a random draw from a uniform distribution determines where on this distribution a woman's most-preferred spouse characteristics lie. These distributions can define an exhaustive set of preferences that can be combined in a two-sex model to match individuals on the marriage market each year, thereby avoiding contradictory or impossible results for male and female marriage rates.

In addition to age preferences, I consider preferences for spouse's education. Marriages are hypergamous if they result in a woman being paired with a male whose completed education exceeds her own. These marriages in China are much more prevalent than marriages of women to men with lower status. I describe hypergamy in terms of average gaps in education attainment between spouses. The average gap between spousal education, weighted by the frequences of marriages between 1999-2000, is +0.5 years, and has been trending downward since the early 1980s. Increasing female education attainment means that constant preferences would result in downward-trending hypergamy measures, since there are fewer hypergamous options the higher one's own education becomes. Changes in the education gaps between spouses, therefore, is a function of switching between the lines on the graph rather than changes in the lines.

[Figure 4 about here].

figure showing average years of age and education gap over time

Marriage markets are complex, but can be reduced to a system for generating matches between two exclusive sets. I execute stochastic stable matching using preference scores, which may be partial and correlated, allowing for incomplete lists and ties (Bouffard et al., 2001; Irving, 1994; Irving et al., 1987; Manlove, 2002; Pini et al., 2011). These modifications to the classic stable marriage model proposed by Gale and Shapley (Gale, 1962; McVitie and

Wilson, 1971) ensure equal treatment of sexes, infrequent occurence of extreme matches, and insensitivity to the sex dominance of the algorithm.

I consider a population with two sexes, men $M = \{m_1, m_2, ..., m_n\}$ and women $W = \{w_1, w_2, ..., w_n\}$. Each member of M, W is characterized by their own age, education, and heterogeneous error term. Each member of M, W also has a preference function $p \in [0, 1) = p(a, e, \eta, \varepsilon)$ which indicates a preference for a spouse of age a and education level e, with error η and a union-specific random disturbance $\varepsilon \in N(\mu = 0, \sigma = .2)$ representing idiosyncracies in tastes. Let A_{mt} equal the set of age-specific nuptiality rates for a male m at time t and E_{mt} the education-specific nuptiality rates.

The utility to any member of either set W, M is defined by the preference function $p(\cdot)$ and the properties of the proposed match candidate w, m and two random error terms. $\pi \in \Pi$ represents a single match of one man and woman m, w out of the total set of possible matches of $M \times W$. The total utility of the match π to the woman and man is given by $u(\pi, w)$ and $u(\pi, m)$, respectively. For each match π and every woman $w \in W$, there is a set $\{m|\phi_w(\pi, m) = 1\}$ that contains the list of men that would be preferred by the woman to the man of match π, where:

There also exists some set F_π of females who are preferred by males to their match under π and who would also prefer to marry those males, where

$$F_\pi = \{w|\exists m \in M \text{ with } \phi_w(\pi, m) = 1, \phi_m(\pi, w) = 1\}.$$

A stable equilibrium $\Pi_i \subset \Pi$ is the one which generates $F_\pi = \emptyset$; *i.e.* one in which there are no two people who would rather be joined with each other than with their current match.

I include the necessary components to the model to project forward the population, incorporating additional uncertainty about the present and future vital rates in the population. Period share of males unmarried is examined under several scenarios. Under all scenarios, fertility, mortality and the sex ratio are determined by a probablistic function.

Starting values of individual education levels e are observed in the 2000 Census. For years beyond 2000, values at time $t + n$ are calculated via a matrix of age- and education-specific transition probabilites. The projected educational transition probabilities implemented here are those produced by the latest revision of the Human Capital Database of IIASA.

Fertility and the sex ratio are subject to uncertainty. I adopt the methodology and variable values from Lutz et al. (2005), itself a meta-analysis of estimates of China's SRB and TFR. E.g., a total fertility rate of 1.59 is adopted starting in 2000, following Retherford et al. (2004). This value is set as the mean of a normal distribution whose 95% confidence interval is [1.39,1.79], intended to capture the uncertainty regarding the true level of fertility.

