I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11466

A New Estimator of Search Duration and Its Application to the Marriage Market

Pavel Jelnov

APRIL 2018

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11466

A New Estimator of Search Duration and Its Application to the Marriage Market

Pavel Jelnov
Leibniz University Hannover and IZA

APRIL 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

A New Estimator of Search Duration and Its Application to the Marriage Market

It is well known that female age at first marriage positively correlates with male income inequality. The common interpretation of this fact is that marital search takes longer when the pool of potential mates is more unequal. This paper challenges that interpretation with a novel econometric method. I utilize the fact that the female age at first marriage was shown to be a sum of a skewed term, possibly related to search, and a normally distributed residual. I estimate search duration as the expected skewed term. I find that in the American data this term does not positively correlate with male income inequality and female education.

JEL Classification: J12, D83

Keywords: marital search, marriage age, inequality, female education

Corresponding author:
Pavel Jelnov
Wirtschaftswissenschaftliche Fakultät
Institut für Arbeitsökonomik
Leibniz University Hannover
Königsworther Platz 1
30167 Hannover
Germany

E-mail: jelnov@aoek.uni-hannover.de

1 Introduction

A large body of studies documents a positive correlation between male income inequality and female age at first marriage (Keeley (1977, 1979), Bloom and Bennett (1990), Danziger and Neuman (1999), Loughran (2002), Gould and Passerman (2002), Coughlin and Drewianka (2011), Yu and Yu (2013), Li (2014)). Generally, the literature interprets this correlation as a positive correlation between male income inequality and female marital search duration. This interpretation is made by assumption, intuitively inspired by the search theory. To the best of my knowledge, no empirical study tested this assumption until now. The reason is simple: search duration is an unobservable variable in almost every dataset. To close this gap, I propose a new econometric method that utilizes the contribution of Ansley Coale, one of the most prominent demographers of the twentieth century, who discovered the common age pattern of marriage across times and countries. Specifically, I rely on the Coale-McNeil decomposition of the female age at first marriage to identify a possible marital search duration. Coale and McNeil decompose the age of female marriage into a skewed term, possibly related to search, and a normally distributed residual. Using the American Vital Statistics for the age of marriage, I estimate the mean skewed term in 35 states for every year between 1968 and 1995. Using the Current Population Survey for income inequality, mean female education, and other covariates on the state-year level, I show that while residual male income inequality is positively correlated with mean female age at first marriage, it may be *negatively* correlated with mean marital search duration. This surprising finding implies that the channel that links female marital search to male income inequality is not necessarily the one intuitively adopted by the literature.

Furthermore, female education is also positively correlated with female age at first marriage, but by my method, it is found to be *negatively* or not at all correlated with marital search duration. This finding may be intuitively explained: while education is sometimes associated with a postponed entrance into the marriage market, the search for a mate is facilitated by interaction with classmates. Moreover, women who enter the market at an older age may be under biological pressure to accept offers. Finally, the shorter marital search of higher-educated women may also be related to their increased attractiveness, leading to more frequent marriage offers.

This paper's identification strategy relies on the fact that female age at first marriage in different countries and times is shown (Coale and McNeil (1972); hereafter CM) to be a sum of two terms: a skewed term (a convolution of a few exponential distributions) and a normally distributed residual. The identification assumption is that the skewed term is related to marital search. This assumption relies on the association of the exponential distribution with the waiting time. The CM paper, Kaneko (2003), and my findings provide empirical support for this assumption. Particularly, CM show that in French data the skewed term is indeed related to duration of stay in the marriage market. Kaneko (2003) shows in Japanese data that the skewed term is shorter when marriages are arranged. In the present paper, I show that variables that may affect search and reservation values, particularly the male-to-female sex ratio and availability of divorce, are indeed correlated with the skewed term but not with the residual one. On the other hand, variables that are likely related only to the age of entrance into the marriage market, particularly the minimal legal age of marriage, are indeed correlated with the residual term but not with the skewed one.

Furthermore, I consider some theoretical concerns, particularly the question of whether within-state variation across years in the residual male income inequality is the correct measure in the case of marriage market. Marriage rates may correlate with within-state inequality across years if they adjust to business cycles. This is a predictable relationship of marriage timing and inequality, but, to the best of my knowledge, there is no certain theoretical prediction about the relationship between search duration and business cycles. Therefore, within-state variation across years may be an inappropriate tool to test the search theory. This concern is addressed by considering long-run levels of income inequality, lagged values, and removal of state fixed effects. While the former two alternatives do not alter the results significantly, removal of state fixed effects inflates many coefficients and some coefficients change their sign. However, the *negative* effect of male inequality on the mean skewed term is found to be quite robust. The relative robustness of the effects on the skewed term also holds for other robustness checks, such as considering raw income instead of residual income. Moreover, the year fixed effects are also more consistent over years once the mean skewed term and not the mean age of marriage is the dependent variable in the regression. All these findings may be seen as supporting evidence for a behavioral interpretation of the mean skewed term.

Kaneko (2003) rewrites the CM marriage age density function such that it has a single shape parameter. This parameter is intuitive as it is directly related to the asymmetry of the marriage-age distribution. The shape parameter has a statistical interpretation. For example, for a certain value of the parameter the CM density function specializes to the extreme value distribution, while for another value it specializes to the normal distribution. In Section 6, I estimate the relationship between the variables of interest and the value of the shape parameter. I also provide examples of trends in the value of this parameter across years. The fact that different states have different trends serves as an additional illustration of the behavioral interpretation of the CM model.

The common link between the age of marriage and inequality is the well-known "mean-preserving spread" mechanism of search models. Under the mean-preserving spread, increasing the dispersion of offers raises the reservation value of the seeker and the higher reservation value leads to a longer search. In other words, the more diverse the population of potential mates is, the more selective the seeker is. This interpretation of the positive correlation between marriage age and inequality was adopted by the literature as being the most reasonable (Gould and Passerman (2002), Loughran (2002), Coughlin and Drewianka (2011), Weiss and Santos (2015)).[1] The present paper challenges this interpretation, emphasizing that age of marriage and marital search duration are not the same. Also, a very large body of research in the social sciences discusses the relationship between female education and age of marriage (Bloom and Bennett (1990), Goldin and Katz (2002), and Fiel and Ambrus (2008) are only few examples). However, to the best of my knowledge, the present paper is the first to show that while female education positively correlates with female age at first marriage, the possible correlation with marital search duration is negative or insignificant. This finding corresponds to recent studies by Chiappori et al. (2009, 2015), who structurally estimate to what extent the wife's education is an important contributor to the husband's utility. This large contribution may be one of the channels to the relatively short marital search due to the increased

[1] A notable exception is Bergstrom and Bagnoli (1993), who related male inequality to uncertainty in the marriage market in a framework without search frictions.

flow of offers when the woman is educated. Turning back to the CM model, it is noteworthy that its parameters are estimated with and without covariates in a few studies (Kaneko (2003), Bloom and Bennett (1990)). However, to the best of my knowledge, no previous study uses the CM decomposition of age of marriage into both skewed and normal terms. Finally, the proposed method may be seen as a simple alternative to structural estimation of search models (see Eckstein and Van der Berg (2007) for a survey of the search-related empirical literature).

The rest of the paper is organized as follows. Section 2 clarifies the identification strategy. Section 3 introduces the CM model of standardized marriage schedules. Section 4 presents the estimation procedure. Section 5 discusses the main results and a set of theoretical considerations and robustness checks. Section 6 examines an alternative estimation procedure that uses a more recent version of the CM density function, namely, the one used in Kaneko (2003). Section 7 concludes.

2 The identification strategy

Let Y be an observed random variable that consists of two unobserved parts, X and u, such that

$$Y = X + u \tag{1}$$

In the context of the present paper, Y is the age of marriage of a woman, X is the duration of search for a spouse, and u is the residual.

Furthermore,

$$X = W\beta + \varepsilon,$$

where W is a vector of observed covariates. The focus of the empirical section of this paper will be on male income inequality and female education, included in W.

We are interested in the estimation of β. However, X is unobservable and the econometrician is restricted to using observable Y. The identification problem is in directly regressing

$$Y = W\beta + \eta,$$

where $\eta = u + \varepsilon$. The issue here is a possible correlation between u and W. Obviously, the correlation between Y and W does not necessarily imply a correlation between X and W because the correlation between Y and W may simply follow from a correlation between u and W. In particular, higher income inequality may lead to a later entry into the marriage market due to, say, a longer investment in education and in premarital savings or an adjustment of the timing of marriage to the business cycle. In such cases, income inequality has little to do with search duration. Similar issues may arise in the case of unemployment duration and job search.

The convenient solution is to try to directly measure X or to seek some exogenous variation in W. Unfortunately, both solutions are not always feasible. This paper proposes a different approach. The idea is that in some cases, X belongs to a known family of distributions. For example, if X is search, it is a convolution of waiting periods, which are likely to be exponential terms. Therefore, the proposed procedure is:

(i) Write Y as a sum of X and a residual term u.

(ii) Using the observed data of Y, calculate a set of estimates $\widehat{E(X)}$.

(iii) Regress $\widehat{E(X)} = W\beta + \varepsilon$.

The identification assumption is that the interpretation of X is unique. In the search context, the assumption is that only search may be a convolution of exponential terms. Specifically, I utilize the CM nuptiality model. They show that the female age at first marriage has the form of Equation (1), where X is a convolution of exponential distributions of different means, and u is a normally distributed residual term. CM back this identification assumption (they use the term "interpretation") with evidence from a survey of French couples, showing that X is a close fit to the duration of their stay in the marriage market. Further supporting evidence appears in Kaneko (2003) and in the present paper (see Sections 5 and 6). I follow steps (i)-(iii) above; in particular:

(i) I adopt the CM decomposition.

(ii) Using the U.S. Vital Statistics, I estimate a panel of $\widehat{E_{jt}(X)}$, where j is a state and t is a year.

