WHO MARRIES WHOM AND WHY

Eugene Choo & Aloysius Siow *

University of Toronto

November 10 2003

Abstract

This paper proposes and estimates a static non-parametric transferable utility model of the marriage market. The model rationalizes the standard interpretation of marriage rate regressions as well as pointing out its limitations. The model was used to estimate US marital behavior in 1971/72 and 1981/82. The estimates show that the gains to marriage for young adults fell substantially over the decade. It also showed that the legalization of abortion had a significant quantitative impact on the fall in the gains to marriage for young adults.

1 Introduction

Thirty years ago, Gary Becker (1973, 1974; summarized in Becker 1981) exposited a static transferable utility model of the marriage market. It is the current benchmark model of the marriage market.[1] While implications of his model have been tested and

*We thank Pierre Andre-Chiappori, Angelo Melino, Shannon Seitz and Adonis Yatchew, and seminar participants at Boston University, Northwestern University, University of Chicago, Center for Applied Microeconomics (Copenhagen), University of Rochester and the University of Toronto for useful comments. We thank Julan Al-Yassin and Jasmin Kantarevic for excellent research assistance. Eugene Choo acknowledges research support from the Connaught Funds. Siow also thanks SSHRC of Canada for research support. Research on this paper was started when the second author was visiting the department of economics at the University of Chicago. He thanks them for their hospitality.

[1] Researchers have used it to study the relationships between sex ratios and marital outcomes such as female labor supply, marriage rates, the determination of dowries and differences in spousal ages (examples include Angrist 2002, Chiappori, et. al. 2001, Edlund 2000, Grossbard-Shectman 1993, Hamilton and Siow 2000, Rao 1993, Seitz 1999, South and Lloyd 1992, South and Trent 1988). Bergstrom 1997 and Weiss 1997 provide surveys of the economics literature up to the mid-nineties. Casper and Bianchi 2002; and Waite, et. al. 2000 show his influence outside economics.

applied, it has seldom been estimated.[2] There are two problems that have to be solved before a transferable utility model of the marriage market can be estimated. First, equilibrium transfers in modern marriages are seldom observed. Second, individuals may differ by age, religion, education, wealth, ethnicity, and so on. Different types of individuals may not agree on the rankings of individuals of the opposite gender as spouses. Thus an empirical model of the marriage market should not impose too much apriori structure on the nature of preferences for marriage partners. However without apriori structure, it is unclear what can be identified from the data.

To understand the identification problem, consider a society with I types of men and J types of women participating in the marriage market. A type is defined by an age range, ethnicity, education, geographic location and so on. Each individual chooses who to marry or to remain single. For each type of man (woman), there are potentially J (I) preference parameters to characterize his (her) utility from each type of spouse and remaining single. In total, there are as many as $2 \times I \times J$ preference parameters. What is observable to a researcher? In principle, the researcher observes the quantity of each type of men in the marriage market, m_i for type i men (I observations), the quantity of each type of women, f_j for type j women (J observations), and the quantity of type i men married to type j women, μ_{ij} ($I \times J$ observations). So the total number of observables are $I + J + I \times J$. For $I, J > 2$, the number of observables are less than the number of unknown preference parameters. Thus **any** behavioral empirical model will need to make identifying assumptions to reduce the number of unknown parameters.[3]

To finesse the identification problem, demographers use a reduced form approach in the form of *marriage matching functions*, to estimate the behavior of the entire marriage market. Marriage matching functions are also a fundamental building block of two-sex models of population growth.[4]

A marriage matching function is defined as follows.[5] Let M be the vector of available men by types, $i = 1, .., I$ at that time. The i'th element of the vector M is denoted by m_i. Let F be the vector of available women by types, $j = 1, ..., J$, where the j'th element

[2]Bergstrom and Lam 1994; Suen and Lui 1999 are exceptions.
[3]Bergstrom and Lam, Hamilton and Siow 2000, Seitz 1999, Suen and Lui, Wong 2003a estimate models of the marriage market with strong identifying assumptions.
[4]See Pollak 1990a for a state of the art study.
[5]Our discussion borrows heavily from the excellent discussions in Pollak 1990b and 1990a.

of the vector is denoted by f_j. Usually, researchers associate types of individuals with their ages. Let Π be a matrix of parameters. A marriage matching function is an $I \times J$ matrix $\mu(M, F; \Pi)$, whose i, j element is μ_{ij}. Denote the number of unmarried men of type i as μ_{i0} and the number of unmarried women of type j as μ_{0j}. The marriage matching function $\mu(M, F; \Pi)$ must satisfy:

$$\mu_{0j} + \sum_{i=1}^{I} \mu_{ij} = f_j \; \forall \; j \tag{1}$$

$$\mu_{i0} + \sum_{j=1}^{J} \mu_{ij} = m_i \; \forall \; i \tag{2}$$

$$\mu_{0j}, \mu_{i0}, \mu_{ij} \geq 0 \; \forall \; i, j \tag{3}$$

Equations (1), (2) and (3) are accounting constraints. (1) says that the total number of men who marry j type women and the number of unmarried j type women must be equal to the number of available j type women for all j. Similarly (2) says that the total number of women who marry i type men and the number of unmarried i type men must be equal to the number of available i type men for all i. (3) holds because the number of unmarrieds of any type and gender, and the number of marriages between type i men and type j women must be non-negative.

Demographers usually work with matching functions with a zero spillover matching rule:

$$\mu_{ij}(M, F;\, \Pi) = \mu_{ij}(m_i, f_j;\, \alpha_{ij})$$

That is, the number of i, j matches only depends on m_i and f_j. Schoen's 1981 harmonic mean mating rule, (the current workhorse in demography) given by

$$\mu_{ij}(M, F) = \frac{\alpha_{ij} m_i f_j}{m_i + f_j} \tag{4}$$

where $\alpha_{ij} > 0$, $\sum_j \alpha_{ij} \leq 1$, and $\sum_j \alpha_{ij} \leq 1$, is a zero spillover matching rule. This matching function will satisfy all the accounting constraints, (1), (2) and (3). While zero spillover marriage matching functions are easy to estimate and use, as Pollak 1990b pointed out, the zero spillover assumption is restrictive. Holding the parameters of the marriage matching function, $\alpha_{ij}'s$, constant, changes in $m_{i'}$ and $f_{j'}$ where $i' \neq i$ or $j' \neq j$

do not affect μ_{ij}.[6] Demographers have of course recognized the importance of spillover (or substitution) effects in marriage matching function (McFarland 1972; Pollard 1997). The problem is to specify marriage matching functions which include substitution effects and yet remain identified.[7]

Another problem with marriage matching functions is that they are mostly specified as reduced form and not derived from a model of the marriage market. It is unclear how they may be used to study the impact of social interventions on the marriage market.

Instead of estimating an entire model of the marriage market, many researchers and policy analysts have used marriage rate regressions to estimate the effects of different interventions on marriage rates. The marriage rate of type j women, ρ_j^f, is:

$$\rho_j^f = \frac{\sum_i \mu_{ij}}{f_j}$$

A marriage rate regression for type j women is defined as:

$$g(\rho_j^f) = X_j{}'\beta + u_j$$

where $g(.)$ is usually the linear or log function.[8] X_j is a vector of characteristics, including policy variables, which affect the marriage rate. u_j is the error term of the regression model. Marriage rate regressions are widely used (E.g. Angrist 2002, Angrist and Evans 1999, Baker, et. al. 2003, Gruber 2000, South and Lloyd 1992, South and Trent 1988). The technique is flexible and easy to implement. The standard interpretation of marriage rate regressions assumes that the marriage rate of a particular type of individual is positively related to factors which increase the welfare gain to marriage for that type.

While easy to use, there are loose ends. First, the standard interpretation of marriage rate regressions has not been derived. Second, how are the estimated effects from male

---

[6] Marriage rate regressions also suffer from a similar defect.

[7] Pollard and Höhn 1993/94 provide the most sophisticated matching function of this kind:

$$\mu_{ij}(M,F) = \frac{m_i f_j a_i b_j}{\frac{1}{2}(\sum_k m_k a_k h_{kj} + f_k b_k h_{ik}))}$$

where a_i, b_j, and h_{ij} are weight functions that are specified by the analyst. When i and j refer to the ages of the participants, the types of individuals are ordered by age. Using this natural ordering, Pollard and Höhn suggested some plausible weight functions. But if types are also defined by ethnicity, religion and other attributes that are not naturally ordered, then it is difficult to apriori specify the weight functions. But without apriori restriction on the weights, the model is not identified.

[8] If individual data is used, then researchers estimate $G(X_j'\beta)$, the probability that a type j individual will marry.

and female marriage rate regressions related? That is, a factor which affects the female marriage rate must also affect the male marriage rate. This interdependence is ignored in marriage rate regressions. In two different applications, Angrist and this paper show that the estimated effects of the same factor on marital behavior from male and female marriage rate regressions can be wildly contradictory.

This paper proposes and estimates a static non-parametric transferable utility model of the marriage market. The model produces a simple marriage matching function with spillover effects which will fit any observed cross-section marriage distribution.[9] Our marriage matching function has an intutitive normative interpretation. Because of its normative interpretation, it can be used to do policy evaluations which are as easy to do as marriage rate regressions but without the limitations discussed earlier.

There are three conceptual benefits for considering transferable utility models of the marriage market. First, taking individual preferences as given, marriage market clearing equilibrium must satisfy all the accounting constraints, (1), (2) and (3).

Second, the reduced form for equilibrium quantities of a market clearing model do not include prices, i.e. equilibrium transfers. Thus the absence of observable transfers to the researcher may not be a problem.

Third, transferable utility models provide a solution to the identification problem discussed above. To see how the identification problem may be resolved, let the marital output of an i type male and a j type female only depends on i and j. Then there are $I \times J$ number of these marital outputs plus $I + J$ outputs of the types being single. If the behavior of the marriage market is characterized by these outputs alone, then we may be able to estimate all the parameters which are necessary to determine marital behavior. In particular, we do not have to estimate separate male and female preferences for spouses. A well known property of transferable utility models of the marriage market is that they maximize the sum of marital output in the society (For example, Roth and Sotomayor 1990; Chapter 8). Thus behavior in transferable utility models can be characterized by knowledge about marital output alone, and knowledge about male and female preferences separately is not necessary. The novelty of this paper is to exploit this property to specify a just identified econometric model of the marriage market, and

[9]Our marriage matching function also satisfies the conditions in Pollak 1990a, sufficient to generate a well posed two-sex model of population growth.

minimize apriori restrictions on male and female preferences for spousal types.[10]

In order to implement the above framework, we use McFadden's (1974) well known extreme value random utility model to generate demand and supply functions for different types of marriages. With our behavioral assumptions, the following marriage matching function is obtained:

$$\mu_{ij} = \Pi_{ij}\sqrt{(m_i - \sum_k \mu_{ik})(f_j - \sum_l \mu_{lj})}$$

The above equation says that the number of i, j marriages is proportional to the geometric average of the unmarrieds of each type. The marriage matching function is homogenous of degree one in M and F.[11]

Our model of the marriage market provides a normative interpretation to Π_{ij}. The model assumes that the total surplus or payoff from a potential i, j marriage depends on a *systematic* payoff that is common to all i, j match, and an *idiosyncratic* payoff related to the two particular potential spouses. Realized i, j marriages generated higher payoffs than other feasible marriages. The higher payoffs may be due to high systematic payoffs related to the i, j match, or high idiosyncratic payoffs, or a combination of the two. All else equal, matches that are frequently observed imply a relatively high systematic payoff to that pairing while matches that are infrequent have low systematic payoffs. We interpret Π_{ij} as measuring the *systematic* payoff to an i, j marriage relative to those types not marrying. The *systematic* qualifier is important. Π_{ij} does not measure the total (*systematic* and *idiosyncratic*) payoff to observed i, j marriages relative to those types not marrying. Our interpretation of Π_{ij} says that the systematic gains to i, j marriages are larger when the number of i, j marriages, suitably scaled, is larger. Scaling is necessary because the number of i, j marriages may be high because there are more type i men and type j women.

An important theoretical antecedent to our work is Dagsvik (2000).[12] A comparison with his work is provided in Section 6.

Using ages as the only types for males and females in the benchmark model, the second part of the paper estimates Π using data from the 1970 and 1980 *US Census,* and

[10]Both Bergstrom and Lam; Suen and Lui used this property to estimate tightly parameterized models.

[11]Pollak 1990a argues that no scale effect is a reasonable requirement for marriage matching functions.
[12]Also see Johansen and Dagsvik 1999; Dagsvik, et. al. 2001.