Under the baseline scenario, preferences for spouses and age-specific nuptiality rates described above are held constant in the future. Because the population distribution by age and education continues to change under the baseline scenario, it exhibits several interesting properties. First, we should expect a tendency in the overall population towards homogamy. As illustrated by Figure 4, the gap between male and female education is diminished for higher levels of female education. Indeed, females with tertiary education, when they marry, are likely to marry males with lower education. The most homogamous group, and the group expected to grow by the greatest amount in the next decades, are females with completed secondary education, who on average marry men with the same years of education. Second, fewer women may choose to marry. Marriage hazards are significantly lower for women at the highest levels of education. First marriage hazard for females decreases as a function of education.

The matching model used in the projections enables us to consider the likely trajectory of mean ages at first marriage under baseline preference functions. Due to increasing education, female age at marriage may increase during the decade 2010-2020, after a period of relative stability around the ages for males and females observed in the 2000 Census (Figure 5). As female age at marriage rises, female aggregate mate preferences change, and women tend to choose older, better-educated men.

The results for the baseline scenario highlight the significance of future educational attainment of women, and the possible unanticipated consequences for male non-marriage. Further, the modeling strategy adopted in this study enables a concrete projection with age and education profiles of the unmarried men and women of China.

References

Bouffard, N., Easther, R., Johnson, T., Morrison, R. and Vink, J. (2001), "Matchmaker, matchmaker, make me a match", *Brazilian Electronic Journal of Economics* , Vol. 4.

Gale, D. (1962), "College admissions and the stability of marriage", *The American Mathematical Monthly* , Vol. 69, pp. 9–15.

Irving, R. W. (1994), "Stable marriage and indifference", *Discrete Applied Mathematics* , Vol. 48, p. 261272.

Irving, R. W., Leather, P. and Gusfield, D. (1987), "An efficient algorithm for the 'optimal' stable marriage", *Journal of the Association for Computing Machinery* , Vol. 34, p. 532543.

Lutz, W., Scherbov, S., Cao, G., Ren, Q. and Zheng, X. (2005), China's uncertain demographic present and future. IIASA Interim Report IR-05-043.

Manlove, D. (2002), "The structure of stable marriage with indifference", *Discrete Applied Mathematics* , Vol. 122, pp. 167–181.

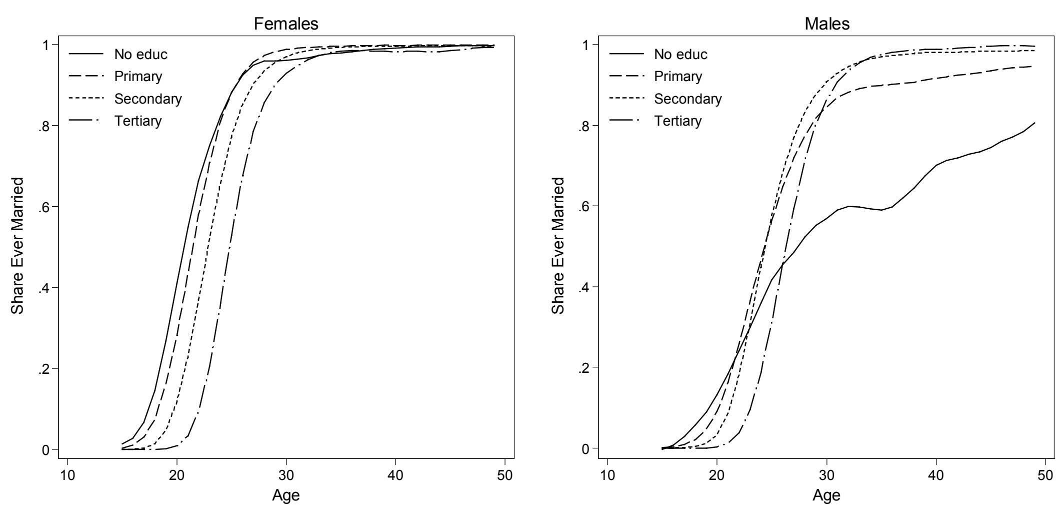
McVitie, D. and Wilson, L. (1971), "The stable marriage problem", *Communications of the ACM* , Vol. 14.