(iii) I regress $\widehat{E_{jt}(X)} = W_{jt}\beta + \varepsilon_{jt}$, where W_{jt} is a vector of socioeconomic covariates, aggregated on the state-year level.

Finally, I compare the estimated coefficients to the coefficients of a naive regression $\bar{Y}_{jt} = W_{jt}\widetilde{\beta} + \widetilde{\varepsilon}_{jt}$. The differences between the coefficients of the two regressions constitute the main contribution of the present paper.

3 The Coale-McNeil nuptiality model

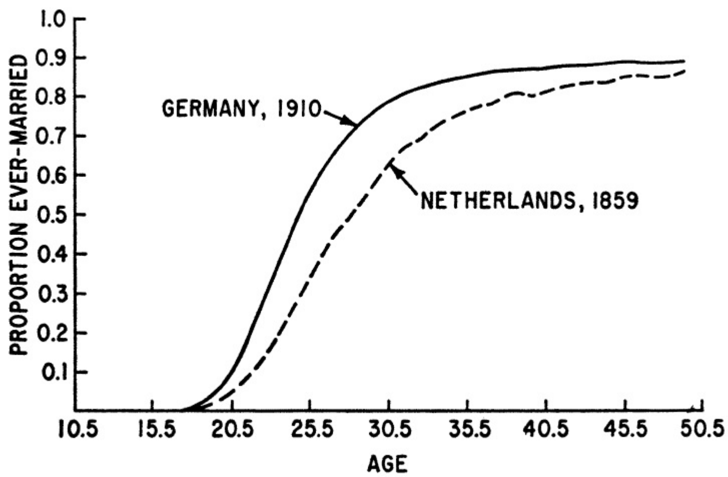
3.1 The standard schedule of the risk of first marriage

In a seminal paper, Coale (1971) summarizes several years of work at the Office of Population Research in Princeton, which he headed for a long tenure. He observes that female age at first marriage follows a standard distribution, identical for different countries and cohorts. The only differences lie in the minimal age of marriage, the final proportion of those who eventually marry, and the age at which this final proportion is reached. In other words, the schedule of marriages by age has a common functional form and the differences between countries and cohorts lie only in the values of three parameters. Figure 1 shows the proportion of ever-married women by age for the Netherlands in 1859 and for Germany in 1910. The two curves seem different. However, Figure 2 redraws the two curves so that they have a common starting point, the final proportion of those who eventually marry is one, and adjusting the horizontal scale such that this final proportion is reached at the same age. As seen in Figure 2, the two standardized curves practically coincide. Furthermore, Coale (1971) fits the standard schedule of risk of first marriage (number of first marriages of women of standard age x divided by the number of single women of standard age x) by the double-exponential function. He fits the function to the marriage schedule of Swedish women in the 1860s and obtains the following probability density function:

$$r_s(x) = 0.174e^{-4.411e^{-0.309x}}$$

Thus, if marriages start at age a_0 and the horizontal scale is compressed by factor k, the risk at

Figure 1: Female age of marriage in the Netherlands, 1859, and Germany, 1910



Source: Coale (1971).

age a is fitted by

$$r(a) = (0.174/k)e^{-4.411e^{(-0.309/k)(a-a_0)}} \tag{2}$$

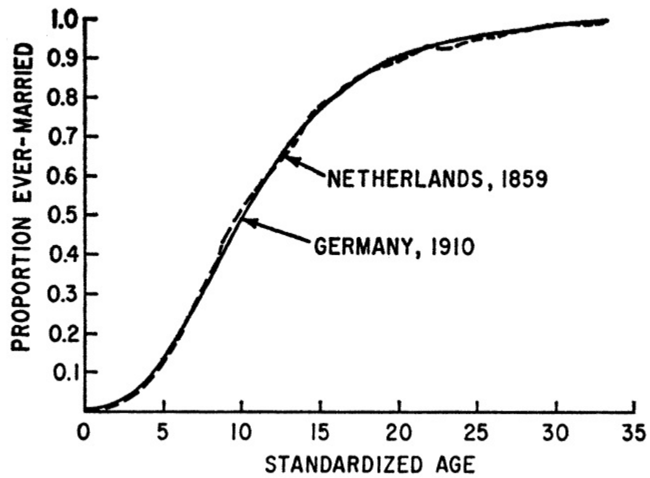
In a subsequent study, CM develop the statistical theory behind the function in Equation (2). The idea is that a random variable with a risk function converging to a constant asymptote r can be stripped of at least one exponential term. This is done according to the identity

$$F_1(Y) = F(Y) + F'(Y)/r,$$

where $F_1(Y)$ is the distribution after the exponential term has been removed. It is possible to repeat this exercise infinitely many times because asymptotic risks in case of a double-exponential risk function remain constant after the removal of exponential terms. This leads to the finding that the female age at first marriage (Y) can be written as

$$Y = \sum_{i=1}^{\infty} Z_i \tag{3}$$

Figure 2: Standardized female age of marriage in Netherlands, 1859, and Germany, 1910



Source: Coale (1971).

where Z_i are latent independent random variables exponentially distributed with $E(Z_i) = \mu_i = \frac{1}{\alpha+(i-1)\lambda}$. The parameters α and λ define the shape of the curve. The role of these parameters turns intuitive by the fact that the ratio of α's and the ratio of λ's in two populations are both equal to the ratio of the standard deviations of the age of marriage in the two populations.

Thus, Y can be rewritten as Equation (1), $Y = X + u$, where

$$X = \sum_{i=1}^{m} Z_i \tag{4}$$

and

$$u = \sum_{i=m+1}^{\infty} Z_i$$

Furthermore, CM show that u converges very fast, as a function of m, to a normal distribution. Therefore, Y consists of an infinite number of independent exponential terms with diminishing means. The sum of the few terms with relatively large means forms X, while the sum of the remaining sequence of exponential terms with negligible means forms the normally distributed residual. In the empirical analysis, CM set $m = 3$ and let u be a normally distributed term with a proper mean and variance. It is found to be an extremely good fit to the empirical distribution of the female age at first marriage (Y).

3.2 Parametrization

Finally, CM prove that from Equation (3), it follows that the probability density function of Y is closely fitted by

$$f_Y(y) = \frac{\lambda}{\Gamma(\alpha/\lambda)} e^{-\alpha(y-\varphi)-e^{-\lambda(y-\varphi)}} \tag{5}$$

where Γ indicates the gamma function, $\varphi = a + (1/\lambda)\psi(\alpha/\lambda)$, where $\psi = \Gamma'/\Gamma$ is the digamma function, and a is simply the mean age at first marriage.[2] CM fit the standard curve by setting the parameters according to the standard schedule of marriages in Sweden in the 1860s. In the present paper, I calibrate the parameters such that the standard curve fits the actual age of marriage distribution for each state-year combination in the U.S. Vital Statistics.

3.3 Relation to search duration

As CM point out, the density function given by Equation (5) has possible behavioral implications. Specifically, CM interpret the standard schedule of first marriage frequencies as the combination of normal distribution of attainment of marriageable age and three exponentially distributed delays.

In Section 5 I show that the (normally distributed) residual is, indeed, correlated with attainment of marriageable age. By contrast, the skewed term is not correlated with the minimal legal age of marriage.

[2] This density function belongs to the family of general log gamma distributions (see Kaneko (2003)).

CM proceed with their discussion by considering empirical examples. First, they estimate the age of "entry" into the marriage market in the U.S. in the 1960s as 15.6 and the mean duration of the skewed part as 73 months. The age of 15.6 corresponds to the average minimal legal marriage age in the U.S. at that time. They proceed to the following interpretation of "entry" and "delay":

"We may conjecture that the age of becoming marriageable is the age at which serious dating, or going steady begins; that the longest delay is the time between becoming marriageable and meeting (or starting to keep frequent company with) the eventual husband; and that the two shorter delays are the period between beginning to date the future husband and engagement, and between engagement and marriage."

Under this interpretation, the longest exponential term is related to search duration. CM test their conjecture with data from a 1959 survey of French married couples. One of the questions in the survey was how long before the marriage had the couple known each other. CM's interpretation of their model is that acquaintance duration should correspond to the sum of the model's second and third exponential terms. They indeed find a very close agreement between the distribution of acquaintance duration in the data and the distribution predicted by the model, except for couples who were acquainted with each other long before marriage, e.g., couples who knew each other from childhood. CM conclude that this finding supports the behavioral interpretation of their model and estimate the mean duration of marital search (the first exponential term) in the French sample to be around four years.

3.4 Further evidence of CM interpretation

Kaneko (2003) finds that the standard marriage schedule of Japanese women born in 1935-1960 is more symmetric than the one implied by the CM standard curve. He figures out that this difference may be explained by the prevalence of arranged marriages in Japan in those years. Once arranged marriages are removed and socio-economic covariates are controlled for, the schedule becomes much closer to the CM one. The suggested interpretation of this finding is that arranged marriages in Japan are behaviorally different from marriages in Western Europe used by CM to calibrate their model. The behavioral interpretation here is similar to the one proposed by CM: the asymmetry of the marriage age density function is due to frictions in the marriage market. In particular, the frictions may be facilitated by arranged marriages.

4 Estimation

4.1 Calibration of the density function parameters

The CM density function, given in Equation (5), has three parameters because the standardization process described above has three degrees of freedom (minimal age of marriage, final proportion of those who eventually marry and the age at which this proportion is reached). I calibrate the set of parameters $\{a, \lambda, \alpha\}_{jt}$ for state j in year t, using the NBER collection of Marriage and Divorce Data of the National Vital Statistics System of the National Center for Health Statistics.[3] I limit the sample to first marriages of white women. The data in the collection covers the period from 1968 to 1995 for 35 states.