1971/72 and 1981/82 *Vital Statistics*. The baby boom generation came into marriageble age between the two decades and thus there were substantial changes in the population vectors between the decades. Our marriage matching function can capture some changes in marital patterns in the US between 1971/72 and 1981/82 due to changes in population vectors between the two periods. However our benchmark model could not capture the drastic fall in the marriage rate among young adults in 1981/82.

Our first attempt to explain the fall in the marriage rate among young adult was to expand the type space to include educational attainment. This expansion of the type space did not explain the fall in the marriage rate. Put another way, the gains to marriage fell between 1971/72 and 1981/82 for young adults of all educational groups.[13]

There were many social changes between 1970 and 1980 which could have affected the gains to marriage over the decade. A major change was the national legalization of abortion in 1973. Legal abortions were partially available in some states by 1970. If the partial legalization of abortions in a state reduced the gains to marriage in that state, we would expect to see lower gains to marriage in the early legalizing states relative to later legalizing states in 1970 but not in 1980. Moreover this difference in difference in the gains to marriage should be concentrated among women of child bearing age. Using marriage rate regressions, Angrist and Evans 1999 showed that the marriage rates of young men and women were lower in early legalizing states relative to later legalization states in the early seventies. We show that the estimates of the number of marriages affected are extremely sensitive to whether we use male or female marriage rate regressions. Using our framework, we show that the partial legalization of abortion in some states can explain up to twenty percent of the drop in the gains to marriage among young adults in the seventies. In doing so, we extend the standard difference in differences estimator to estimate the effect of a policy change on bivariate distributions.[14]

It should be clear that the empirical issues discussed here and the ability of transferable utility models of the marriage market to resolve many of these issues are known. The main contribution of this paper is to provide an elementary parametrization which exploits all the power of this framework.

[13]Demographers already noticed this decline in the marriage rate of young adults (E.g. Qian and Preston 1993; Qian 1998).

[14]Difference in differences estimators are usually used to estimate effects on the moments of a univariate distribution.

2 The model

We begin by describing a transferable utility model of marriage. There are I types of men and J types of women. For a type i man to marry a type j woman, he must transfer τ_{ij} amount of income to her. There are $I \times J$ sub-marriage markets for every combination of types of men and women. The marriage market clears when given equilibrium transfers, τ_{ij}, the demand by men of type i for type j spouses is equal to the supply of type j women for type i men for all i, j.

To implement the above framework empirically, we adopt the extreme value random utility model of McFadden to generate market demands for marriage partners. Each individual considers matching with a member of the opposite gender. Let the utility of male g of type i who marries a female of type j be:

$$V_{ijg} = \widetilde{\alpha}_{ij} - \tau_{ij} + \varepsilon_{ijg}, \qquad \text{where} \tag{5}$$

$\widetilde{\alpha}_{ij}$: Systematic gross return to male of type i married to female of type j.

τ_{ij} : Equilibrium transfer made by male of type i to spouse of type j.

ε_{ijg} : i.i.d. random variable with type I extreme value distribution.[15]

Equation (5) says that the payoff to person g from marrying a female of type j consists of two components, a systematic and an idiosyncratic component. The systematic component, $\widetilde{\alpha}_{ij} - \tau_{ij}$, is common to all males of type i married to type j females. The systematic return is reduced when τ_{ij}, the equilibrium transfer, is increased.

The idiosyncratic component, ε_{ijg}, measures the departure of his individual specific match payoff, V_{ijg}, from the systematic component. We assume that the distribution of ε_{ijg} does not depend on the number of type j females, f_j. Put another way, there are sufficient number of females of type j such that his idiosyncratic payoff from choosing to marry a type j female does not depend on f_j. The payoff to g from remaining unmarried, denoted by $j = 0$, is:

$$V_{i0g} = \widetilde{\alpha}_{i0} + \varepsilon_{i0g} \tag{6}$$

where ε_{i0g} is also an i.i.d. random variable with type I extreme value distribution.

Individual g will choose according to:

$$V_{ig} = \max_j \{V_{i0g}, .., V_{ijg}, .., V_{iJg}\} \tag{7}$$

[15]The random variable $\varepsilon_{ijg} \sim EV(0,1)$, with the cumulative distribution given by $F(\varepsilon) = e^{-e^{-\varepsilon}}$.

We assume that the numbers of men and women of each type is large. Let μ^d_{ij} be the number of i,j marriages demanded by i type men and μ^d_{i0} be the number of unmarried i type men. Then McFadden showed that (Appendix A includes a proof for convenience):

$$\begin{aligned}\ln \mu^d_{ij} &= \ln \mu^d_{i0} + \widetilde{\alpha}_{ij} - \widetilde{\alpha}_{i0} - \tau_{ij} \\ &= \ln \mu^d_{i0} + \alpha_{ij} - \tau_{ij}\end{aligned} \tag{8}$$

The term $\alpha_{ij} = \widetilde{\alpha}_{ij} - \widetilde{\alpha}_{i0}$, is the systematic gross return to a i type male from an i,j marriage relative to being unmarried. The above equation is a quasi-demand equation by type i men for type j spouses. [16] Unlike the usual demand equation, the transfers for non-type j women appear nominally absent in equation (8). But they are not absent as these other transfers are all embodied in $\ln \mu^d_{i0}$.

Let Γ be Euler's constant. Appendix A shows another well known result:

$$\mathbb{E}V_{ig} = \Gamma + \widetilde{\alpha}_{i0} + \ln(\frac{m_i}{\mu^d_{i0}}) \tag{9}$$

$\mathbb{E}V_{ig}$ is the expected utility of a male of type i before he sees his realizations of his ε_{ijg} for all j. Equation (9) shows that it is proportional to the log of the ratio of the number of available type i men relative to the number of type i men who choose to remain single. The expected payoff if being single is the only option is given by $\mathbb{E}V_{i0g} = \Gamma + \widetilde{\alpha}_{i0}$. If $\ln(\frac{m_i}{\mu^d_{i0}})$ is observable, it measures the expected benefit of a type i male from being able to participate in the marriage market where non participation means only choosing to be single. Let

$$q_i = \ln\left(\frac{m_i}{\mu^d_{i0}}\right) \tag{10}$$

denote the expected gains to entering the marriage market for a type i male. As shown in the appendix and section 3, the expected gains depends on preference parameters, $\widetilde{\alpha}_{ij}$ and $\widetilde{\alpha}_{i0}$, as well as transfers, τ_{ij}.

The random utility function for women is similar to that for men except that in marriage with a type i men, a type j women receives a transfer, τ_{ij}. Let $\widetilde{\gamma}_{ij}$ denote the systematic gross gain that j type women get from marrying i type men, and $\widetilde{\gamma}_{0j}$ be the systematic payoff that j type women get from remaining single. The term $\gamma_{ij} = \widetilde{\gamma}_{ij} - \widetilde{\gamma}_{0j}$, is the systematic gross gain that j type women get from marrying i type men relative to not marrying.

[16] It is not a demand curve because $\mu_{i0} = m_i - \sum_j \mu_{ij}$.

Let μ_{ij}^s be the number of i, j marriages demanded by j type women and μ_{0j}^s the number of type j women who want to remain unmarried. The quasi-supply equation of type j women who marry type i men is be given by:

$$\ln \mu_{ij}^s = \ln \mu_{0j}^s + \gamma_{ij} + \tau_{ij}. \tag{11}$$

Again, the transfers for all the other types of men other than i is embodied in $\ln \mu_{0j}^s$.

Following (9), the expected gains to entering the marriage market for a type j female is:

$$Q_j = \ln\left(\frac{f_j}{\mu_{0j}^s}\right)$$

There are $I \times J$ sub-marriage markets for every combination of types of men and women. The marriage market clears when given equilibrium transfers, τ_{ij}, the demand by men of type i for type j spouses is equal to the supply of type j women for type i men for all i, j.[17]

When the competitive marriage market for all i, j pair clears, the demand for i, j marriages is equal to the supply:

$$\mu_{ij} = \mu_{ij}^d = \mu_{ij}^s \tag{12}$$

Substituting (12) into equations (8) and (11) to get:

$$\tau_{ij} = \frac{\ln \mu_{i0} - \ln \mu_{0j} + \alpha_{ij} - \gamma_{ij}}{2} \tag{13}$$

Substituting (13) into (11), we get:

$$\ln \mu_{ij} - \frac{\ln \mu_{i0} + \ln \mu_{0j}}{2} = \frac{\alpha_{ij} + \gamma_{ij}}{2} \tag{14}$$

If we let $\pi_{ij} = \ln \Pi_{ij} = \frac{\alpha_{ij}+\gamma_{ij}}{2}$, we can rewrite Equation (14) as:

$$\Pi_{ij} = \frac{\mu_{ij}}{\sqrt{\mu_{i0}\mu_{0j}}} \tag{15}$$

which is our marriage matching function.

Equation (15) has an intuitive interpretation. The right hand side of (15) is the ratio of the number of i, j marriages to the geometric average of those types who are

[17] Chapter 9 of Roth and Sotomayor (1990) has a proof of the existence of market equilibrium for a general transferable utilities model of marriage of which ours is a special case.

unmarried. The log of the left hand side, $\ln \Pi_{ij} = \pi_{ij}$, has the interpretation as the total systematic gain to marriage per partner for **any** i, j pair relative to the total systematic gain per partner from remaining single. Put another way, one expects the systematic gains to marriage to be large for i, j pairs if one observes many i, j marriages. However there are two other explanations for numerous i, j marriages. First, there are lots of i type men and j type women in the population. Second, there are relatively more i type men and j type women in the population than other types of participants. Scaling the number of i, j marriages by the geometric average of the numbers of unmarrieds of those types control for these effects.[18]

Equation (15) is homogeneous of degree zero in population vectors and the number of marriages. From the point of view of the marriage matching function, if we assume the systematic returns as defined by π_{ij} stays fixed, doubling M and F will result in a doubling of μ. Thus our marriage matching function has no scale effect in population vectors.

2.1 Identification

A point estimate for Π_{ij} is given by $\frac{\mu_{ij}}{\sqrt{\mu_{i0}\mu_{0j}}}$. Equation (15) is non-parametric in the sense that it fits any observed marriage distribution. That is, we do not impose any apriori structure on the systematic gains to marriage. However our approach is completely parametric with respect to the idiosyncratic gains to marriage. The marriage matching function is also fully saturated in the sense that there are $I \times J$ elements in μ and there are $I \times J$ parameters in Π. In order to maintain identification of the marriage matching function, the behavioral restrictions underlying Equation (15) can only be relaxed by imposing other restrictions.

Observing Π_{ij} however, is not sufficient for us to identify the individual specific systematic returns, α_{ij} and γ_{ij}. It is also not sufficient to estimate $(\alpha_{ij} - \gamma_{ij})$, which is needed to identify the equilibrium transfers in Equation (13). In other words, knowing the systematic gains to a match is not sufficient to determine whether men pay positive

[18] The term $2\pi_{ij}$ is not the expected total gain to marriage for an i, j couple that chooses to marry each other. Observed i, j married couples get in total $2\pi_{ij}$ plus the idiosyncratic payoffs of each spouse which is the result of optimizing behavior. Since they could have married other types or not marry, the average total payoff of i, j couples who married each other relative to not marrying is weakly larger than $2\pi_{ij}$.

or negative transfers to women in equilibrium. On the other hand, in applications where τ_{ij} is also observed, then equations (13) and (15) would allow us to identify α_{ij} and γ_{ij}.[19] In these cases, equations (13) and (15) are able to fit any observed marriage distribution and transfer function. So the current model can be used to fit any finite type competitive bilateral matching market if the matching distribution and the equilibrium pricing functions are observed.

In addition to π_{ij}, equations (8) and (11) allows us to identify $\alpha_{ij} - \tau_{ij}$ and $\gamma_{ij} + \tau_{ij}$, that is:

$$\ln\left(\frac{\mu_{ij}}{\mu_{i0}}\right) = \alpha_{ij} - \tau_{ij} = n_{ij}$$
$$\ln\left(\frac{\mu_{ij}}{\mu_{0j}}\right) = \gamma_{ij} + \tau_{ij} = N_{ij}$$

We will refer to n_{ij} as the systematic gain to marriage for a type i male in an i, j marriage relative to not marrying, and N_{ij} as the systematic gain to marriage for a type j female in an i, j marriage relative to not marrying.