Pini, M. S., Rossi, F. and Venable, K. B. (2011), Stability in matching problems with weighted preferences. Paper Presented at ICAART 2011 - 3rd International Conference on Agents and Artificial Intelligence.

Retherford, R. D., Choe, M. K., Chen, J., Li, X. and Cui, H. (2004), "Fertility in China: How much has it really declined?", *Chinese Journal of Population Research* , Vol. 28, pp. 3–15.

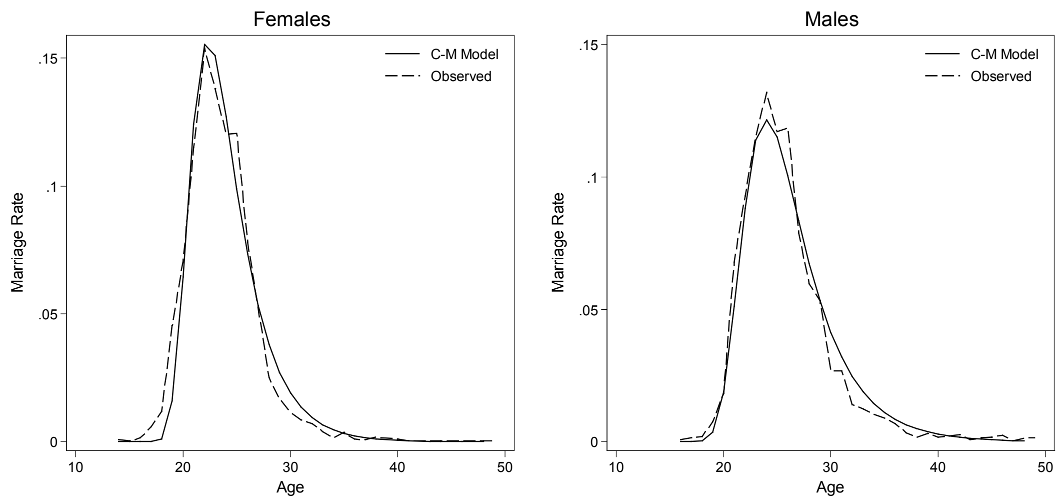
Tables and Figures

Figure 1: Cumulative Share Ever Married



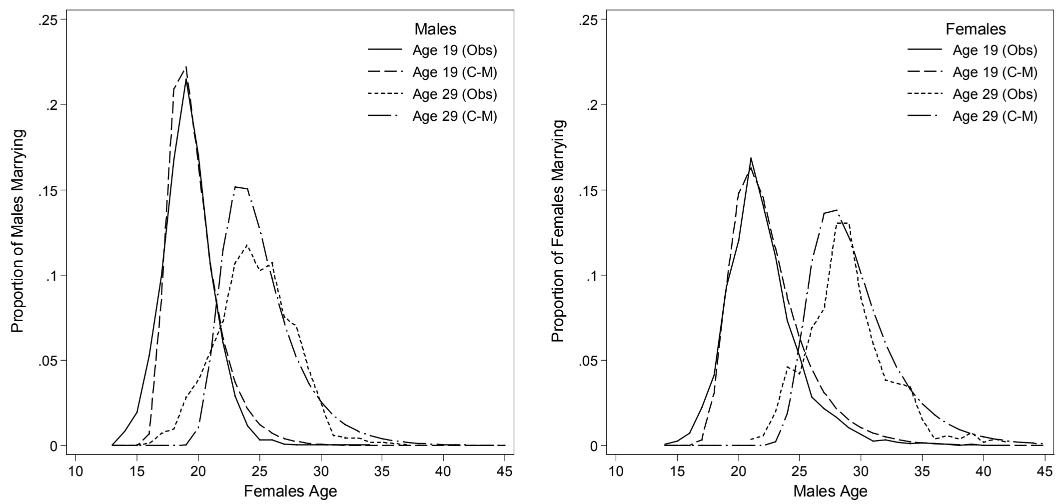
Note: Period (cross-section) measure. Smoothed via locally weighted regression.
Source: China 2000 Census 0.1% Sample.