CM fit the density function (Equation (5)) to the observed marriage schedules of cohorts of women. By contrast, I am interested in the year-place specific pattern, and, therefore, interpret the density function as a cross-state distribution of all first marriages formed in state j in year t. As discussed already in Coale (1971), the cohort schedule and the cross-section distribution are quite similarly well described by the density function given in Equation (5). Theoretically, one could calibrate CM's parameters for completed cohorts and then regress the estimated mean duration of the skewed term on the covariates in the year when the cohort was 16 years old (i.e., the age of "entry" according to CM's calibration). However, such a procedure would generate only few observations because only cohorts born until 1955 may be considered as complete by 1995, the last year in the dataset. Moreover, the results of my analysis show that despite the different interpretation, the CM model fits the cross-state data very well, as we shall see later in this section.

The calibration procedure is:

[3] Downloadable at http://www.nber.org/data/marrdivo.html.

(i) Calibrate a_{jt} as the mean age at first marriage in state j in year t.

(ii) Calibrate $\{\lambda, \alpha\}_{jt}$ by maximizing the log-likelihood of the density function (Equation (5)) using data of first marriages in state j in year t.

The quality of fit is measured by the Kolmogorov-Smirnov-style statistic

$$q_{jt} = max_x(|\hat{F}_{jt}(x) - \widetilde{F}_{jt}(x)|)$$

where $\hat{F}_{jt}(x)$ is the observed cumulative frequency of first marriages up to age x, and $\widetilde{F}_{jt}(x)$ is the calibrated cumulative density.

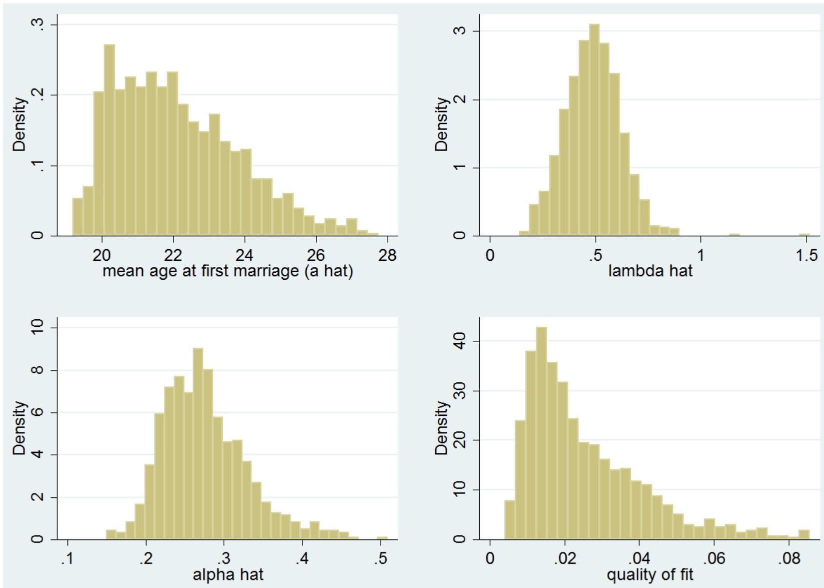
Figure 3 summarizes the calibration results. The first panel of the figure shows the distribution of the mean age at first marriage ($\widehat{a}_{jt}$). The following two panels show, respectively, the distributions of $\widehat{\lambda}_{jt}$ and $\widehat{\alpha}_{jt}$. The fourth panel shows the distribution of the quality of fit statistic q_{jt}. The figure shows that the mean age at first marriage is positively skewed, while the estimates of $\widehat{\lambda}_{jt}$ and $\widehat{\alpha}_{jt}$ are distributed almost symmetrically. Despite the cross-state interpretation of the model, different from the original CM interpretation, and despite the fit to the actual age rather than the standard one, the fit is almost as good as in the CM paper. The quality of fit statistic never reaches 0.1 and its median is only 0.02, while in the CM paper the absolute value of the area between the empirical and the fitted standard curves is, similarly, 0.016.

4.2 Regressions

The main purpose of this paper is to compare the coefficients of a naive regression, which means directly regressing the mean age of marriage $\bar{Y}_{jt}$ on a vector of covariates W_{jt} (with main interest in male income inequality and female education) to the coefficients of the regression of the estimated mean skewed term $\widehat{E_{jt}(X)}$ on the same set of covariates. The two regressions are

$$\bar{Y}_{jt} = W_{jt}\widetilde{\beta} + \widetilde{\gamma}_j + \widetilde{\delta}_t + \widetilde{\varepsilon}_{jt} \tag{6}$$

Figure 3: Calibration results



and

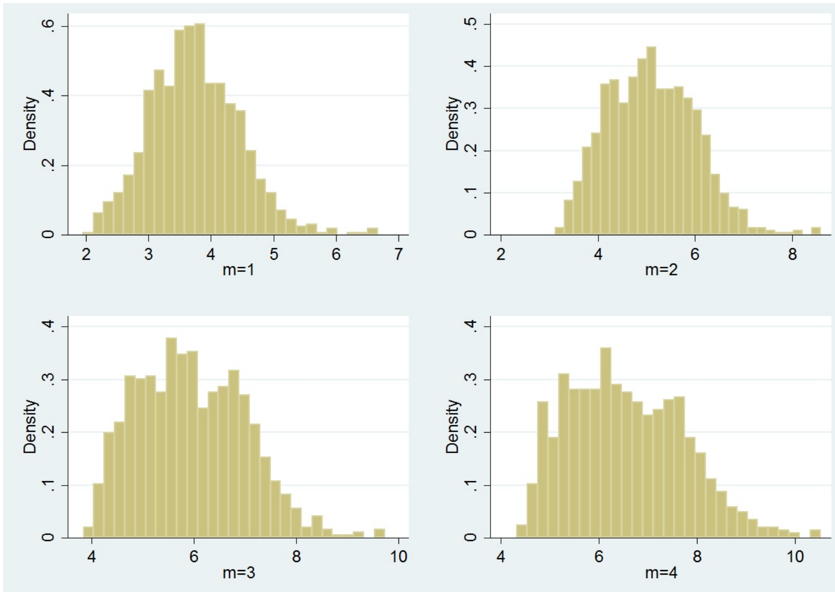
$$\widehat{E_{jt}(X)} = W_{jt}\beta + \gamma_j + \delta_t + \varepsilon_{jt}, \tag{7}$$

where γ_j and δ_t are, respectively, the state and year fixed effects. The estimated mean of the skewed term, by Equation (4), is

$$\widehat{E_{jt}(X)} = \sum_{i=1}^{m} \widehat{\mu}_{jt} = \sum_{i=1}^{m} \frac{1}{\widehat{\alpha}_{jt} + (1-i)\widehat{\lambda}_{jt}}$$

I estimate Equation (7) for $m = 1$ and $m = 3$. The former value provides the longest skewed term, and the latter one is CM's suggestion for the total delay between entry into the marriage market and marriage. The larger m is, the smaller the residual term is, such that $\widehat{E(X)}$ approaches $\bar{Y}$. The distribution of $\widehat{E(X)}$ is shown in Figure 4 for different values of m. The histograms show that the estimated means of the skewed term vary from two to ten years. The estimated mean of the longest component ($m = 1$) is distributed almost symmetrically around 3.5 years but may be as long as 6.5

Figure 4: Distribution of $\widehat{E(X)}$, the estimated mean of the skewed term



years.

4.3 Covariates

The vector of covariates W consists of 24 variables, divided into four groups. All covariates relate to state j in year t. The first three groups are calculated from the Integrated Public Use Microseries of the Current Population Survey (CPS) for years 1968-1995 (Flood et al. (2015)). The fourth group is adopted from various studies. The covariates are:

Group 1: Income. Six variables are included: mean, standard deviation, and cubic root of the third central moment of male and female residual logged personal income distributions. The residual is derived by regressing

$$ln(I_{ijt}) = \theta_1 married_{ijt} + \theta_2 male_{ijt} + \theta_3 age_{ijt} + \theta_4 age^2_{ijt} + educ_{ijt}\pi + year_t + state_j + \nu_{ijt}, \quad (8)$$

where I_{ijt} is the personal income (restricted to positive values[4]) of individual i in state j in year t, $married_{ijt}$ and $male_{ijt}$ are dummy variables, and $educ_{ijt}$ is a vector of three dummies for advanced levels of education: high school diploma, some college, and college degree. The fixed effects are $year_t$ and $state_j$.[5]

Group 2: Education. Six variables are included, which represent the proportion of men and women in each of the three educational groups listed above, in state j in year t.

Group 3: Sex ratio. Five variables are included, representing the male-to-female odds ratio at ages 16-19, 20-23, 24-27, 28-31, and 32-35, in state j in year t.

Group 4: Laws. Seven variables are included:

-The minimal legal age of marriage. It consists of four variables: minimal age of marriage of males and females, with and without parental consent (Blank et al. (2009)).

-A dummy for Early Legal Access (ELA), i.e., the availability of oral contraception for single childless women below age 21 (Bailey et al. (2011)).

-A dummy for the possibility of no-fault divorce (Ashbaugh et al. (2002)).

-A dummy for legal abortion (Levine et al. (1999)).

Because not all states appear in both CPS and the Vital Statistics for the same years, the final dataset consists of 734 observations (year-state combinations). Summary statistics are presented in Table I.

[4]An alternative is to set $ln(I_{ijt}+1)$ as the dependent variable and to restrict the sample to non-negative incomes. However, the case of zero income is very age-related. Thus, such a regression would be sensitive to the restriction of the range of ages in the sample. Generally, the $ln(I_{ijt}+1)$ specification provides similar signs of the coefficients in the main regressions (Equations (6) and (7)) but of a smaller magnitude and statistical significance than the $ln(I_{ijt})$ specification.