3 Marriage rate regressions and policy evaluations

The expected gain to entering the marriage market for a type j female denoted by Q_j, is related to the marriage rate by:

$$Q_j = \ln\left(\frac{f_j}{\mu_{0j}}\right) = -\ln\left(1 - \frac{\sum_i \mu_{ij}}{f_j}\right) \approx \frac{\sum_i \mu_{ij}}{f_j} = \rho_j^f \tag{16}$$

This approximation is accurate for small marriage rates. The marriage rate for type j females is also related to the systematic net gains N_{ij} in (**??**) according to:

$$\rho_j^f \approx Q_j = \ln\Big(1 + \sum_j \frac{\mu_{ij}}{\mu_{i0}}\Big) = \ln\Big(1 + \sum_j \exp(\gamma_{ij} + \tau_{ij})\Big) = \ln\Big(1 + \sum_j \exp(N_{ij})\Big). \tag{17}$$

Equation (17) says that the marriage rate of type j women depends positively on the systematic gross gains to marriage, γ_{ij}, and equilibrium transfers, τ_{ij}. Thus (17) provides

[19] In general, dowries should not be regarded as proxies for τ_{ij}. Variations in dowry prices reflect variations in τ_{ij} only if the variations in dowry prices are not due to changes in the value of dowry as a means of providing bridal wealth (Botticini and Siow 2003). See Edlund (2000) for an example of the problems that arise when this caveat is ignored.

a formal justification for the standard interpretation of marriage rate regressions, where the marriage rate of type j females is assumed to vary positively with factors which increase the gains to marriage for these women. Researchers estimate (17) with proxies for γ_{ij} and τ_{ij}. For example, researchers who use the sex ratio of type i men to type j women, $\frac{m_i}{f_j}$, as a regressor are assuming that $\frac{m_i}{f_j}$ and τ_{ij} are positively correlated. However, when $\frac{m_i}{f_j}$ varies, in general the entire population vectors, M and F are also varying. Thus it is difficult to generalize with the estimated coefficient of the marriage rate, ρ_j^f, on the sex ratio, $\frac{m_i}{f_j}$, from any particular sample.[20] The difficulties with marriage rate regressions discussed in the introduction also remain.

To avoid these difficulties and retain the convenience of marriage rate regressions, consider the following regression model for the total systematic gains to an i, j marriage:

$$\pi_{ij} = X_{ij}{}'\beta + u_{ij}, \tag{18}$$

where X_{ij} denote the vector of variables (including policy variables) that affect the total systematic gains to an i, j marriage. u_{ij} is an error term with mean zero and uncorrelated with X_{ij}. Since we can construct π_{ij} from equation (15), we can estimate β in equation (18).

Policy changes will induce changes in π_{ij} as captured by (18). Changes in π_{ij} will affect marital behavior via the marriage matching function described in equation (15). So given estimates of β, one can predict the effect of changes in X_{ij} on marriage behavior including marriage rates.

Estimating (18) is as easy as estimating marriage rate regressions. So we preserve the advantages of marriage rate regressions for doing policy evaluations of factors which affect the marriage market. Unlike marriage rate regressions, we do not estimate separate sets of regressions for different types of individuals or genders. Thus we will not run into the interpretive difficulties discussed in the introduction with marriage rate regressions. Also, we do not include the sex ratio as a regressor in (18). Instead, the section below considers the impact of changes in population vectors on the marriage distribution.

[20]Instrumenting the sex ratio does not solve this problem.

4 How M and F affect μ

Given the preference parameters of the system, Π_{ij}, we are often interested in how variations in the supply population vectors, M and F, affect the distribution of marriages as represented by μ. Let M^t and F^t be time varying population vectors. Then μ^t will also be time varying. Our marriage matching function may be rewritten as

$$\mu_{ij}^t = \Pi_{ij}\sqrt{\mu_{i0}^t \times \mu_{0j}^t} \tag{19}$$

$$= \Pi_{ij}\sqrt{\left(m_i^t - \sum_{k=1}^{J}\mu_{ik}^t\right)\left(f_j^t - \sum_{g=1}^{I}\mu_{gj}^t\right)} \tag{20}$$

If we take Π_{ij}, M^t and F^t as exogenously given, equation (20) defines a $I \times J$ system of quadratic equations with the $I \times J$ elements of μ^t as unknowns. This system can be reduced to an $I + J$ system with $I + J$ number of unmarrieds of each type, μ_{i0}^t and μ_{0j}^t, as unknowns. This reduced system is defined by equations (21) and (22) below. If we can solve for μ_{i0}^t and μ_{0j}^t, then the μ_{ij}^t's are fully determined by equation (19). To derive this system of equations, we sum equation (19) over all i's to get:

$$\sum_{i=1}^{I}\mu_{ij}^t = \sum_{i=1}^{I}\Pi_{ij}\sqrt{\mu_{i0}^t \times \mu_{0j}^t}$$

$$f_j^t - \mu_{0j}^t = \sum_{i=1}^{I}\Pi_{ij}\sqrt{\mu_{i0}^t \times \mu_{0j}^t} \tag{21}$$

Similarly, summing equation (19) over all j's, we get:

$$m_i^t - \mu_{i0}^t = \sum_{j=1}^{J}\Pi_{ij}\sqrt{\mu_{i0}^t \times \mu_{0j}^t} \tag{22}$$

Given population quantities M, F, μ and Π as defined in equation (15), local uniqueness of μ^* for new values of $M^* \neq M$, $F^* \neq F$ and holding Π fixed is given by the following result.

Proposition 1 *Let* $\Pi_{ij} = \frac{\mu_{ij}}{\sqrt{(m_i - \sum_{k=1}^{I}\mu_{ik})(f_j - \sum_{g=1}^{J}\mu_{gj})}}$ *and* M *and* F *be the vectors of* m_i *and* f_j *respectively. For* M^* *and* F^* *close to* M *and* F, μ^* *is uniquely determined.*

The proof using the implicit function theorem is given in Appendix B.

5 Limitations

In this section, we would like to draw attention to two limitations of our approach. The first arise from using the extreme value random utility model of McFadden to model demand. The "independence of irrelevant alternative" limitation on substitution patterns in that model of demand is well known.[21] At present, we do not know how restrictive our substitution patterns are on the marriage matching function. However a marriage matching function which allows for spillover effects and remains econometrically identified will need to have strong restrictions on these effects.

Another limitation of our static approach to the marriage market is that it ignores dynamic considerations. In particular, the value of delaying marriage at time t depends on future opportunities for marriage. Future opportunities are related to current population vectors M^t and F^t and the decisions that these individuals make. If future opportunities affect the value of not marrying, then Π_{ij}^t should be a function of these future opportunities and some exogenous preference parameters, which we denote by Ω, i.e. $\Pi_{ij}^t = \Pi(M^t, F^t, \Omega)$.

A simple way to test for the presence of these dynamic considerations is to test if Π_{ij}^t is related to M^t and F^t. We will not want to simply regress Π_{ij}^t on M^t and F^t because Π_{ij}^t is constructed using M^t and F^t. So if there is measurement error in observed population vectors, the measurement error will induce a correlation between observed population vectors and our constructed Π_{ij}^t. One way to get around this measurement error problem is to use $M^{t'}$ and $F^{t'}$ as instruments for M^t and F^t, $t \neq t'$.

If we find that Π_{ij}^t is correlated with population vectors at time t, after controlling for measurement error, this correlation is consistent with individuals being concerned about future opportunities in the marriage market. Of course the correlation may also be due to other forms of misspecification of our marriage matching model.

If there are scale effects, Π_{ij}^t may also be correlated with population vectors at time t. Scale effects are not ruled out by our model per se. Since Π_{ij}^t measures the systematic gain to marriage for an i, j pair, this gain can in principle depend on the population vectors. But unless we know the form of this dependence, forecasting with the model becomes infeasible.

[21] For example, refer to page 113 of McFadden (1974) for a discussion.

A discussion of the results from these specification tests is given in Section 7.3 of the paper.

6 Dagsvik's model

The marriage matching function in Dagsvik(2000) is defined by

$$\theta_{ij} = \frac{\mu_{ij}}{\mu_{i0}\mu_{0j}} \tag{23}$$

where θ_{ij} are unrestricted. The term θ_{ij} has a similar normative interpretation as our Π_{ij}. His model is also non-parametric and will fit any observed marriage distribution. Thus given data from a single cross section, we cannot differentiate between his model and ours in terms of fit of the data.

Empirically the two marriage matching functions differ in that Dagsvik's model has scale effects. For the simple case of one type of male and one type of female, it is easy to check that Dagsvik's model satisfies increasing returns to scale in the population vectors. The two models also employ different specification of payoffs to marriage. In his model, the payoff that male g of type i gets from marriage to female k of type j is defined by:

$$V'_{ijgk} = \widetilde{\alpha}'_{ij} + \varepsilon_{ijgk}.$$

$\widetilde{\alpha}'_{ij}$ denotes the systematic return to i type male from an i, j match. ε_{ijgk} denotes the idiosyncratic returns from a match between individual g and the j type female individual k.[22] So if he is matched with another female k' of type j, he will get a different payoff. Likewise for the payoffs of the females when they choose between different males. Since individuals in Dagsvik's model value every potential spouse differently, he cannot use price taking behavior (equilibrium transfers) to clear the marriage market. Instead, he uses the deferred acceptance algorithm and stable matching as an equilibrating device. Stability per se is not the difference between his model and ours because our equilibrium is also stable.

In contrast, our model assumes that for any type j, there are sufficient number of females of that type such that male g is indifferent between them. Likewise for any

[22] The random variable ε_{ijgk} is also assumed to have type I extreme value distribution.

type i males, female k has enough males of that type to choose from such that she is indifferent between them. So f_j does not directly affect the idiosyncratic payoff that male g gets from choosing to marry a female of type j. Likewise for female k. Given these indifference assumptions about within type spouses, we can use types specific transfers to clear the marriage market. Thus we have a transferable utilities model of the marriage market whereas Dagsvik (2000) has a non-transferable utilities model.

Analytically, our model is easier to derive. While there are differences between the two models, we are more similar to each other than other marriage matching functions. Both matching functions are built from explicit, albeit different, models of the marriage market. We follow his lead in using extreme value random utility functions.[23]

7 Changes in the estimated gains to marriage over the seventies

The objective of the empirical work is to estimate the marriage distributions by ages in 1971/72 and 1981/82. Data from the 1970 and 1980 *US Census* were used to construct the population vectors. Marriage records from the 1971/72 and 1981/82 *Vital Statistics* were used to construct the bivariate distributions of marriages. A state has to report the number of marriages to *Vital Statistics* to be in the sample. This requirement eliminated 10 states in 71/72 and 9 states in 81/82.[24]

For each period, we investigate a two year rather than one year marriage distribution because the two year distribution has less thin cells. For each period, we examine the marital behavior of individuals between the ages of 16 and 75 implied by the population vectors and preference parameters estimated from our model. Details on the construction of the data used are left to Appendix A.

In our sample of states, there were 16.0 million and 19.6 million available men and women respectively between the ages of 16 to 75 in 1970 (that is, these individuals were unmarried at the time of the census). There were 3.24 million marriages in 1971/2.

[23] Logan, et. al. 2001 also used extreme value utility functions and stable matching to construct their model of a small marriage market. Because they do not derive a closed form marriage matching function, their empirical model is significantly more difficult to estimate.

[24] Arizona, Arkansas, Colorado, Nevada, New Mexico, New York, North Dakota, Oklahoma, Texas, and Washington were excluded in 71/72 and 81/82. Colorado was added in 81/82.

Table 1: Data Summary

	US Census data in 1970	1980	Δ
Number of Available Males (M^t)	16.0 mil	23.4 mil	46%
Number of Available Females (F^t)	19.6 mil	27.2 mil	39%
Average age of Available Males	30.4	29.6	
Average age of Available Females	39.1	37.1	
	Vital Statistics data in 1971/72	1981/82	Δ
Number of Marrieds (μ^t)	3.24mil	3.45 mil	6.5%
Average age Married Males	27.1	29.1	
Average age Married Females	24.5	26.4	

There were 23.4 million and 27.2 million available men and women respectively in 1980. Although the available population increased by more than 39% over the decade, there were only 3.45 million marriages in 1981/2, an increase of 6.5%. A summary of the data set is in Table 1 below.

Figure 1a and 1b show the bivariate age distributions of the marrieds in 1971/2 and 1981/2 respectively. In both years, most marriages occured between young adults and there was strong positive assortative matching by age.

In Figure 2, we graph the 1970 and 1980 age distributions of the population vectors.[25] For both decades, there are more available men than women in the early ages and the reverse is true in the later ages. These gender differences are due to the fact that there are relatively more widows and the lower remarriage rate of divorced women. The higher remarriage rate of divorced men reduced the availability of younger women. The arrival

[25] The average age of available men and women in 1970 were 30.4 and 39.1 respectively. This gender difference reflected the larger fraction of available older women. The average age of the married men and women in 1971/2 were 27.1 and 24.5 respectively, reflecting the usual gender difference in ages of marriage. The statistics for 1980 are similar, as shown in Table 1.

of the baby boomers to the marriage market in 1980 is readily visible from the increase in the population of the availables. This arrival should have had a substantial impact on the marriage market. However, the number of young marrieds in 1980 barely increased.