Figure 2: Observed and modeled first marriage propensities



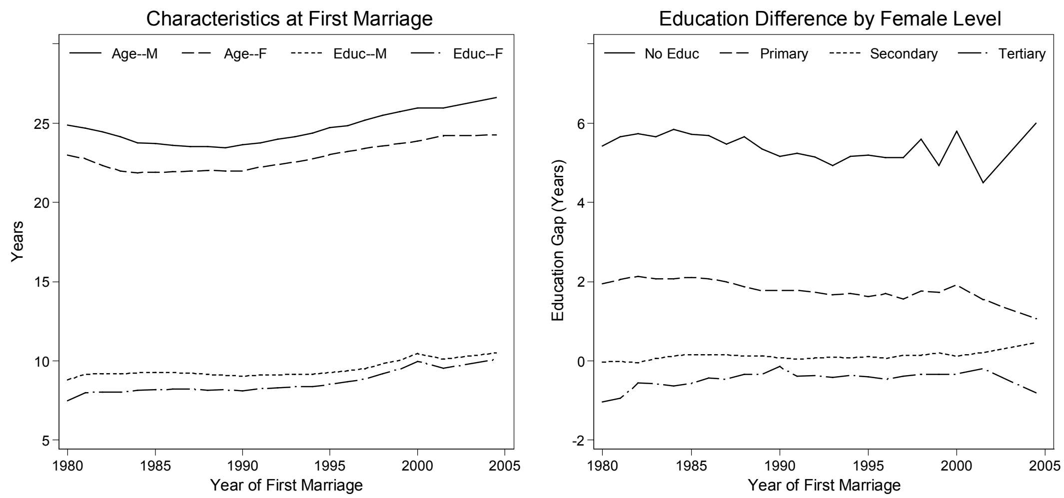
Note: C-M refers to Coale-McNeil nuptiality model, a constrained gamma distribution.
Source: China 2000 Census 0.1% Sample; author's calculation. **drop in favor of the aasnr figure?**.

Figure 3: Observed and modeled first marriage preferences



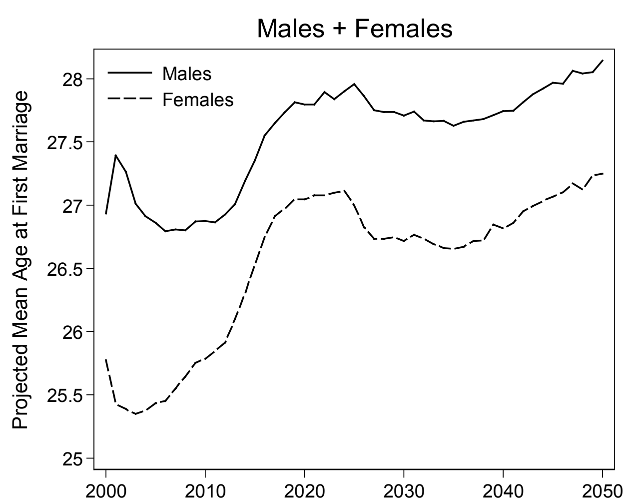
Note: C-M refers to Coale-McNeil nuptiality model, a constrained gamma distribution. Frequency describes share of one sex who marry ($\sum = 1$), categorized by the age of their chosen spouse.
Source: China 2000 Census 0.1% Sample; author's calculation.

Figure 4: Average Spousal Age and Education Gaps, 1980-2005



Note: Marriages combined in 3-year groups for 2000-2005 data due to low counts. Only first marriages intact at the time of the Census are included.
Source: China 2000 Census, 2005 Inter-census Survey; author's calculation.

Figure 5: Projections of Mean Age at First Marriage, 2000-2050



Note: Confidence intervals not shown.
Source: China 2000 Census; author's calculations.