[5]The specification in Equation (8) includes socio-economic variables that are generally associated in the economic literature with income premiums. The idea is that the residual income should be idiosyncratic and more closely associated with the search process. However, alternative definitions of residual income may exclude endogenous variables (education and marital status) and selection of women into the labor force (excluding women from the estimation of Equation (8)). Results with these modifications are reported in the Supplementary Appendix. The limit case of these modifications is the consideration of the raw income, rather than the residual income, discussed in Section 5.4. These modifications have only negligible effect on the results.

Table I: Summary statistics

Variable			mean	sd	min	max
		year			1968	1995
		mean female age at first marriage	22.54	1.657	19.52	27.73
log income	male	mean	9.282	0.442	7.997	10.18
		standard deviation	1.246	0.114	0.881	1.698
		$\sqrt[3]{E(\widehat{ln(I) - ln(I)})^3}$	-1.615	0.230	-2.454	-0.824
	female	mean	8.305	0.531	6.985	9.377
		standard deviation	1.586	0.195	1.097	2.21
		$\sqrt[3]{E(ln(I) - \widehat{E}(ln(I)))^3}$	-1.931	0.279	-2.560	-0.989
residual log income	male	mean	0.002	0.100	-0.261	0.457
		standard deviation	1.039	0.112	0.697	1.562
		$\sqrt[3]{E(\nu - \widehat{E}(\nu))^3}$	-1.381	0.258	-2.425	-0.575
	female	mean	-0.007	0.11	-0.424	0.292
		standard deviation	1.528	0.205	1.003	2.155
		$\sqrt[3]{E(\nu - \widehat{E}(\nu))^3}$	-1.888	0.302	-2.575	-0.888
education	male	high school	0.086	0.112	0.006	0.437
		some college	0.254	0.056	0.103	0.421
		college degree	0.101	0.061	0.021	0.368
	female	high school	0.090	0.126	0	0.464
		some college	0.248	0.065	0.096	0.475
		college degree	0.065	0.057	0.002	0.324
N			734			

Variable		mean	sd	min	max
Sex ratio	16-19	1.034	0.237	0.147	2.609
	20-23	0.973	0.2	0.413	2.067
	24-27	1.024	0.21	0.513	3.548
	28-31	1.008	0.174	0.555	1.924
	32-35	0.998	0.170	0.347	1.699
minimal age of marriage	no consent (f)	18.07	0.788	15	21
	consent (f)	15.52	2.928	0	18
	no consent (m)	18.28	0.862	17	21
	consent (m)	16	3.052	0	19
ELA		0.941	0.235	0	1
no-fault divorce		0.860	0.348	0	1
abortion		0.931	0.254	0	1
$\widehat{\alpha}$		0.263	0.042	0.167	0.438
$\widehat{\lambda}$		0.460	0.125	0.141	1.517
q		0.021	0.014	0.004	0.085
calibrated length of search $\widehat{E(X)}$	$m = 1$	3.905	0.613	2.285	6.001
	$m = 2$	5.341	0.778	3.389	7.674
	$m = 3$	6.229	0.918	4.109	8.879
	$m = 4$	6.873	1.038	4.534	9.746
	$m = 5$	7.379	1.143	4.864	10.72
N		734			

5 Results

The results of estimating the regressions of the marriage age and its mean skewed part (Equations (6) and (7)) are reported in Tables II and III.

Table II shows the results of the fixed-effects regressions with the standard errors clustered by state. Table III shows the results without state fixed effects and the differences between the two estimations are discussed in Section 5.3. Column 1 in each of the tables presents the results for Equation (7) for $m = 1$, i.e., the mean longest skewed term as the dependent variable. Column 2 presents the results for $m = 3$. Column 3 presents the results for Equation (6) , where the dependent variable is the mean age at first marriage. Columns 4-6 repeat the estimation with the sample restricted to the 1968-1990 period, excluding the years of increasing non-marital cohabitation.

5.1 Identification

Before discussing the coefficients of male income inequality and female education, let us focus on the coefficients of some of the control variables. While these variables are not the main variables of interest in this paper, they are helpful in testing the identification assumption. These variables plausibly discriminate between "age of entry into the marriage market" and "duration of marital search." They include sex ratio, age of legal marriage, and possibility of divorce. Sex ratio affects search as it determines the flow of meetings with available mates of the opposite sex. The results show that sex ratio (number of men per woman) is indeed correlated only with $\widehat{E(X)}$ (columns 1-2 and 4-5). The correlation is mostly positive, indicating that a higher male-to-female ratio is associated with a longer waiting time (or a longer search). Theoretically, the longer search can be explained by the fact that the higher the male-to-female ratio is, the higher the selectiveness of the women is.

Furthermore, legal marriage age is related to the entrance into the market (see Iyigun and Lafortune (2016), who use the minimal legal marriage age as an instrumental variable). Indeed, the results show that it is correlated with $\bar{Y}$ but not with $\widehat{E(X)}$. Interestingly, while the coefficient of the minimal female legal marriage age without parental consent is intuitively positive, the coefficient of the corresponding minimal male legal marriage age is negative. A possible explanation is that this effect is due to an omitted variable bias: a higher male legal marriage age (conditional on the female one) is associated with a more conservative social structure, which, in turn, is associated with a lower female age of marriage. Regarding other laws (ELA, divorce, abortion), no statistically significant relationship with the female age at first marriage is found (columns 3 and 6), probably because during the period covered in data the variation in these variables is small (see the summary statistics in Table I). However, the coefficient of the availability of no-fault divorce in columns 1-2 and 4-5 is statistically significant and negative. Intuitively, in the context of search, women are more likely both to receive and to accept marriage proposals if marriage can be relatively easily dissolved. Similar evidence of the effect no-fault divorce has on family formation and marital capital appears in Alesina and Giuliano (2006), Drewianka (2006), and Stevenson (2007).

5.2 Main variables of interest

The main coefficients of interest include those of male residual income and female education. Let us first observe columns 3 and 6 of Table II, which correspond to the naive regression of the mean age at first marriage $\bar{Y}$ (Equation (6)). Consistently with the literature, the first three moments of residual male income are positively correlated with female age at first marriage. On the other hand, female income is found not to be related. The coefficients are similar in the full and restricted samples. Regarding education, the results show different coefficients in the full and restricted samples. While for the period 1968-1995, male and female education has a strong and positive correlation with mean female age at first marriage, for the period 1968-1990 the effects are smaller.

Let us now turn to the regressions of Equation (7), where the dependent variable is $\widehat{E(X)}$ (columns 1-2 and 4-5). First, it is clearly observed how the results converge to the results of Equation (6) (columns 3 and 6) as m increases, because removing more and more exponential terms with relatively large means erodes the residual term. Note that CM set $m = 3$ in their empirical

Table II: Regressions of the mean and the mean of the skewed term of female age at first marriage with state fixed effects

		(1)	(2)	(3)	(4)	(5)	(6)
		full sample (1968-1995)			subsample (1968-1990)		
		$\widehat{E(X)}$		$\bar{Y}$	$\widehat{E(X)}$		$\bar{Y}$
		$m=1$	$m=3$		$m=1$	$m=3$	
residual logged income	mean (m)	-1.842***	-0.891**	0.922***	-1.503***	-0.591*	0.846**
		(0.394)	(0.334)	(0.332)	(0.352)	(0.319)	(0.368)
	standard deviation (m)	-1.917***	-0.981**	1.310***	-1.333**	-0.600	1.223***
		(0.578)	(0.456)	(0.479)	(0.517)	(0.410)	(0.383)
	$\sqrt[3]{E(\nu-\widehat{E(\nu)})^3}$ (m)	-0.505***	-0.250*	0.419**	-0.365**	-0.169	0.422***
		(0.177)	(0.141)	(0.182)	(0.173)	(0.140)	(0.123)
	mean (f)	-0.227	-0.0125	0.221	-0.316	-0.147	-0.0433
		(0.253)	(0.252)	(0.308)	(0.195)	(0.195)	(0.218)
	standard deviation (f)	-0.729**	-0.368	0.234	-0.501	-0.304	-0.225
		(0.355)	(0.227)	(0.402)	(0.431)	(0.291)	(0.390)
	$\sqrt[3]{E(\nu-\widehat{E(\nu)})^3}$ (f)	-0.331**	-0.192*	0.158	-0.231	-0.160	-0.0697
		(0.131)	(0.103)	(0.166)	(0.206)	(0.146)	(0.171)
education	high school (m)	0.550	0.268	2.008	0.770	-1.949*	-3.708**
		(0.755)	(1.193)	(1.406)	(1.387)	(1.141)	(1.693)
	some college (m)	0.550	0.207	0.549	-0.154	-0.445	0.808
		(0.597)	(0.601)	(0.612)	(0.899)	(0.548)	(0.672)
	college degree (m)	-0.451	0.668	4.055***	0.405	0.813	2.086**
		(0.935)	(0.908)	(0.713)	(1.129)	(0.931)	(0.830)
	high school (f)	-1.772**	0.493	3.108***	0.709	1.213	1.224
		(0.735)	(0.843)	(1.029)	(1.373)	(1.083)	(0.920)
	some college (f)	-0.0211	0.452	1.115*	-0.848	0.172	2.144**
		(0.588)	(0.598)	(0.611)	(0.962)	(0.619)	(0.804)
	college degree (f)	-2.327**	0.626	4.994***	1.373	2.038*	1.707
		(0.857)	(0.987)	(1.351)	(1.519)	(1.156)	(1.352)
sex ratio	16-19	0.0976*	0.0362	-0.0604	0.0748	0.0534	0.0145
		(0.0559)	(0.0381)	(0.0544)	(0.0464)	(0.0355)	(0.0440)
	20-23	0.112	0.0643	0.00532	0.168**	0.101*	-0.0120
		(0.0741)	(0.0549)	(0.0778)	(0.0696)	(0.0566)	(0.0640)
	24-27	-0.0903*	-0.0620	0.0337	0.128**	0.0909***	0.000114
		(0.0523)	(0.0439)	(0.0419)	(0.0628)	(0.0294)	(0.0784)
	28-31	0.0752	0.0473	-0.00984	0.0967	0.0522	-0.0473
		(0.0728)	(0.0518)	(0.0601)	(0.0654)	(0.0519)	(0.0550)
	32-35	0.000352	0.0266	0.0213	0.0676	0.0636	0.0556
		(0.0654)	(0.0539)	(0.0657)	(0.0702)	(0.0531)	(0.0573)
minimal age of marriage	no consent (f)	0.00827	0.0184	0.143***	0.0245	0.0339	0.144***
		(0.0658)	(0.0580)	(0.0497)	(0.0599)	(0.0522)	(0.0444)
	consent (f)	0.0644	0.00640	-0.0773	0.0650	0.00867	-0.0803
		(0.0517)	(0.0524)	(0.0720)	(0.0508)	(0.0478)	(0.0649)
	no consent (m)	0.0374	0.0103	-0.199***	0.0346	0.0149	-0.190***
		(0.0512)	(0.0474)	(0.0558)	(0.0467)	(0.0382)	(0.0457)
	consent (m)	-0.0663	-0.0264	0.0759	-0.0692	-0.0304	0.0758
		(0.0494)	(0.0502)	(0.0681)	(0.0485)	(0.0458)	(0.0616)
other	ELA	-0.142	-0.143	-0.0564	-0.135	-0.126	-0.0808
		(0.141)	(0.153)	(0.213)	(0.126)	(0.133)	(0.171)
	no-fault divorce	-0.180**	-0.207*	-0.151	-0.180**	-0.161**	-0.0979
		(0.0729)	(0.108)	(0.133)	(0.0738)	(0.0703)	(0.0861)
	abortion	-0.202	-0.160	-0.263	-0.206	-0.112	-0.228
		(0.257)	(0.155)	(0.242)	(0.238)	(0.145)	(0.197)
	year and state FE	Yes	Yes	Yes	Yes	Yes	Yes
	constant	3.95	5.30	20.40	2.97	4.57	20.72
	observations	734	734	734	569	569	569