7.1 Estimating the net gains to marriage by gender

Our model allows us to estimate the systematic net gain relative to not marrying, for each party in any i, j marriage. The 1971/72 estimates for type i males, given by $n_{ij}^{71} = \ln(\frac{\mu_{ij}^{71}}{\mu_{i0}^{71}})$, and j type females, given by $N_{ij}^{71} = \ln(\frac{\mu_{ij}^{71}}{\mu_{0j}^{71}})$ are compared in Figure 3.

In the 1971/2 and 1981/2 marital records, there were many age pairs which had no marriage. This is a common problem in empirical discrete choice applications and is encountered throughout the empirical section of this paper. We employ kernel smoothers to deal with this thin cell problem. The smoother estimates a function at any age by averaging or smoothing local neighbouring data points. For example the non-parametric estimator of the net gains in 1971/2 from an i, j marriage to a j type female is given by,

$$\widehat{N_{ij}^{71}} = \sum_{k=1}^{I} \omega_i(k) \cdot \ln\left(\frac{\mu_{kj}^{71}}{\mu_{0j}^{71}}\right) \tag{24}$$

$$\text{where } \omega_i(k) = \frac{(Ib)^{-1} K(\frac{i-k}{b})}{(Ib)^{-1} \sum_{m=1}^{I} K(\frac{i-m}{b})}$$

We used the normal kernel as the kernel weighting function, $K(\cdot)$. The bandwidth for the estimator denoted by b defines the width of the interval over which local averaging takes place.[26] For an age pair where zero marriage is encountered, a weighted average of neigbouring points with positive marriages are used to construct an estimate at that point. The non-parametric estimator proposed in equation [24] is standard and discussions of the approach may be found in a number references. A recent text by Yatchew 2003 provides an excellent overview. A similar estimator can be constructed for age i males.

Figure 3a plots $\widehat{n_{ij}^{71}}$ and $\widehat{N_{ij}^{71}}$ for 20 and 30 year old males and females by the ages of their spouses and Figures 3b plots them for 40 and 50 year old males and females. In Figure 3a, the distribution of $\widehat{N_{i,20}^{71}}$ $\widehat{n_{20,j}^{71}}$ are right skewed, with the 20 years old female

[26] To ensure that we do not under or over-smooth, numerous values of the smoothing parameter were attempted.

receiving the largest systematic net gain when she marries a slightly older male while the 20 years old male receiving the largest systematic net gain when he marries a slightly younger female.

Comparing the distribution of systematic net gain for a 30 years old female, $\widehat{N_{i,30}^{71}}$, with her 20 years old counterpart, we find the distribution for a 30 years old female to be more dispersed. Again she receives the largest net gain when she marries someone slightly older. If we consider the distribution for 30 year old males, $\widehat{n_{30,i}^{71}}$, we also find the distribution to be more dispersed than for his 20 year old counterpart. Again his largest net gain is to marrying someone slightly younger. [27]

Figure 3b compares the systematic net gains to marriage for 40 and 50 years old males and females. We observe that the net gains to marriage fell substantially by age. The net gains to marriage for 40 old males were higher than for 40 year old females and the marriage rate is also higher. The distribution of net gains to marriage for 40 year old females is similar to that of 50 year old males! Put another way, the age distribution of spouses of 40 year old females in 1970 is similar to the age distribution of spouses of 50 year old males. Finally, the net gains to marriage for 50 year old females were lower than the other groups.

From Figure 3, we also observe that the estimated net gains are negative which reflects the fact that the systematic net gains to marriage is smaller than not marrying. This is not surprising since at any age, most individuals do not marry.[28] The behavioral assumption of the model predicts a match to occur only when the match specific idiosyncratic utility is large.

Most of the features of the empirical distributions in Figures 3a and 3b are expected; What is new is that our model provides a normative interpretation of these empirical distributions. It is important to remember that our estimates of net gains reflect both preferences and equilibrium transfers.

[27] According to equations (16) and (17) stated earlier, the area below the transformed net gains, $\exp(n_{ij})$ and $\exp(N_{ij})$ is proportional to the type specific marriage rates. Comparing the areas under the respective transformed distributions, the area under $\exp(\widehat{n_{20,j}^{71}})$ is smaller relative to that of $\exp(\widehat{N_{i,20}^{71}})$ suggesting that the marriage rate of 20 year old females is larger than that of 20 year old males. For 30 year old individuals we observe the converse, that is, the marriage rate for males is higher than his 30 year old female counterpart. These transformed distributions not shown in the current version are available from the authors on request. These qualitative results are also apparent from Figure 4.

[28] $n_{ij} > 0$ implies $\mu_{ij} > \mu_{i0}$ which is counterfactual for all i, j.

7.2 Estimating the systematic gains to marriage

Figure 4 shows the 1971/2 and 1981/2 marriage rates by age. For both decades, the marriage rates for women are higher than that for men in the early ages and lower in the later ages. Marriage rates for both men and women were noticebly lower in 1981. The decline in marriage rates was particular sharp for young adults.

Within the context of our model, the marriage rate of a type of individual measures the expected gains to entering the marriage market relative to not entering for that type. In this light, Figure 4 shows two relevant features. First, the expected gains to entry rise rapidly by age, peak around mid-twenties and then slowly decline with age. The expected gains are larger for young women than men and reverse for older individuals. Second, the expected gains to entering the marriage market for both men and women fell substantially over the decade. It is clear that a constant total systematic gains to marriage model over the decade cannot fit the data. It should also be clear that any increasing returns to scale marriage matching function, where the marriage rate should have increased over the decade, will provide a even worse fit.

In order to quantify the changes in the systematic gains to marriage over the decade, we first estimate a benchmark model of the marriage market where the type space only consists of the ages of individuals. That is, $i = 16, .., 75$ denotes the age for men, $j = 16, ..75$ denotes the age for women, $t = 71$ and $t = 81$ denotes the period 1971/72 and 1981/82 respectively. For year t, we estimate the total systematic gains to marriage by:

$$\pi_{ij}^t = \ln \left[\frac{\mu_{ij}^t}{\sqrt{\mu_{i0}^t \mu_{0j}^t}} \right]$$

To deal with the thin cell problem, we employ a non-parametric locally smoothed estimate of $\pi^t(\cdot)$ at any age pair (k, l).[29]

---

[29]

$$\widehat{\pi}^t(k,l) = \sum_{j=1}^{J} \sum_{i=1}^{I} \omega_{ij}(k,l) \cdot \pi_{ij}^t \tag{25}$$

$$\text{where} \qquad \omega_{ij}(k,l) = \frac{(IJb^2)^{-1} K(\frac{i-k}{b}) \cdot K(\frac{j-l}{b})}{(IJb^2)^{-1} \sum_{i=1}^{I} \sum_{j=1}^{J} K(\frac{i-k}{b}) \cdot K(\frac{j-l}{b})}$$

The parameter b denotes the bandwidth and $K(\cdot)$ denotes the kernel weight. Like in the previous application, we used the normal kernel and attempted numerous bandwidth parameters.

Figure 5 shows the smoothed non-parametric plot of $\widehat{\pi}_{ij}^{71}$. Compared with Figure 1a, the distribution of the estimated total gains are less peaked and less concentrated. In particular, the total gains are larger off the age diagonal and for older individuals than would be predicted from bivariate marriage distribution of Figure 1a. Like the estimates of the net gains from marriage in the previous section, we observe that the estimated systematic total gains relative to remaining single is also negative. This reflect the empirical fact that most available individuals do not marry. As far as we know, we have just presented the first estimates of the systematic gains to marriage relative to not marrying between any two age pairs for the US.

As discussed earlier, Figure 1a shows the standard result that there is strong positive assortative matching by age. Beginning with Becker, economists have investigated conditions for observing positive assortative matching by traits in the marriage market.[30] These investigations assume that if a trait of an individual has a natural ordering, then marital output is increasing or decreasing in that trait.[31] Figure 5 shows that systematic marital output is not monotonic in either male or female ages. Thus the theoretical investigations into the conditions for positive assortative matching do not apply in our context. The rationale for observing positive assortative matching by age in Figure 5 is more mundane. Approximately along the age diagonal, systematic marital output is higher than off the age diagonal. That is, individuals match assortatively by age because they 'suit' each other best! Although not shown, the plot of $\widehat{\pi}_{ij}^{81}$ is qualitatively similar.

7.3 Testing model mispecification

A strong implication of our model, as given in equation (15), is that π^t only reflect preference parameters and is independent of population vectors. To the extent that dynamic considerations and scale effects are important, π^t will be a function of the population vectors at time t. Table 2 presents some regressions of $\widehat{\pi}_{ij}^{71}$ on demographics and 1970 population vectors. We also include results using $\ln \widehat{\theta}_{ij}^{71}$ from equation (23) as a dependent variable to test Dagvik's model as well.

[30] Also see Shimer and Smith 2000, Legros and Newman 2002.

[31] Most of these investigations implicitly or explicitly focus on income of the participants. Holding other factors constant, the assumption that a spouse with more income is better is natural.

TABLE 2

	OLS	IV	OLS	IV
Dependent var.	$\ln\left(\frac{\mu_{ij}}{\sqrt{\mu_{i0}\mu_{0j}}}\right)$		$\ln\left(\frac{\mu_{ij}}{\mu_{i0}\mu_{0j}}\right)$	
$\ln(m_i^{70})$	0.0321	0.2136	-0.6156	-0.1484
	(0.12)	(0.09)	(2.27)	(0.06)
$\ln(f_j^{70})$	-0.7038	-6.7777	-1.3368	-6.6350
	(2.81)	(1.54)	(5.32)	(1.52)
Observations	2771	2771	2771	2771
R^2	0.93		0.93	
Instruments		$\ln(m_i^{80})$		$\ln(m_i^{80})$
		$\ln(f_j^{80})$		$\ln(f_j^{80})$

Robust t statistics in parentheses. Ninth order polynomial present in all regressions.

We present results from OLS and instrumental variable (IV) regressions.[32] IV regressions were carried out because $\widehat{\pi}_{ij}^{71}$ and $\ln\widehat{\theta}_{ij}^{71}$ are constructed with 1970 population vectors. So if there is measurement error in our measure of the population vectors, this may induce correlations between $\widehat{\pi}_{ij}^{71}$, $\ln\widehat{\theta}_{ij}^{71}$ and the population vectors even when there is no true relationship. We use the 1980 population vectors as instruments for the 1970 population vectors. In all regressions, we also include ninth order age polynomials of ages, i and j.

Columns (1) and (2) show OLS and IV results for our model. Columns (3) and (4) show OLS and IV results for Dagsvik's model. In column (1), the 1970 female population vector can still explain variations in $\widehat{\pi}_{ij}^{71}$ even after controlling for demographics (using a ninth order polynomial in ages). This result suggests that $\widehat{\pi}_{ij}^{71}$ is possibly correlated with the population vectors in 1970, a violation of our model. The estimated negative coefficient however suggests that this correlation could also be induced by measurement error in population vectors. In the IV regression (column (2)), the correlation between

[32]We also estimated models where the observations are weighted by the geometric average of the population vectors, and with median regressions. The results are similar to those obtained here.

our estimate of preferences and population vectors become statistically insignificant. However, given the large standard errors, it is premature to conclude that there is no model mispecification.

The results for Dagsvik's model in columns (3) and (4) are similar to what we found for our model. In fact in the OLS results (column (3)), the estimated coefficients for both population vectors are statistically different from zero. Thus there is some marginal evidence to prefer our model to his. The evidence in Table 2 does not support the hypothesis that scale effects are quantitatively important in the US marriage market in the early seventies.

8 Drop in the gains to marriage

Figure 6 shows the plot of the change in the gains to marriage over the decade, $\Delta\widehat{\pi}_{ij} = \widehat{\pi}_{ij}^{81} - \widehat{\pi}_{ij}^{71}$. As alluded to earlier, the striking feature of the data is the sharp drop in the estimated total gains to marriage to young adults in 81/2. The drop is particular visible along the age diagonal as shown in Figure 6b.

We explored two factors which may have caused the fall in the total gains to marriage. First, in an earlier draft of this paper, we expanded the type space to include three levels of education (less than high school, high school graduate and college graduate). College graduates delay marriage relative to non-college graduates. We expected more adults to have obtained college degrees in 1980. So if we account for educational attainment, we may be able to explain part of the drop in the gains to marriage. Introducing educational attainment in an expanded type space significantly improved the fit of the model. Nonetheless, we were still unable to account for most of the estimated drop in the gains to marriage among young adults over the decade. The reason why including educational attainment did not explain the drop is shown in Figure 7. While the fraction of adults who have a college degree in 1980 was much larger than in 1970, the increase was concentrated among older adults. Among adults younger than 25 years old, there was essentially no difference in college attainment between 1970 and 1980. Thus educational attainment did not play a large role in affecting the drop in the gains to marriage among young adults over the seventies.

9 Legalizing abortion and the fall in the gains to marriage

Technological innovations and social changes like the invention of the pill and the legalization of abortion in the seventies has affected the gains to marriage by changing the opportunites available to women. In this section, we explore the role of differential access to legal abortions across states in the seventies in affecting the gains to marriage.[33] Before 1967, legal abortion was generally unavailable. Between 1967 and 1973, legal abortion became easier to obtain in several states (reform states).[34] The reform states included in our analysis are: Alaska, California, Delware, Florida, Georgia, Hawaii, Kansas, Maryland, North Carolina, Oregon, South Carolina and Virginia.

In January 22, 1973, due to the United States Supreme Court ruling in *Roe v. Wade,* legal abortions became available in the entire country. This ruling is less restrictive on access to abortion than what were available previously in the reform states.

If partial availability of legal abortions in a state reduced the gains to marriage in that state, we would expect to see lower gains to marriage in reform states relative to non-reform states in 1971/2 but not in 1981/2. Moreover this difference in difference in the gains to marriage should be concentrated among women of child bearing age and the men who marry them.

In order to empirically study the impact of the partial legalization of abortions on the gains to marriage, consider an expansion of the type space of individuals. A type of an individual is now defined by his or her age, whether the individual lives in a reform state (r for male and R for female), or non-reform state (n for male and N for female), and time, t. We will use the convention s and S to denote the states of residences for a male and female respectively, where $s \in \{r, n\}$ and $S \in \{R, N\}$. We assume that all individuals at time t, living in a reform state or otherwise, are available to other individuals at the same time t in one national marriage market. So an individual living in a reform state at time t may marry someone living at same time in either a reform

[33] Akerlof, et. al. 1996 argued that the legalization of abortion may substantially reduce the gains to marriage. Also see Goldin and Katz 2002 and Siow 2002.

[34] Thirteen states passed "Model Penal Code" legislation. Alaska, Florida, Hawaii, New York and Washington enacted even more liberal laws. California's restrictive abortion laws were struck down by the state courts. See Merz, Jackson and Klerman 1995 for details.

or non reform state, and vice versa.

Let $t = 71$ refer to the marriage market in the years 1971 and 1972, and $t = 81$ refer to 1981 and 1982. The number of i, j marriages between male and female individuals from states (s, S) respectively at time t is denoted by μ_{ijt}^{sS}.

To provide a benchmark for our analysis, consider the marriage rate regression, where ρ_{jt}^{S} is the marriage rate of age j females living in state S at time t:

$$\rho_{jt}^{S} = h(j) + h_t(j) \cdot (1 - D_{jt}) + h^R(j) \cdot D_j^r + h_t^R(j) \cdot D_j^R \cdot D_{jt} + v_{jt}^S \tag{26}$$

We use the notation D to denote dummy variables and $h(x)$ to denote some general nonparametric function which has x as its argument. The variable D_{jt} takes a value of 1 for $t = 1971/72$ and zero otherwise; D_j^R takes a value of 1 if the female individual is from a reform state, and zero otherwise; v_{jt}^S is an error term with mean zero.

The terms $h_t(j)$ and $h^R(j)$ allow for age specific time trend and age specific state effect respectively. Then $h_t^R(j)$ measures the impact of living in a reform state at $t = 71$ on the marriage rate of type j females. The function $h_t^R(j)$ can be estimated non-parametrically by the difference in differences (DD) estimator. Let

$$\Delta^2 \rho_j^f = [\rho_{j71}^R - \rho_{j81}^R] - [\rho_{j71}^N - \rho_{j81}^N], \tag{27}$$

Then a non-parametric estimator of $h_t^R(k)$ at age k is given by,

$$\widehat{h_t^R}(k) = \sum_{j=1}^{J} \omega_j(k) \cdot \Delta^2 \rho_j^f \tag{28}$$

$$\text{where } \omega_j(k) = \frac{(Jb)^{-1} K(\frac{j-k}{b})}{(Jb)^{-1} \sum_{j=1}^{J} K(\frac{j-k}{b})}$$

The function $K(\cdot)$ is the kernel weights and b is the appropriate bandwith for the estimator. A similar estimator can be constructed for age i males.

The systematic gains to marriage can be parameterized in a similar manner. Let the marriage gains to an age i male living in state s with an age j female living in state S at time t, π_{ijt}^{sS}, be given by:

$$\begin{aligned}\pi_{ijt}^{sS} = {} & g(i,j) + g_t(i,j) \cdot (1 - D_{ijt}) + g^{rR}(i,j) \cdot D_{ij}^{rR} + g^{nR}(i,j) \cdot D_{ij}^{nR} \\ & + g^{rN}(i,j) \cdot D_{ij}^{rN} + g_t^{rR}(i,j) \cdot D_{ij}^{rR} \cdot D_{ijt} + g_t^{nR}(i,j) \cdot D_{ij}^{nR} \cdot D_{ijt} \\ & + g_t^{rN}(i,j) \cdot D_{ij}^{rN} \cdot D_{ijt} + \varepsilon_{ijt}^{sS}\end{aligned} \tag{29}$$

The notational convention adopted in equation [26] applies. The dummy variable D_{ijt} takes a value of 1 for age combinations in years $t = 1971/72$ and zero otherwise; the variable D_{ij}^{rN} takes a value of 1 for couples where the male resides in the reform states, r, and the females in the non-reform state, N and zeros otherwise, and so on. The function $g(i, j)$ captures the systematic gain to marriage for an age i male in a non-reform state with an age j female in a non-reform state in 1971/72. It forms the base gains to marriage that varies according to the ages of the couples, (i, j). The functions $g^{rR}(i, j)$, $g^{nR}(i, j)$ and $g^{rN}(i, j)$, captures the remaining fixed effects arising from the state of residence of the couple. For example, $g^{rR}(i, j)$ is the increment in systematic gains added to the base $g(i, j)$ if the couples are both from the reform states.

The increment to the gains to marriage in years 1981/82 for an (i, j) pair is captured by the function $g_t(i, j)$. This time effect is assumed to be independent of the state of residence. The function $g_t^{sS}(i, j)$ is the increment to the gains to marriage in $t = 1971/72$ between a male in state s and female in state S for state combinations $sS \neq nN$. If we expect the legalization of abortion in the reform states to have lowered the gains to marriages among young adults who both reside in those states, then $g_t^{rR}(i, j) < 0$ for young couples. The mean zero error term is denoted by ε_{ijt}^{sS}.

Our model for the systematic gains to marriage in equation [29] has some advantages over the marriage rate formulation in equation [26]. First, the formulation using the systematic gains satisfies all the restrictions of a marriage matching function while the marriage rate models of the form in equation [26] do not impose any restriction between different marriage rates. Second, our model can distinguish between the effect of the legalization of abortion on the systematic gains to marriage for age i males with different types of females. For example, $g_t^{rR}(i, j)$ need not be the same as $g_t^{rR}(i, j')$.

For any age combination (i, j) with observed marriages, the systematic gains, π_{ijt}^{sS} is estimated by $\widehat{\pi_{ijt}^{sS}} = \ln\left(\mu_{ijt}^{sS} / \sqrt{\mu_{i0t}^{sS}\mu_{0jt}^{sS}}\right)$. The increment in the gain to marriage for an i, j pair in 1971/72 who lived in reform states, $g_t^{rR}(i, j)$ can be estimated by the DD estimator:

$$\Delta^2\pi_{ij}^{rR} = (\widehat{\pi_{ij71}^{rR}} - \widehat{\pi_{ij81}^{rR}}) - (\widehat{\pi_{ij71}^{nN}} - \widehat{\pi_{ij81}^{nN}}) \tag{30}$$

Similarly, we can define a non-parametric estimate of $g_{71}^{rR}(\cdot, \cdot)$ at any age combination

(k,l) as

$$\widehat{g_{71}^{rR}}(k,l) = \sum_{j=1}^{J}\sum_{i=1}^{I} \omega_{ij}(k,l) \cdot \Delta^2 \pi_{ij}^{rR} \tag{31}$$

$$\text{where} \qquad \omega_{ij}(k,l) = \frac{(IJb^2)^{-1} K(\frac{i-k}{b}) \cdot K(\frac{j-l}{b})}{(IJb^2)^{-1} \sum_{i=1}^{I} \sum_{j=1}^{J} K(\frac{i-k}{b}) \cdot K(\frac{j-l}{b})}$$

Note the similarity between $\Delta^2 \pi_{ij}^{sS}$ and the standard DD marriage rate estimator, $\Delta^2 \rho_l^f$, for $l = i, j$. Although $\Delta^2 \pi_{ij}^{sS}$ is defined for an age pair (i,j), and state pair (s,S) rather than for male or female ages alone, it is as easy to estimate as equation [27].

Data:

Using information on the place of residence from the *US Census,* and the marriage records from the *Vital Statistics,* we classify the data described in Section 7 according to whether the place of residence of an individual is a reform or non-reform state. Table 3 provides a summary of the data used.

The sample of available males and females on the marriage market from the non-reform states is considerably larger than that of the reform states. The increase in the population observed over the decade in the two groups of states also differ in magnitude. In the reform states, the population of available males and females increased by 50.4 % and 54.6 % respectively, compared to a more modest increase of 38.2 % and 30.7 % for available males and females respectively in the non-reform states. The average age of males and females in the two groups of states are comparable to the numbers of the entire sample reported in Table 1.

As expected, marriages between individuals in the same state of residence are more likely relative to marriage between individuals living in different states. There are 2.1 million marriages between couples in the non-reform states, (μ^{nN}), compared to 1.05 million between couples in the reform states, (μ^{rR}), in 1971/72. The number of cross-marriages in 1971/72, $(\mu^{rN},\ \mu^{nR})$, is around 40,000. The changes in the total number of marriages across the four groups over the decade differ in magnitude and sign. Marriages between reform state males and non-reform state females decreased by 14.8 % while marriages between males from non-reform states and females from the reform states decreased by almost 30 %. In the reform states where there was little change in access to legalized abortion over the decade, we find total marriages increase by 17 %

Table 3: Data summary based on place of residence

	US Census data in 1970	1980	Δ
No. of available males in reform states (M^r)	5.76 mil	9.24 mil	60.41%
No. of available females in reform states (F^r)	6.70 mil	10.36 mil	54.63%
No. of available males in non-reform states (M^n)	10.25 mil	14.17 mil	38.24%
No. of available females in non-reform states (F^n)	12.90 mil	16.86 mil	30.70%
Aver. age of available males in reform states	30.00	29.62	
Aver. age of available females in reform states	38.93	36.93	
Aver. age of available males in non-reform states	30.64	29.53	
Aver. age of available females in non-reform states	39.22	37.24	

	Vital Statistics data in 1971/72	1981/82	Δ
No. of marriages in rR states (μ^{rR})	1.05 mil	1.26 mil	17.17%
No. of marriages in rN states (μ^{rN})	45,456	38,730	-14.80%
No. of marriages in nR states (μ^{nR})	39,367	30,358	-29.68%
No. of marriages in nN states (μ^{nN})	2.10 mil	2.11 mil	0.56%
Aver. age married males in reform states	27.5	29.6	
Aver. age married males in non-reform states	26.9	28.9	
Aver. age married females in reform states	24.8	26.8	
Aver. age married females in non-reform states	24.4	26.2	

while in the non-reform states where legalized abortion became more accessible, total marriages only increased by .56 %. It is this differential change in marriage patterns in the four groups and the changes in the population of marriage market participants that provide identification of the fall in marriage gains due to legalizing abortion.

Results:

Figure 8(a) shows estimates of the decrease in marriage rates in the reform states from the DD marriage rate estimators, $\widehat{h_t^N}(k)$. Consistent with the findings in Angrist and Evans 1999, $\widehat{h_t^l}(k)$ are negative for both young males and young females. There is evidence of a small increase in the marriage rate of males, and a smaller increase in the marriage rates of females, between the ages of 30 to 40. As explained later, it is problematic that the estimated effects for males are significantly larger than that for females.

Figure 9 shows estimates of $\widehat{g_t^{rR}(i,j)}$. The figure shows the systematic gains to marriage for young adults in 1971/72 living in reformed states fell relative to those living in non-reformed states. Other effects are less easy to detect from the figure. Figure 8(b) shows a slice along the diagonal of the $\widehat{g_t^{rR}(\cdot,\cdot)}$ distribution, that is, $\widehat{g_t^{rR}(i,i)}$ for same aged spouses. This slice of the distribution is informative because there are many same aged spouses. The drop in the gains to marriage for same age spouses, between the ages of 19 to 26, in reform states is substantial. We also see a small increase in the gains to marriage for same age spouses, between the ages of 27 to 40. An explanation of these gains is that these are young individuals who would have gotten married young had abortion not been legalized. This social change allow these individuals to delay marriage to an older age.