Table III: Regressions of the mean and the mean of the skewed term of female age at first marriage without state fixed effects

		(1)	(2)	(3)	(4)	(5)	(6)
		full sample (1968-1995)			subsample (1968-1990)		
		$\widehat{E(X)}$		$\bar{Y}$	$\widehat{E(X)}$		$\bar{Y}$
		$m=1$	$m=3$		$m=1$	$m=3$	
residual logged income	mean (m)	-1.704***	-2.366***	-3.878***	-1.526***	-1.790***	-2.921***
		(0.470)	(0.455)	(0.813)	(0.456)	(0.391)	(0.883)
	standard deviation (m)	-0.977	-3.566***	-6.873***	-1.074*	-2.811***	-4.686***
		(0.724)	(0.675)	(1.271)	(0.562)	(0.442)	(1.305)
	$\sqrt[3]{E(\nu-\widehat{E}(\nu))^3}$ (m)	-0.0930	-0.972***	-2.173***	-0.137	-0.723***	-1.470***
		(0.268)	(0.282)	(0.472)	(0.208)	(0.201)	(0.432)
	mean (f)	0.113	0.581	2.603***	0.312	0.158	1.026
		(0.465)	(0.574)	(0.932)	(0.415)	(0.480)	(0.828)
	standard deviation (f)	0.401	0.595	2.021	0.532	0.344	1.394
		(0.643)	(0.840)	(1.526)	(0.725)	(0.646)	(1.389)
	$\sqrt[3]{E(\nu-\widehat{E}(\nu))^3}$ (f)	0.299	0.239	0.497	0.239	0.0771	0.326
		(0.327)	(0.361)	(0.744)	(0.361)	(0.288)	(0.704)
education	high school (m)	1.805	1.604	7.268**	-0.909	-7.781**	-11.26**
		(1.283)	(1.436)	(3.407)	(3.123)	(3.324)	(5.505)
	some college (m)	1.548	1.203	0.339	0.689	1.847	3.772
		(1.548)	(1.902)	(3.060)	(1.518)	(1.854)	(3.397)
	college degree (m)	1.030	7.507***	18.76***	2.154	7.750***	16.25***
		(2.052)	(2.224)	(4.837)	(2.121)	(1.970)	(5.008)
	high school (f)	-4.284***	-3.202**	-1.894	-4.283*	-10.48***	-24.25***
		(1.294)	(1.501)	(2.859)	(2.507)	(2.631)	(6.104)
	some college (f)	0.0822	-1.306	-1.952	-0.199	-0.776	-0.245
		(1.149)	(1.551)	(3.399)	(1.124)	(1.357)	(3.257)
	college degree (f)	0.0958	1.798	7.474	5.571	6.175	9.363
		(2.987)	(3.227)	(4.755)	(4.470)	(4.060)	(5.883)
sex ratio	16-19	0.258***	0.276***	0.0933	0.210***	0.298***	0.165
		(0.0759)	(0.0850)	(0.136)	(0.0681)	(0.0800)	(0.131)
	20-23	0.223**	0.163**	-0.0705	0.352***	0.178*	-0.238
		(0.0866)	(0.0730)	(0.214)	(0.104)	(0.0933)	(0.240)
	24-27	-0.122	0.0127	0.278	0.0420	0.0201	-0.0922
		(0.0859)	(0.110)	(0.247)	(0.101)	(0.100)	(0.224)
	28-31	-0.0589	0.0303	0.235	-0.0442	0.0140	0.105
		(0.0905)	(0.0953)	(0.249)	(0.110)	(0.123)	(0.222)
	32-35	0.0458	0.181	0.280	0.0979	0.172	0.217
		(0.111)	(0.119)	(0.237)	(0.107)	(0.108)	(0.246)
minimal age of marriage	no consent (f)	0.0221	0.0639	0.175	0.0284	0.0366	0.0689
		(0.0431)	(0.0519)	(0.124)	(0.0457)	(0.0506)	(0.119)
	consent (f)	0.0684	-0.0193	0.0663	0.0324	-0.0193	0.146
		(0.0532)	(0.0423)	(0.162)	(0.0558)	(0.0452)	(0.160)
	no consent (m)	0.0646	0.0330	0.112	0.0676	0.0635	0.184
		(0.0669)	(0.0853)	(0.135)	(0.0595)	(0.0667)	(0.118)
	consent (m)	-0.0743	-0.00109	-0.00918	-0.0405	-0.00610	-0.0976
		(0.0520)	(0.0384)	(0.156)	(0.0540)	(0.0418)	(0.155)
other	ELA	-0.0794	-0.217	-0.0829	-0.114	-0.211	-0.0358
		(0.131)	(0.146)	(0.204)	(0.118)	(0.131)	(0.189)
	no-fault divorce	0.0419	-0.0639	-0.0446	-0.0103	-0.0999	-0.0426
		(0.145)	(0.123)	(0.227)	(0.129)	(0.111)	(0.203)
	abortion	0.220	0.152	-0.341	0.254	0.176	-0.300
		(0.205)	(0.176)	(0.291)	(0.204)	(0.168)	(0.262)
	year and state FE	Yes	Yes	Yes	Yes	Yes	Yes
	constant	1.809	4.350	14.91	1.254	4.074	15.68
	observations	734	734	734	569	569	569

analysis, as the residual term for $m = 3$ is close to following a normal distribution.

Surprisingly, and contrary to the results of the naive regression in columns 3 and 6, the moments of male residual income are negatively associated with $\widehat{E(X)}$. Furthermore, female residual income moments are statistically significant, and the coefficients are also negative (but smaller, in absolute terms, than the male income coefficients). Regarding education, male education is not related to $\widehat{E(X)}$, while female education is negatively related for $m = 1$, in contrast to its relation to $\bar{Y}$. For $m = 3$, education is mostly not statistically significant. The interpretation is intuitive. While education sometimes implies a postponed entrance into the marriage market, it may be associated with a shorter search. First, women interact with their male classmates, which increases the likelihood of meeting a future husband. Second, women may be under biological pressure to accept offers if they enter the market later. Third, a more educated woman is a more attractive mate, which leads to more frequent marriage offers. The fact that the statistically significant negative correlation is observed only for $m = 1$ implies that only the largest exponential component (the longest delay) is negatively correlated with female education.

Importantly, the opposite signs are not a result of some mechanical negative correlation between $\bar{Y}$ and $\widehat{E(X)}$. On the contrary, the correlation is positive and high: 0.92 for $m = 1$ and 0.87 for $m = 3$. The correlation between $\widehat{E(X)}$ and the residual term $\widehat{u} = \bar{Y} - \widehat{E(X)}$ is also positive, but smaller: 0.34 for $m = 1$ and 0.46 for $m = 3$.

5.3 Sensitivity to state fixed effects

Equations (6) and (7) include state fixed effects. This is done for two reasons. First, considering geographical fixed effects follows the previous studies (Gould and Passerman (2002), Loughran (2002)). Second, state fixed effects capture long-run social and economic differences between states, correlated with inequality. However, in this case, state fixed effects may also raise a problem. While state fixed effects capture long-run or permanent differences between states, regressions with these effects may fail to capture the true structural relationship between income inequality and age of marriage. The within-state variation across years in income inequality is related to business cycles. Because participation in the marriage market is a dynamic process, individuals may endogenously respond to business cycles by postponing marriage to better times. Thus, the positive relationship between marriage age and inequality may reflect variation across business cycles while the theoretical relationship between business cycles and search duration is not evident.