We interpret the effects displayed in Figures 8 and 9 as due to the partial legalization of abortion on marriage rates and the gains to marrage. The standard DD argument for identification is based on the claim that the policy intervention of interest generates year and location specific interaction effects that would otherwise not be there. In addition to the standard argument, we also expect partial legalization to affect young adults more than older adults which is consistent with the evidence in Figures 8 and 9.

In order to quantify the effect of the partial legalization of abortion on marriage rates, we use the two estimators, $\Delta^2\pi_{ij}^{sS}$ and $\Delta^2\rho_l^k$ $(k = m, f)$, to do a counterfactual experiment. Consider an experiment where the non-reform states also partially legalize abortion in 1971/72 like the reform states. The estimates from the DD marriage rate equation [27], allow us to contruct a counterfactual marriage rate for male and female in the non-reform states. Using our estimate of $\widehat{h^s_{71}}(k)$, let the counterfactual marriage rates for males and females be denoted by $\widetilde{\rho^n_{i71}}$ and $\widetilde{\rho^N_{j71}}$ respectively, where

$$\widetilde{\rho^s_{k71}} = \rho^s_{k71} - \widehat{h^s_{71}}(k) \quad \text{where } (s,k) \in \{(n,i),(N,j)\}.$$

Using the counterfactual marriage rates in the non-reform states and the observed rates in the reform states, we construct an aggregate male and female marriage rate in the scenario where there was no differential access to abortion in 1971/72.

A comparable counterfactual marriage rate can be constructed using the DD marriage gains estimator. We first estimate $\widehat{g^{rR}_{71}}(i,j)$, $\widehat{g^{rN}_{71}}(i,j)$, and $\widehat{g^{nR}_{71}}(i,j)$ according to equation [30]. Using these estimates, we construct gains to marriage in the non-reform states in the counterfactual scenario that abortion was partially legalized in these states in 1971/72. Using our estimates of $\widehat{g^{sS}_{71}(i,j)}$, these counterfactual marriage gains, $\widetilde{\pi^{nN}_{ij71}}$, $\widetilde{\pi^{rN}_{ij71}}$, and $\widetilde{\pi^{nR}_{ij71}}$ are estimated according to this equation,

$$\widetilde{\pi^{sS}_{ij71}} = \widehat{\pi^{sS}_{ij71}} - \widehat{g^{sS}_{71}(i,j)} \qquad \text{where} \quad sS \in \{nN, rN, Rn\}, \quad \forall\ i,j$$

We subsequently compute the number of marriages that would have been observed using these counterfactual marriage gains and the observed marriage gains for the reform states, $\widehat{\pi^{rR}_{ij71}}$.

Let the counterfactual aggregate marriage rates in 1971/72 constructed using the DD marriage rate and DD marriage gains estimator be denoted by C^{ρ}_{j71} and C^{π}_{j71} respectively. The graphs in Figure 10 (a) and (b) compares the change in actual marriage rates for age k, $\Delta\rho^l_k = \rho^l_{k81} - \rho^l_{k71}$, $l = m, f$, with $\Delta^{\rho}\rho^l_k = C^l_{k71} - \rho^l_{k71}$ and $\Delta^{\pi}\rho^l_k = C^{\pi}_{k71} - \rho^l_{k71}$.[35] As discussed earlier, marriage rates for males and females fell over the decade. Both $\Delta^{\rho}\rho^l_k$ and $\Delta^{\pi}\rho^l_k$ suggest that a quantitatively significant part of the fall in aggregate marriage rates for young adults over the decade is attributable to the lack of partial legalization in the non-reform states in 1970. $\Delta^{\pi}\rho^m_{22}$ suggests that 20% of the observed fall in the 22

[35] Non-reform states went from no legalization to full legalization between 1970 and 1980. This change can be conceptually decomposed into (1) no legalization to partial legalization, and (2) partial legalization to full legalization. We are asking how much of the change in marriage rates over the decade can be attributed to the conceptual change from no legalization to partial legalization.

year old male marriage rates can be attributed to partial legalization compared to the 31% estimate from $\Delta^{\rho}\rho_{22}^{m}$. The estimates of female marriage rate decrease attributable to the partial legalization of abortion is more modest.

While the $\Delta^{\rho}\rho_{k}^{l}$ and $\Delta^{\pi}\rho_{k}^{l}$ estimators provide qualitatively similar results, the quantitative predictions of the two estimators are very different. As mentioned earlier, a shortcoming of the DD marriage rate estimator is the lack of consistency between the estimated number of marriages from the male and female marriage rate equations. The estimate from the female DD marriage rates estimator suggest that legalizing abortion in the non-reform states would have resulted in 7080 less marriages in 1971/72 while the estimate using the male marriage rates is 196,270.[36] The latter estimate is larger by a factor of 27 times![37] This kind of discrepancy from male and female marriage rate regressions is not unusual. So while marriage regressions are easy to use and interpret, the biases in these estimators can be substantial.

The estimate from the DD marriage gains estimator is around 45,440 less marriages among individuals aged 16 to 75 years of age. In other words, the legalization of abortion in the non-reform states in 1971/72 would have resulted in 1.4 % less marriages in that period. Among young individuals the decrease is more pronounced. For males aged 16 to 25 years old, abortion legalization in the non-reform states would have lowered the number of marriages in this group by 4.2 % while among 16 to 25 years old females, the decrease is around 3.6 %. As suggested by the graphs in Figure 10, the effects is reversed for older aged individuals. For males older than 26 years of age, this social change would have increased the number of marriages in this group by 3.8 % and for females older than 26 years of age, the increase is around 5.2 %.

10 Conclusion

This paper proposed and estimated a non-parametric transferable utility model of the marriage market. The model was used to estimate US marital behavior in 1971/72 and 1981/82. The estimates show that the gains to marriage for young adults fell

[36]The total number of recorded marriages in 1971/72 is 3,235,806.

[37]While smaller, significant disparity remains if we limit ourselves to marriages for individuals less than age thirty.

substantially over the decade. The legalization of abortion had a significant quantitative impact on the fall in the gains to marriage.

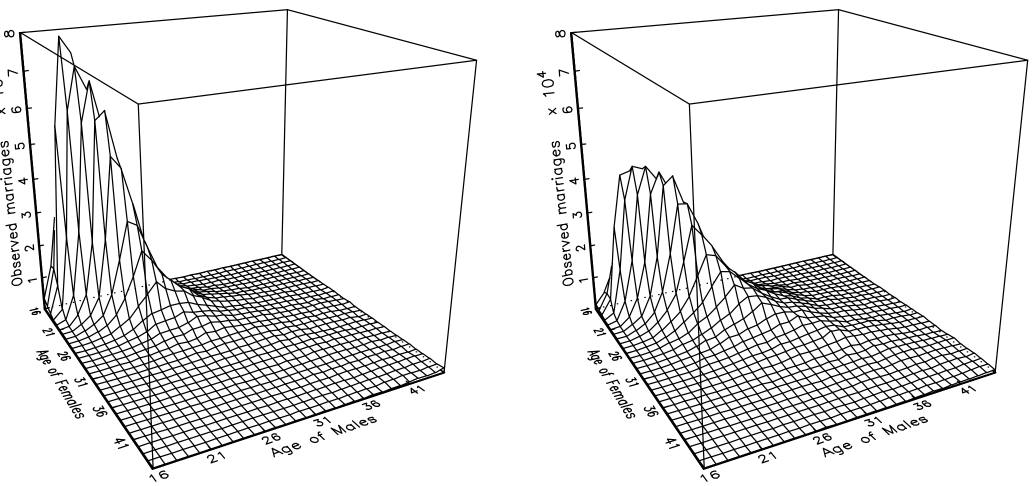
We discuss some avenues for future research. First, we need to understand better the substitution effects in this model.

Second, we considered a static model of the marriage market. A dynamic transferable utility model of the marriage market is needed. This research is currently in progress.

Third, an issue that arises in modelling dynamic marriage market models with search frictions is the specification of the meeting technology. To date, economists have primarily considered random meeting technology (E.g. Ayagari, et. al. 2000, Seitz, Hamilton and Siow, Wong 2003a and 2003b). Our marriage matching function provides a methodology for generating non-random meeting technologies.

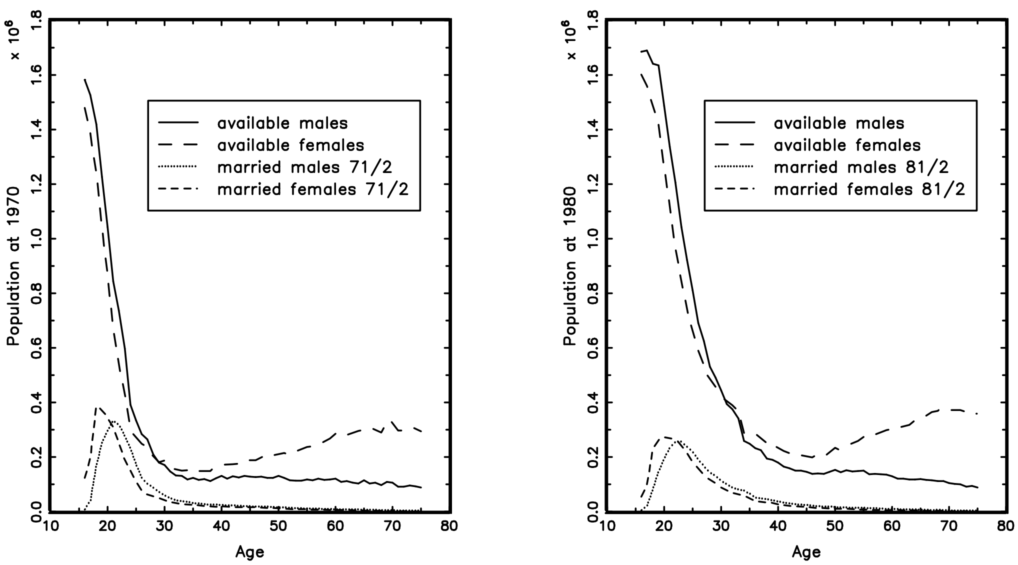
Fourth, we ignored cohabitation in this paper. Methodologically, cohabitation is easy to incorporate. Cohabitation is another type of match in the marriage market. The reason why we ignored cohabitation in the current empirical analysis is because we cannot calculate the number of new entrants into cohabitation from census data.

Figure 1



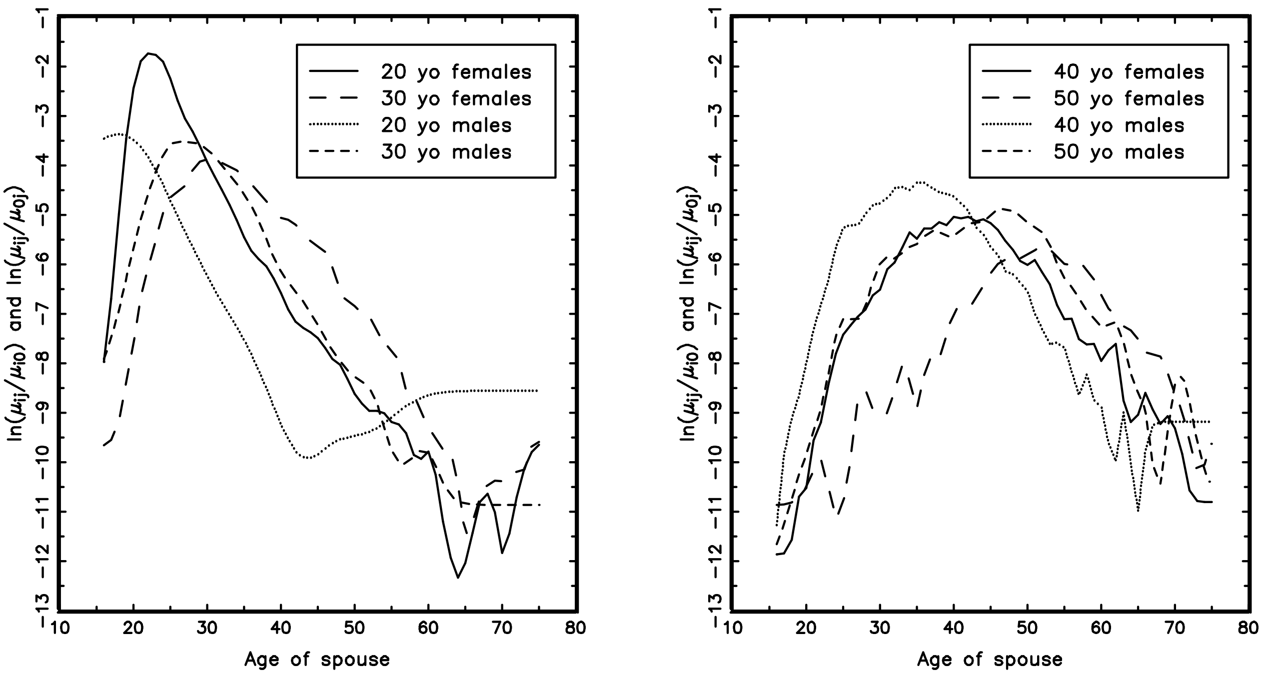
a) Surface of observed μ_{ij} for 1971/72 b) Surface of observed μ_{ij} for 1981/82