Table III presents the results of Equations (6) and (7) after they have been reestimated without

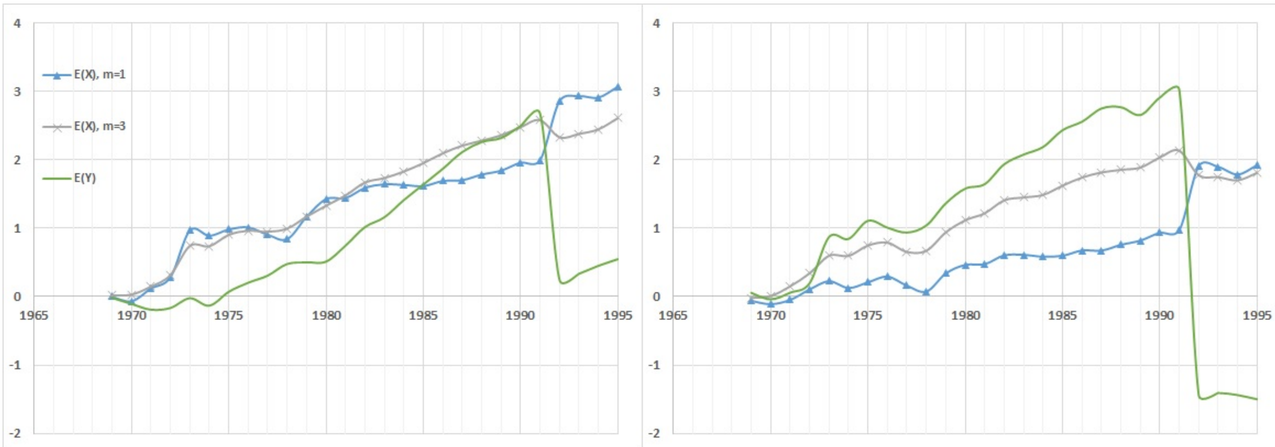
state fixed effects. Some coefficients are heavily inflated by removal of state fixed effects. This is not surprising because state fixed effects explain some 15-20% of the cross-state variation in the mean age at first marriage and in the mean skewed term. Moreover, some coefficients change their sign. This is especially remarkable in columns 3 and 6, where the dependent variable is the mean age at first marriage. In particular, the signs of the coefficients of the moments of male residual income turn negative, contrary to what is observed in columns 3 and 6 of Table II. This is a result of omitting many of the variables previously absorbed by state fixed effects. For example, adding the rate of urbanization to the regression without state fixed effects reduces the amplitude of the negative coefficient of male income inequality by 20%.

However, the results in columns 1, 2, 4, and 5 are more robust to excluding state fixed effects

than the results in columns 3 and 6. The mean residual income is still negatively related to the longest skewed term ($m = 1$) and the coefficient of the residual income standard deviation is negative but not statistically significant. In particular, the coefficients of the income moments in column 4, where the sample excludes the 1990s, are very similar to the corresponding coefficients in Table II. Furthermore, the coefficients of the educational variables are also more robust to excluding state fixed effects in columns 1, 2, 4, and 5 than in columns 3 and 6. These results mean that the long-term or permanent differences between states which are correlated with inequality are less related to the skewed term than to the residual one.

One way to visualize the difference between the results with and without state fixed effects as well as the difference between the years before and after 1990 is to plot the *year* fixed effects. These effects

Figure 5: Year fixed effects with (left) and without (right) state fixed effects



are plotted in Figure 5. We observe a positive trend in the year effects on both $\bar{Y}$ and $\widehat{E(X)}$ starting in the early 1970s. However, the trend is sharper for $\bar{Y}$ than for $\widehat{E(X)}$. Under the identification assumption, this finding implies that the common factors that change over time are more related to the entrance into the market than to search frictions. There is a downward discontinuity in the year effects on $\bar{Y}$ after 1990, which is not observed in the year effects on $\widehat{E(X)}$. Furthermore, this discontinuity is sharper without state fixed effects. The figure clearly shows that $\widehat{E(X)}$ is a more consistent measure than mean age of marriage $\bar{Y}$. It serves as an additional justification for using $\widehat{E(X)}$ to test theoretical hypotheses with regard to duration of stay in the marriage market.

5.4 Residual versus raw income

The measure of income inequality is computed as a residual that removes effects associated with the standard Mincer Equation as well as year and state fixed effects (Equation (8)). Use of the residual serves two purposes. First, it is a better measure of individuals' permanent incomes, as opposed to their current incomes. For example, the cross-sectional standard deviation of incomes in a given state and year is affected by the population age distribution, but that variation should not be relevant to marital search decisions because each individual (who does not die first) will pass through the entire age distribution during his or her lifetime. Residual income after removal of covariates such as age and education is likely to be permanent and thus more relevant to marital search. Moreover, residual income is associated with (unobservable by the econometrician) personal traits that are not as easy to search for as age and education. Therefore, the relationship between residual inequality and marital search duration is directly related to search efforts, as emphasized in search theory.

Table IV presents the results of estimation of Equations (6) and (7), with and without state fixed effects, using moments of raw income instead of residual income. Again, the coefficients in Equation (7) reported in columns 1, 2, 4, and 5 are found to be more robust than the coefficients in Equation (6) reported in columns 3 and 6. While the effects of the mean and standard deviation of male income on the mean age of marriage, reported in columns 3 and 6, are not statistically significant, the effects on the mean skewed term, reported in columns 1, 2, 4, and 5, are quite similar to the results in Table II. Without state fixed effects, the coefficients of Equation (7) are mostly smaller in absolute terms than those in Table III.

5.5 Estimation using long-run and lagged income inequality

An issue related to state fixed effects is considering long-run average levels of income and income inequality and considering lagged mean and standard deviation of income. The long-run income mean and standard deviation remove fluctuations related to business cycles. Furthermore, the dynamic nature of the marriage process should incorporate the lagged levels of inequality because individuals are single only if they did not marry earlier. Table V presents the results of the regressions with five-year average moments of male residual income distribution. Table VI presents the estimation results of the baseline model but with the addition of the moments of male residual income from three years earlier. Columns 1-3 in each of the tables correspond to the regressions with state fixed effects. Columns 4-6 correspond to the regressions without such effects. The results are in line with

Table IV: Regressions of the mean and the mean of the skewed term of female age at first marriage with raw income moments of male income distribution, 1968 to 1995

		(1)	(2)	(3)	(4)	(5)	(6)
		With state fixed effects			Without state fixed effects		
		$\widehat{E(X)}$		$\bar{Y}$	$\widehat{E(X)}$		$\bar{Y}$
		$m=1$	$m=3$		$m=1$	$m=3$	
logged income	mean (m)	-1.722***	-0.965***	0.470	-0.898*	-0.918*	-0.628
		(0.434)	(0.303)	(0.395)	(0.453)	(0.490)	(0.868)
	standard deviation (m)	-1.310**	-0.822***	0.360	-0.658	-1.673*	-2.389
		(0.530)	(0.284)	(0.551)	(0.686)	(0.876)	(1.527)
	mean (f)	-0.286	0.168	0.699***	0.882**	1.710***	2.649***
		(0.209)	(0.199)	(0.238)	(0.358)	(0.411)	(0.722)
	standard deviation (f)	-1.050**	-0.264	0.871**	0.818	1.767**	2.727*
		(0.387)	(0.277)	(0.413)	(0.576)	(0.655)	(1.454)
all other variables in Equations (6) and (7)		yes	yes	yes	yes	yes	yes
state fixed effects		yes	yes	yes	no	no	no
year fixed effects		yes	yes	yes	yes	yes	yes
observations		734	734	734	734	734	734

Note: Standard errors are clustered by state. Statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

the ones in Tables II and III. The lagged mean residual income (Table VI) provides coefficients of the same sign as the current mean residual income but of a smaller magnitude. The lagged standard deviation is not statistically significant in a regression with fixed effects and has a positive effect on the skewed term in the regression without state fixed effects (column 4 in Table VI). This is the only place throughout the regressions in this paper where we observe a positive and statistically significant effect of male income inequality on the mean skewed term of female marriage age.

5.6 Age structure of the population

Young participants in the marriage market may be more patient than older ones. This notion suggests that the effect of inequality on the average age at marriage should vary with the age distribution of the population. In particular, it would suggest the need to add to the regressions

Table V: Regressions of the mean and the mean of the skewed term of female age at first marriage with five-years-average moments of male income distribution, 1968 to 1995

		(1)	(2)	(3)	(4)	(5)	(6)
		With state fixed effects			Without state fixed effects		
		$\widehat{E(X)}$		$\bar{Y}$	$\widehat{E(X)}$		$\bar{Y}$
		$m=1$	$m=3$		$m=1$	$m=3$	
residual male logged income (5 years average moments)	mean	-1.515***	-0.655*	0.569	-1.805***	-2.614***	-4.686***
		(0.463)	(0.358)	(0.410)	(0.536)	(0.574)	(0.926)
	standard deviation	-1.929**	-0.710	2.772**	0.0880	-4.689***	-11.79***
		(0.735)	(0.764)	(1.018)	(0.949)	(1.011)	(2.335)
all other variables in Equations (6) and (7)		yes	yes	yes	yes	yes	yes
controls (laws)		yes	yes	yes	yes	yes	yes
state fixed effects		yes	yes	yes	no	no	no
year fixed effects		yes	yes	yes	yes	yes	yes
observations		734	734	734	734	734	734

Note: Standard errors are clustered by state. Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

Table VI: Regressions of the mean and the mean of the skewed term of female age at first marriage with lagged moments of male income distribution, 1968 to 1995

		(1)	(2)	(3)	(4)	(5)	(6)
		With state fixed effects			Without state fixed effects		
		$\widehat{E(X)}$		$\bar{Y}$	$\widehat{E(X)}$		$\bar{Y}$
		$m=1$	$m=3$		$m=1$	$m=3$	
residual male logged income	mean	-1.796***	-0.932**	0.838***	-1.717***	-2.140***	-2.971***
		(0.450)	(0.346)	(0.297)	(0.511)	(0.412)	(0.721)
	standard deviation	-1.858***	-0.991**	1.213**	-1.068	-3.118***	-5.503***
		(0.602)	(0.463)	(0.469)	(0.717)	(0.755)	(1.105)
residual male logged income 3 years earlier	mean	-0.0179	0.0780	0.00856	-0.113	-0.330	-1.226**
		(0.282)	(0.174)	(0.191)	(0.319)	(0.338)	(0.497)
	standard deviation	0.270	0.204	0.296	0.689**	-0.755**	-3.487***
		(0.360)	(0.213)	(0.440)	(0.319)	(0.370)	(0.739)
all other variables in Equations (6) and (7)		yes	yes	yes	yes	yes	yes
controls (laws)		yes	yes	yes	yes	yes	yes
state fixed effects		yes	yes	yes	no	no	no
year fixed effects		yes	yes	yes	yes	yes	yes
observations		734	734	734	734	734	734

Note: Standard errors are clustered by state. Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

interaction terms between inequality and the mean age of young men and women.

Thus, I reestimate regressions including, in addition to the baseline variables, the mean age of men and women between 16 and 40 years old. These two variables are added separately and in interaction with the standard deviation of the residual income of males of the same age range.

The results with state fixed effects are clearer than the results without them, and they are in line with the prediction. A larger age gap between men and women in the population (older men, younger women) is associated with a shorter search: the coefficients of age without interaction with inequality in columns 1 and 2 of the table are negative for the male age and positive for the female age. However, this effect diminishes when male income inequality rises: the coefficients of the interaction terms have the opposite sign of the coefficients without interaction. In particular, the effect of the interaction between male income inequality and mean female age on the mean skewed term is negative. Under the identification assumption, this result means that an older female population has a lower reservation value in the marital search paradigm, conditional on the mean male age. The mean male age has the opposite effect: the coefficient of the interaction with inequality is positive. The interpretation is that the search of women for a husband takes longer when the male population is both older and more unequal, conditional on the mean female age. The corresponding effects on the mean age of marriage (columns 3 and 6) are not statistically significant. Estimation without state fixed effects provides results that are not statistically significant with regard to the female age but are similar to the fixed effects regression results with regard to the male age.

6 Estimation using single shape parameter

An intuitive alternative way to see the relationship between the asymmetry of marriage age distribution and covariates such as male income inequality and female education is to utilize the fact that the CM density function given in Equation (5) belongs to the family of general log gamma

Table VII: Regressions of the mean and the mean of the skewed term of female age at first marriage with mean age of men and women between 16 and 40 years old, 1968 to 1995

	(1)	(2)	(3)	(4)	(5)	(6)
	With state fixed effects			Without state fixed effects		
	$\widehat{E(X)}$		$\bar{Y}$	$\widehat{E(X)}$		$\bar{Y}$
	$m=1$	$m=3$		$m=1$	$m=3$	
standard deviation × mean male age	0.407***	0.270***	-0.0228	0.357**	0.343**	0.238
	(0.104)	(0.0976)	(0.170)	(0.155)	(0.168)	(0.308)
standard deviation × mean female age	-0.282**	-0.209*	-0.0243	-0.0105	-0.0762	-0.295
	(0.109)	(0.107)	(0.166)	(0.153)	(0.161)	(0.318)
residual log income	-1.321***	-0.684***	0.597**	-1.325***	-1.975***	-2.903***
	(0.289)	(0.247)	(0.226)	(0.380)	(0.389)	(0.690)
standard deviation of the residual log income	-4.701**	-2.446	1.577	-9.856**	-9.094**	-2.620
	(2.192)	(1.798)	(2.445)	(3.672)	(4.162)	(7.676)
mean male age	-0.452***	-0.263**	0.0996	-0.127	-0.106	-0.0421
	(0.122)	(0.101)	(0.175)	(0.179)	(0.198)	(0.365)
mean female age	0.347***	0.273**	0.0517	0.0500	0.151	0.366
	(0.118)	(0.110)	(0.178)	(0.165)	(0.179)	(0.326)
all other variables in Equations (6) and (7)	yes	yes	yes	yes	yes	yes
state fixed effects	yes	yes	yes	no	no	no
year fixed effects	yes	yes	yes	yes	yes	yes
observations	734	734	734	734	734	734

Note: The mean male and female ages are in the range of 16 to 40 years old. Standard errors are clustered by state. Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.

distributions (GLG; see Kaneko (2003)). The shape parameter of the GLG distribution is related to CM's shape parameters α and γ in the following way:

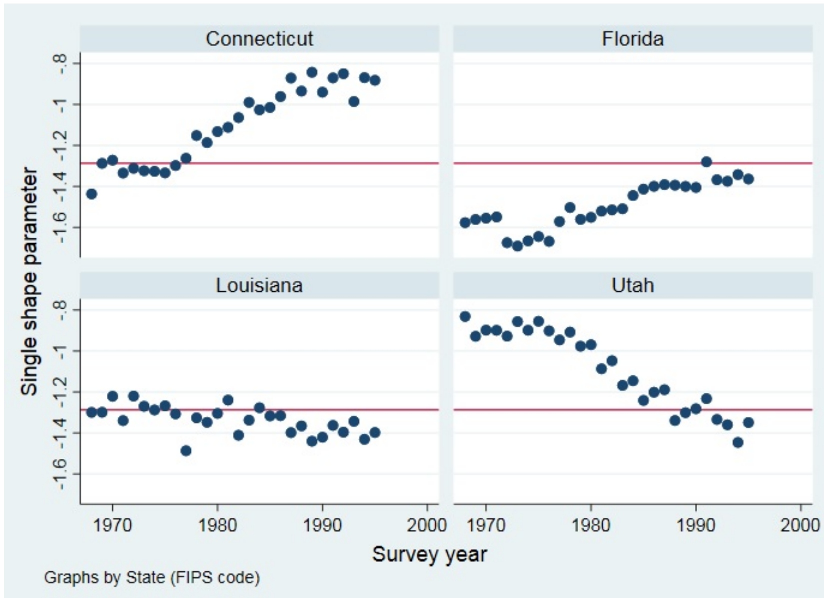
$$\delta = -(\frac{\alpha}{\gamma})^{-0.5}$$

The GLG family includes many well-known distributions. For instance, when $\delta \rightarrow 0$, it converges to the normal distribution and when δ=-1 it specializes to the extreme value distribution. According to CM's original calibration of the parameters that fit the Swedish cohorts of the 1860s, δ=-1.287. The closer δ is to zero, the more symmetric the distribution is. For instance, Kaneko (2003) finds that Japanese women who married after World War II have a value of δ slightly above -1, which means a more symmetric distribution than that in CM. This, he argues, can be explained by behavioral differences between those cohorts of Japanese women and their Western European women counterparts, such as the prominence of arranged marriages in post-war Japan. This finding supports the interpretation of the shape of CM's distribution as being related to marital search. Under this interpretation of the findings regarding Japanese women, arranged marriages are associated with a shorter search.

Values of $\hat{\delta}$ with respect to the calibration results in my data show different trends across states.

Figure 6 presents four examples: Connecticut, Florida, Louisiana, and Utah. The horizontal line corresponds to the CM value (-1.287). In Connecticut, the values of $\hat{\delta}$ in the early 1970s are quite stable and correspond to the CM value. However, starting in the mid-1970s, $\hat{\delta}$ starts to rise, making the distribution more asymmetric. From the mid-1980s on the value stabilizes to close to -0.8, which is quite similar to the value of Japanese women estimated in Kaneko (2003). In Florida, the upward trend is also observed but the values are lower than in Connecticut and never reach the CM value. In Louisiana, $\hat{\delta}$ is quite constant and is very similar to the CM value. Finally, in Utah the trend is downward: it starts close to the Japanese (or the later Connecticut) value and decreases to the CM value.

Figure 6: The trends in the single shape parameter



I further estimate Equation (7) with $\hat{\delta}$ as the dependent variable. Table VIII presents the estimation results. Columns 1 and 3 report the regression results with residual logged income (with and without state fixed effects), while columns 2 and 4 report the results with raw logged income. The results are very much in line with the results in Table II, but deviate from the results in Table III. In regressions with state fixed effects, a higher male income, male income inequality, and the third moment of the male income distribution are associated with a higher value of $\hat{\delta}$, which means a more symmetrical distribution of the age of marriage and, according to the behavioral interpretation, a shorter search. Without state fixed effects, the effect of male income inequality is negative, which means a longer search. Female education is also associated with a more symmetrical distribution, with a moderate difference when state fixed effects are removed. In summary, the results imply a less asymmetric distribution of female marriage age as a function of the main variables of interest (male inequality and female education) when the regression is estimated with state fixed effects. Under the identification assumption, this relationship means a shorter marital search.

Table VIII: Regressions of the shape parameter (Kaneko delta), 1968 to 1995

		(1)	(3)	(3)	(4)
		With state FE		Without state FE	
residual male logged income	mean	0.700***		0.0891	
		(0.145)		(0.166)	
	standard deviation	0.691***		-0.838***	
		(0.206)		(0.285)	
male logged income	mean		0.611***		0.281
			(0.188)		(0.195)
	standard deviation		0.428*		-0.394
			(0.235)		(0.323)
education	high school (m)				
	some college (m)				
	college degree (m)				
	high school (f)	1.386***	1.539***	1.614***	1.675***
		(0.321)	(0.313)	(0.572)	(0.533)
	some college (f)	0.200	0.159	-0.454	-0.280
		(0.300)	(0.344)	(0.522)	(0.549)
	college degree (f)	1.526***	1.229**	0.479	0.163
		(0.474)	(0.591)	(1.027)	(0.988)
all other variables in Equations (6) and (7)		yes	yes	yes	yes
state fixed effects		yes	yes	no	no
year fixed effects		yes	yes	yes	yes
N		734	734	734	734

Note: Standard errors are clustered by state. Statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

7 Concluding remarks

This paper's empirical results show that contrary to the intuition adopted by the literature, female marital search duration is not necessarily positively correlated with male income inequality and female education. The actual correlation may be negative or at least ambiguous. One possible explanation of this finding is that increasing inequality is associated not only with a different reservation value but also with a different search strategy. Furthermore, the results correspond to recent research on female marriage return to education, as women in regions with a larger share of educated females are suggestively found to experience a shorter marital search. This finding may be attributed to a lower selectiveness of these women or to their increased attractiveness.