Figure 2



a) Availables and Marrieds at 1970 b) Availables and Marrieds at 1980

Figure 3



a) Systematic net returns for 20 and 30 years old

b) Systematic net returns for 40 and 50 years old

Figure 4

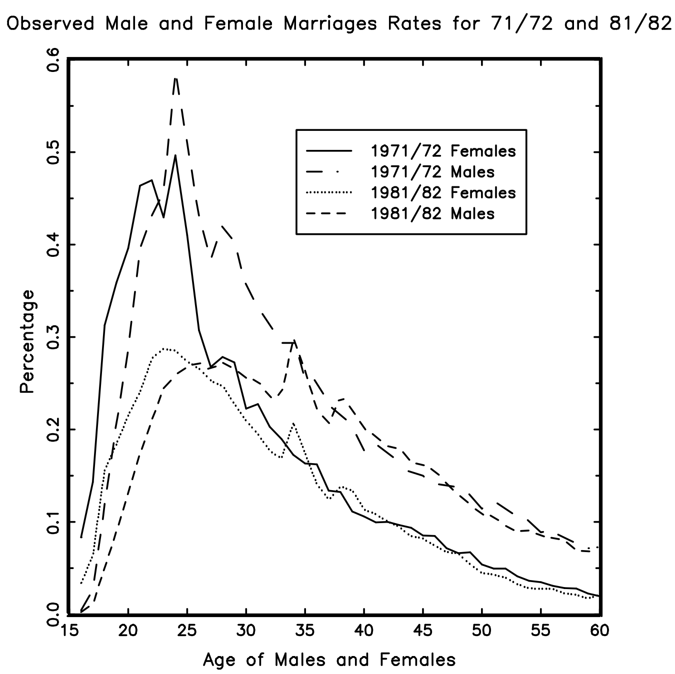


Figure 5

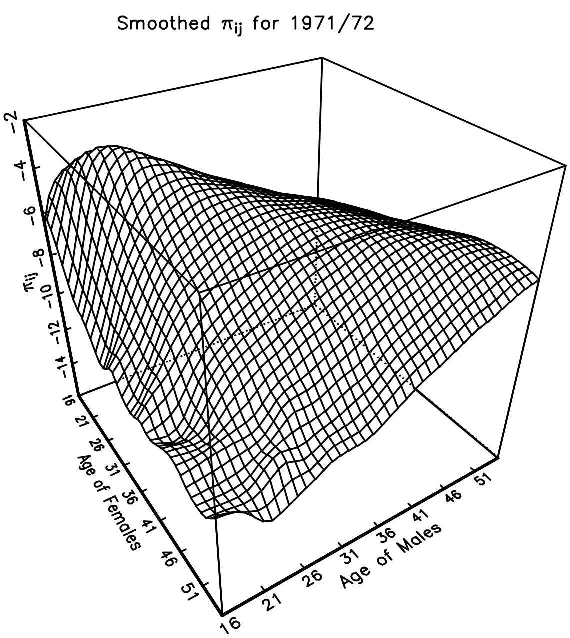
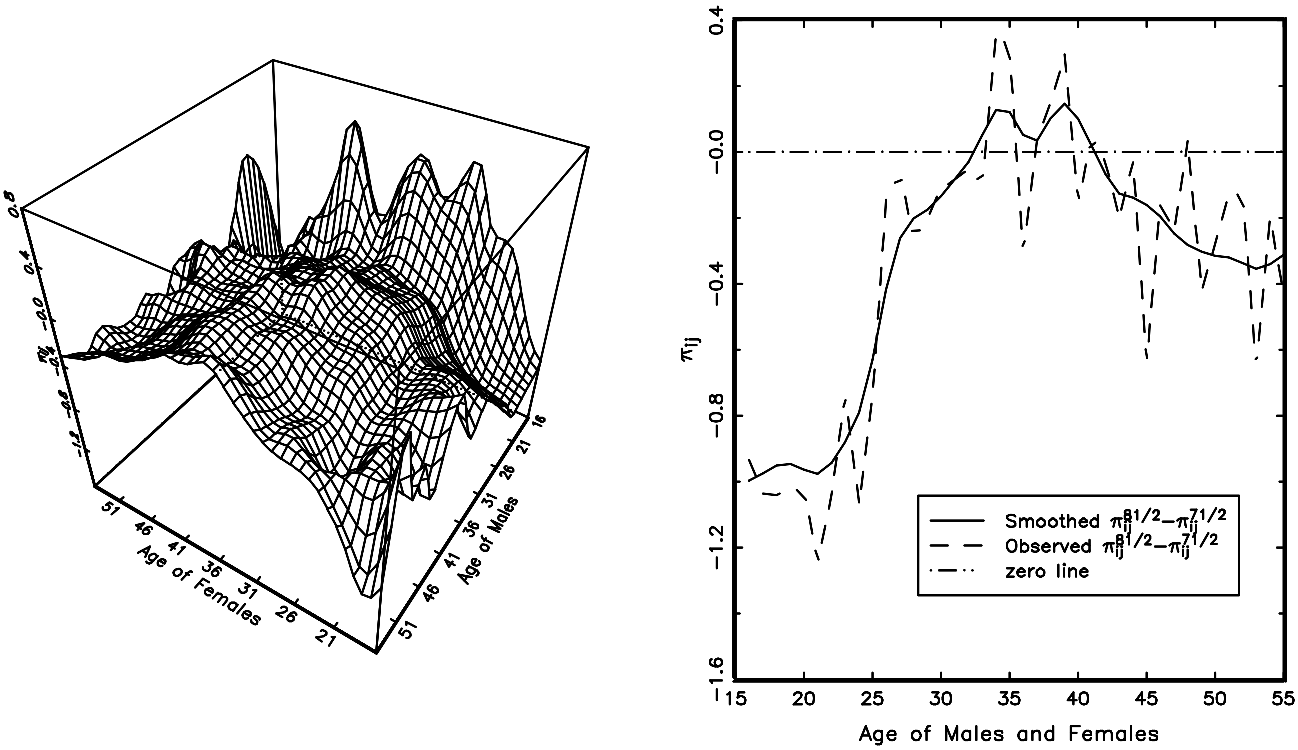


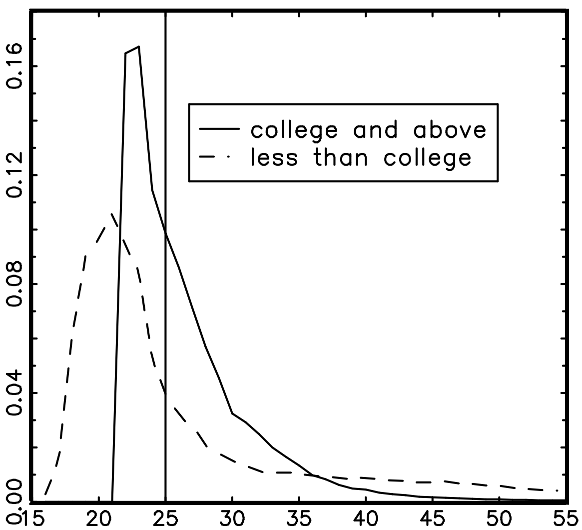
Figure 6



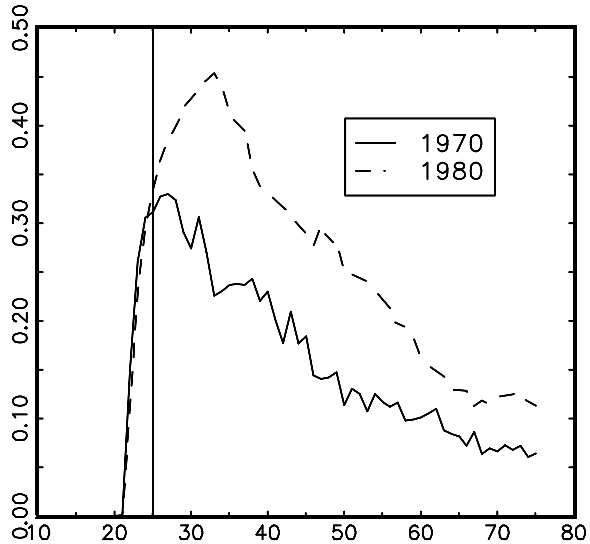
a) Differenced in aggregate π, $\pi_{ij}^{81/2}-\pi_{ij}^{71/2}$

b) Differenced in aggregate π, i.e.$(\pi_{ij}^{81/2}-\pi_{ij}^{71/2})$ among same aged married individuals

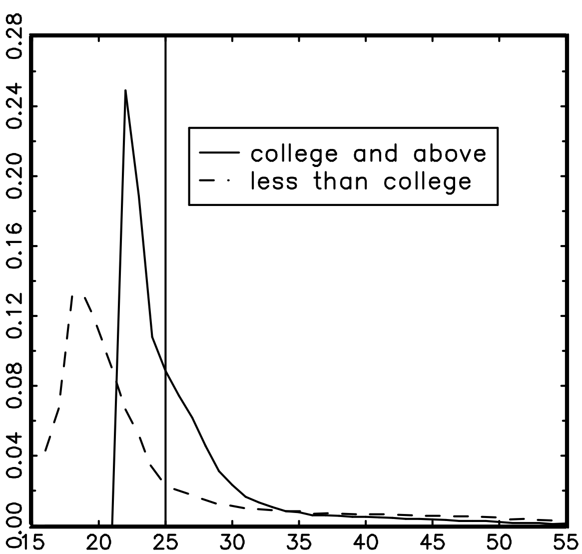
Figure 7



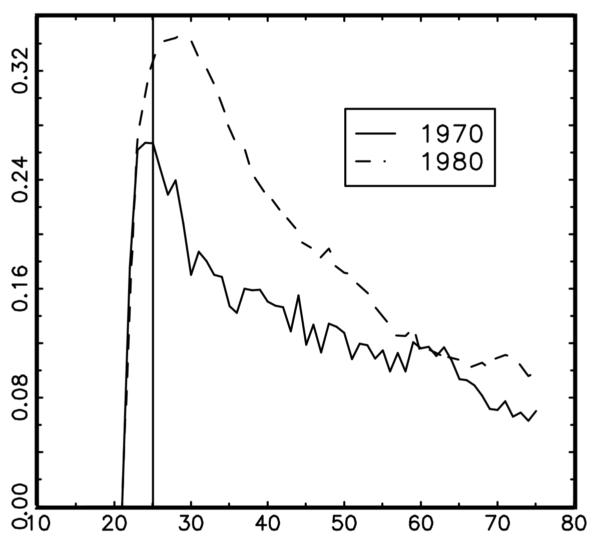
a) Histogram of 1970 Married Males
by age and education