The empirical strategy proposed here relies on the identification of a latent variable (marital search duration). The identification assumption is that a certain component of the female age-at-first-marriage distribution has a unique interpretation. Specifically, the identification assumption is that the skewed convolution of exponential terms relates to search. This identification assumption is supported by the empirical finding indicating that the sex ratio and divorce laws are related only to this skewed term. On the other hand, the minimal legal marriage age is found to be related only to the residual term, which may be interpreted as the age of entry into the marriage market.

The alternative specifications and robustness checks show that the coefficients of the variables of interest on the mean skewed term are more robust than the coefficients on the mean age of marriage. Moreover, the year fixed effects are also more consistent once the mean skewed term and not the mean age of marriage is considered as the dependent variable. These findings imply that the mean skewed term is a relatively robust measure of the marriage market paradigm.

A question that is left for further research is why the age of marriage positively correlates with residual male income inequality. If it is not explained by marital search duration, this positive correlation needs an alternative explanation. My findings imply that this correlation is mediated by the residual term, which may be related to the age of entry into the marriage market. A possible channel that links male inequality with later marriage is uncertainty. Empirically, Gottschalk and Moffitt (2009) and Moffitt and Gottschalk (2002, 2011, 2012), who analyze data from the 1960s to 1990s, find that almost half of the inequality in the income of American men is due to transitory shocks. The residual income is stripped of some of the permanent factors and gives more weight to the transitory shocks. Indeed, the coefficients in Table II where residual income is considered are of a larger amplitude than in Table IV where raw income is considered. From a theoretical perspective, Bergstrom and Bagnoli (1993) show that one does not need to assume search frictions to explain a positive relationship between inequality and age of marriage if at least some of the inequality is related to uncertainty about future incomes of young men. However, their model assumes an exogenous difference between men and women in order to address a large spousal age gap. In my data from American Vital Statistics from 1968 to 1995, the mean spousal age gap is relatively small and not very volatile. It is 2.5 years in the late 1960s, increases to 2.9 in the mid-1980s, and then decreases to 2.6 in the 1990s. The relatively low level and small volatility of the spousal age gap is consistent with the idea that entry into the marriage market is delayed when uncertainty is high. Women do not marry men who are much older; instead, they wait until the uncertainty about their coevals is resolved.

The second important result is the negative relationship between male inequality and female duration of search. This result is not surprising once we take into account that some fundamental parameters of the search mechanism may be different when the level of inequality is higher. A more unequal society should be more stratified. If search is within "classes" (Burdett and Coles (1997) and Smith (2006) develop a theory where marital search is endogenously within classes) the search pools are smaller in the case of a more stratified society. Smaller pools may be easily associated with a shorter search. Moreover, a very unequal society converges to one where the marginal cost of marrying a "wrong" partner is very high and competition over "proper" marital partners is intense. Even though an extremely stratified society is more related to the pre-industrial epoch than to America between the 1960s and the 1990s, the period analyzed in this paper, it can be seen as a theoretical limit case. Because of the intense competition over partners within the narrow class, an extremely stratified society is more likely to observe arranged marriage. In such a case, the search is under pressure to be short and takes place as early as possible, even before the mate (especially if it is a girl) reaches a marriageable age.

Finally, the proposed method is in fact a simple alternative to structural estimation. A serious limitation of structural models is the difficulty of the estimation. A model is generally a set of simultaneous equations with implicit functions of interest. Because of the complicated nature of economics, structural models that produce a closed-form probability density function of the variables of interest are rare. Thus, estimation generally cannot be easily replicated by an inexperienced reader. Moreover, some of the model's structural assumptions are motivated by computational concerns. Therefore, the need for a simple alternative is compelling. The method proposed in this paper may serve, in some cases, as such an alternative.

Compliance with Ethical Standards: The author declares that he has no conflict of interest.

References

Alesina, Alberto F., and Paola Giuliano. "Divorce, fertility and the value of marriage," *NBER Working Paper No. 12375*, (2006).

Bailey, M. J., Guldi, M., Davido, A., & Buzuviz, E. (2011). "Early Legal Access: Laws and Policies Governing Contraceptive Access, 1960-1980," *Unpublished manuscript.*

Bergstrom, T. C., & Bagnoli, M. (1993). "Courtship as a waiting game," *Journal of Political Economy*, 101(1), 185-202.

Blank, R. M., Charles, K. K., & Sallee, J. M. (2009). "A cautionary tale about the use of administrative data: Evidence from age of marriage laws," *American Economic Journal: Applied Economics*, 1(2), 128-149.

Blau, F. D., Kahn, L. M., & Waldfogel, J. (2000). "Understanding young women's marriage decisions: The role of labor and marriage market conditions," *Industrial & Labor Relations Review*, 53(4), 624-647.

Bloom, D. E., & Bennett, N. G. (1990). "Modeling American marriage patterns," *Journal of the American Statistical Association*, 85(412), 1009-1017.

Burdett, K., & Coles, M. G. (1997). "Marriage and class," *The Quarterly Journal of Economics*, 112(1), 141-168.

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

Chiappori, P. A., Salanié, B., & Weiss, Y. (2015). "Partner choice and the marital college premium: Analyzing marital patterns over several decades," *unpublished manuscript.*

Coale, A. J. (1971). "Age patterns of marriage," *Population Studies*, 25(2), 193-214.

Coale, A. J., & McNeil, D. R. (1972). “The distribution by age of the frequency of first marriage in a female cohort,” *Journal of the American Statistical Association*, 67(340), 743-749.

Coughlin, T. P., & Drewianka, S. D. (2011). “Can rising inequality explain aggregate trends in marriage? Evidence from U.S. states, 1977-2005,” *The BE Journal of Economic Analysis & Policy*, 11(1).

Danziger, L., & Neuman, S. (1999). “On the age at marriage: theory and evidence from Jews and Moslems in Israel,” *Journal of economic behavior & organization*, 40(2), 179-193.

Drewianka, Scott. (2008). "Divorce law and family formation." *Journal of Population Economics* , 21(2), 485-503.

Eckstein, Z., & Van den Berg, G. J. (2007). “Empirical labor search: A survey,” *Journal of Econometrics*, 136(2), 531-564.

Field E., & Ambrus, A. (2008). “Early marriage, age of menarche, and female schooling attainment in Bangladesh,” *Journal of Political Economy,* 116(5), 881-930.

Flood, S., King, M., Ruggles, S., & Warren, J. R. (2015). Integrated Public Use Microdata Series, Current Population Survey: Version 4.0, University of Minnesota.

Goldin, C., & Katz, L.F. (2002). “The power of the pill: Oral contraceptives and women’s career and marriage decisions,” *Journal of Political Economy*, 110(4), 730-770.

Gottschalk, P., & Moffitt, R. (2009). “The rising instability of US earnings,” *Journal of Economic Perspectives*, 23(4), 3-24.

Gould, E. D., & Paserman, M. D. (2003). Waiting for Mr. Right: Rising inequality and declining marriage rates,” *Journal of Urban Economics*, 53(2), 257-281.

Iyigun, M., & and Lafortune, J. (2016). “Why wait? A century of education, marriage timing and gender roles,” IZA Discussion Paper 9671.

Kaneko, R. (2003). "Elaboration of the Coale-McNeil nuptiality model as the generalized log gamma distribution: A new identity and empirical enhancements," *Demographic Research*, 9, 223-262.

Keeley M. C. (1977). "The economics of family formation: An investigation of the age at first marriage," *Economic Inquiry*, 15(2), 238-250.

Keeley M. C. (1979). "An analysis of the age pattern of first marriage," *International Economic Review*, 20(2), 527-544.

Laumann, E. O., Gagnon J. H., Michael R. T., & Michaels S. (1992). *National Health and Social Life Survey, 1992: United States*, Ann Arbor, MI: Inter-university Consortium for Political and Social Research, ICPSR06647-v2.

Levine, P. B., Staiger, D., Kane, T. J., & Zimmerman, D. J. (1999). "Roe v Wade and American fertility," *American Journal of Public Health*, 89(2), 199-203.

Li, Xue (2014). *Economic Inequality and Marriage Formation*, Ph.D. thesis, Louisiana State University.

Loughran, D. S. (2002). "The effect of male wage inequality on female age at first marriage," *Review of Economics and Statistics*, 84(2), 237-250.

Moffitt, R. A., & Gottschalk, P. (2002). "Trends in the transitory variance of earnings in the United States," *The Economic Journal*, 112(478), C68-C73.

Moffitt, R. A., & Gottschalk, P. (2011). "Trends in the covariance structure of earnings in the US: 1969-1987," *The Journal of Economic Inequality*, 9(3), 439-459.

Moffitt, R. A., & Gottschalk, P. (2012). "Trends in the transitory variance of male earnings methods and evidence," *Journal of Human Resources*, 47(1), 204-236.

Smith, L. (2006). “The marriage model with search frictions,” *Journal of political Economy*, 114(6), 1124-1144.

Stevenson, B. (2007). “The impact of divorce laws on marriage-specific capital,” *Journal of Labor Economics,* 25(1), 75-94.

Yu, J., & Xie, Y. (2015). “Changes in the determinants of marriage entry in post-reform urban China,” *Demography*, 52(6), 1869-1892.

Vlosky, D. A., & Monroe, P. A. (2002). “The effective dates of no-fault divorce laws in the 50 states,” *Family Relations*, 51(4), 317-324.

Santos, C., & Weiss, D. (2016). “’Why not settle down already?’ A quantitative analysis of the delay in marriage,” *International Economic Review*, 57(2), 425-452.