b) Proportion of Males with
college and above education



c) Histogram of 1970 Married Females
by age and education



d) Proportion of Females with
college and above education

FIGURE 8

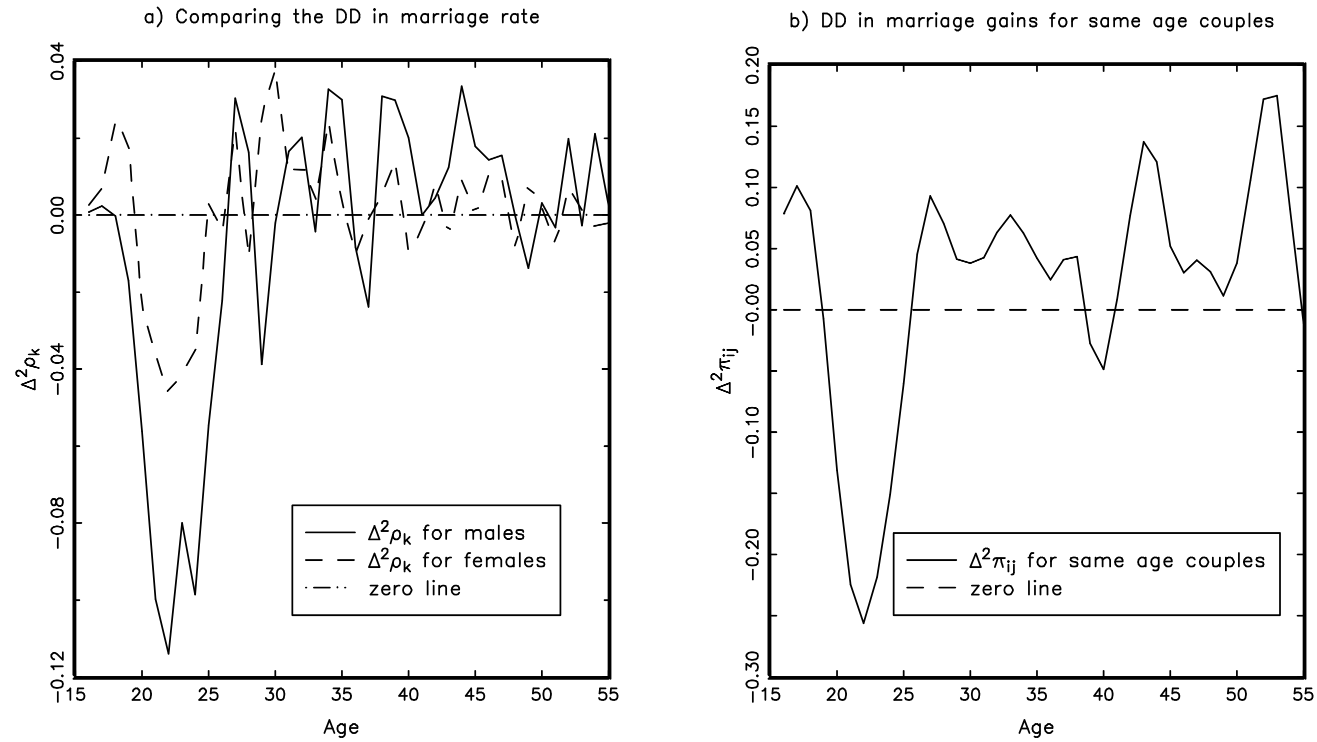


FIGURE 9

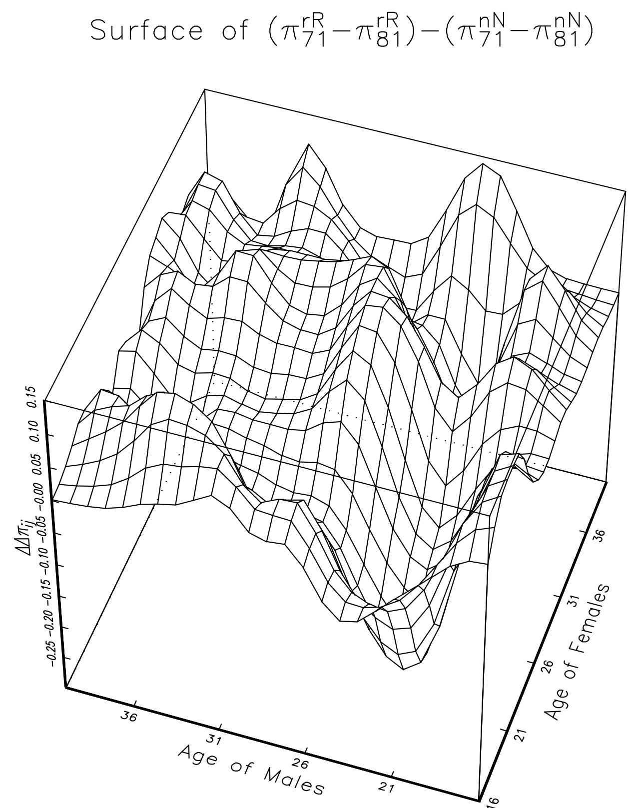
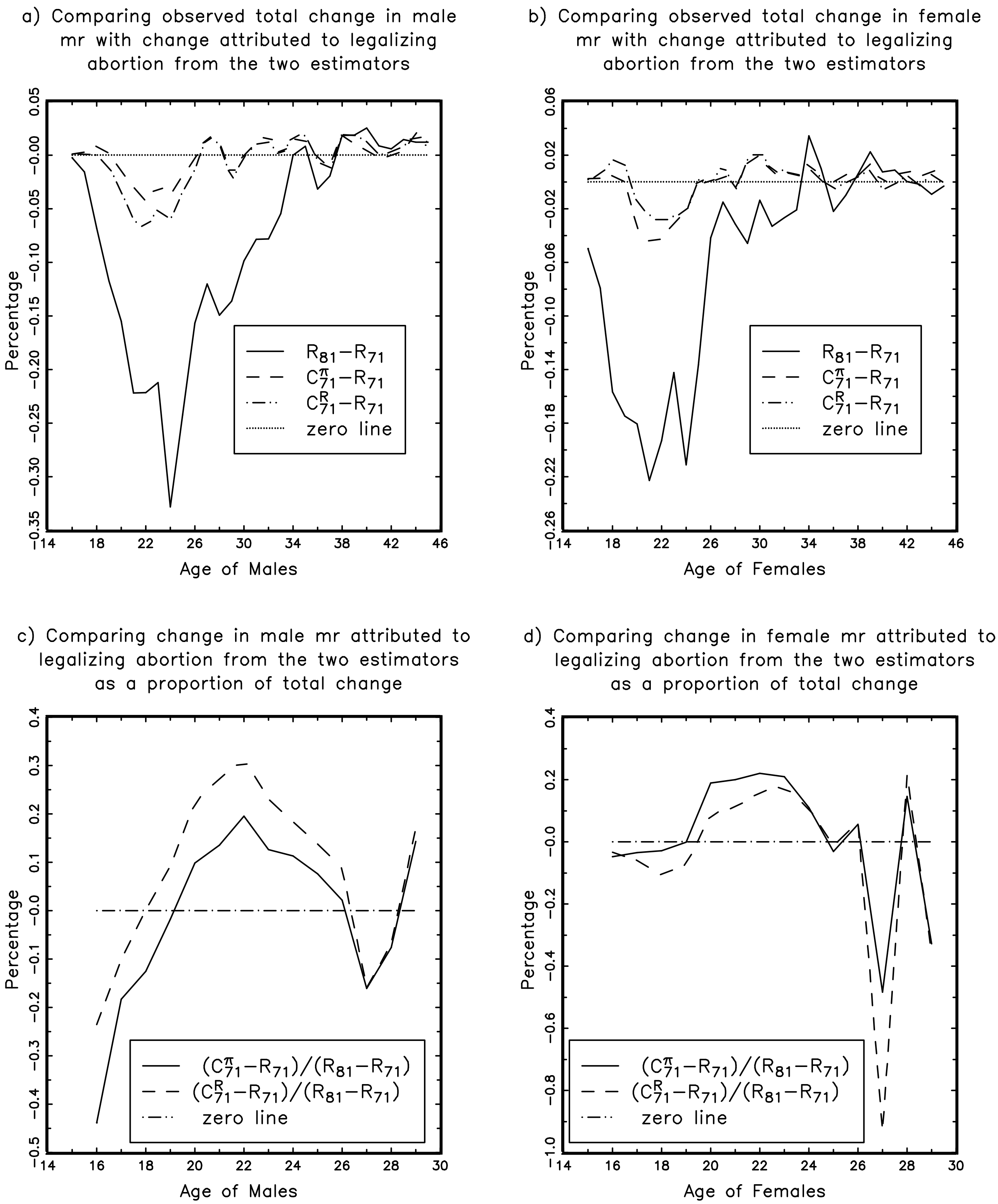


FIGURE 10



a) Comparing observed total change in male mr with change attributed to legalizing abortion from the two estimators
Percentage
Age of Males
$R_{81}-R_{71}$
$C^{\pi}_{71}-R_{71}$
$C^{R}_{71}-R_{71}$
zero line
b) Comparing observed total change in female mr with change attributed to legalizing abortion from the two estimators
Percentage
Age of Females
$R_{81}-R_{71}$
$C^{\pi}_{71}-R_{71}$
$C^{R}_{71}-R_{71}$
zero line
c) Comparing change in male mr attributed to legalizing abortion from the two estimators as a proportion of total change
Percentage
Age of Males
$(C^{\pi}_{71}-R_{71})/(R_{81}-R_{71})$
$(C^{R}_{71}-R_{71})/(R_{81}-R_{71})$
zero line
d) Comparing change in female mr attributed to legalizing abortion from the two estimators as a proportion of total change
Percentage
Age of Females
$(C^{\pi}_{71}-R_{71})/(R_{81}-R_{71})$
$(C^{R}_{71}-R_{71})/(R_{81}-R_{71})$
zero line

11 Appendix A

The derivations of (8) and (9) are known and included here for completeness.

11.1 Derivation of (8)

(5) may be rewritten as

$$V_{ijg} = \widetilde{\alpha}_{ij} - \tau_{ij} + \varepsilon_{ijg} = \eta_{ij} + \varepsilon_{ijg}$$

As specified by (7), g solves

$$V_{ig} = \max_j \{V_{i0g}, .., V_{ijg}, .., V_{iJg}\}$$

The probability that a type j woman is chosen is:

$$\begin{aligned} \Pr\{V_{ig} = V_{ijg}|\eta\} &= E\{\Pi_{k \neq j} F(\varepsilon_{ijg} + \eta_{ij} - \eta_{ik})\} \\ &= \int_{-\infty}^{\infty} \exp\{-\sum_{k \neq j} e^{-\varepsilon - \eta_{ij} + \eta_{ik}}\} e^{-\varepsilon - e^{-\varepsilon}} d\varepsilon \end{aligned}$$

The index k runs from 0 to J.

Let

$$c = 1 + \sum_{k \neq j} e^{-\eta_{ij} + \eta_{ik}}$$

Also note that

$$\int e^{-\varepsilon - ce^{-\varepsilon}} d\varepsilon = \frac{e^{-ce^{-\varepsilon}}}{c} \tag{32}$$

Then (**??**) becomes:

$$\begin{aligned} \Pr\{V_{ig} = V_{ijg}|\eta\} &= \int_{-\infty}^{\infty} \exp\{-\varepsilon - c\exp(-\varepsilon)\} d\varepsilon \\ &= \left[\frac{e^{-ce^{-\varepsilon}}}{c}\right]_{-\infty}^{\infty} \\ &= \frac{\exp \eta_{ij}}{\sum_k \exp \eta_{ik}} \end{aligned}$$

So

$$\frac{\Pr\{V_{ig} = V_{ijg}|\eta\}}{\Pr\{V_{ig} = V_{i0g}|\eta\}} = \exp(\eta_{ij} - \eta_{i0}) = \exp(\widetilde{\alpha}_{ij} - \alpha_{i0} - \tau_{ij}) \tag{33}$$

When there are many men of each type, we may approximate $\Pr\{V_{ig} = V_{ijg}|\eta\}$ with $\frac{\mu_{ij}}{m_i}$. Then (8) follows from (33).

11.2 Derivation of (9)

The index k runs from 0 to J. Observing male g of type i choose choice j, the expected utility of that individual is:

$$EV_{ijg} = \eta_{ij} + E(\varepsilon_{ijg} | \varepsilon_{ijg} + \eta_{ij} > \eta_{ik} + \varepsilon_{ikg} \; \forall k \neq j)$$

$$\begin{aligned} & E(\varepsilon_{ijg} | \varepsilon_{ijg} + \eta_{ij} > \eta_{ik} + \varepsilon_{ikg} \; \forall k \neq j) \\ = & \frac{\int_{-\infty}^{\infty} \varepsilon \exp\{-\sum_{k \neq j} e^{-\varepsilon - \eta_{ij} + \eta_{ik}}\} e^{-\varepsilon - e^{-\varepsilon}} d\varepsilon}{\Pr\{V_{ig} = V_{ijg} | \eta\}} \end{aligned} \tag{34}$$

Using (32) and the fact

$$\int_{-\infty}^{\infty} x e^x \exp(-\phi e^x) dx = -\frac{\Gamma + \ln \phi}{\phi}$$

where Γ is Euler's constant, $\simeq 0.577215$, (34) may be expressed as

$$E(\varepsilon_{ijg} | \varepsilon_{ijg} + \eta_{ij} > \eta_{ik} + \varepsilon_{ikg} \; \forall k \neq j) = \Gamma + \ln(\sum_k \exp \eta_{ik}) - \eta_{ij}$$

Thus

$$EV_{ijg} = \eta_{ij} + E(\varepsilon_{ijg} | \varepsilon_{ijg} + \eta_{ij} > \eta_{ik} + \varepsilon_{ikg} \; \forall k \neq j) = \Gamma + \ln(\sum_k \exp \eta_{ik}) \tag{35}$$

which is independent of j. Since knowing the optimal choice of the individual is not informative about his expected payoff, $EV_{ig} = EV_{ijg}$. Then (35) and (8) imply:

$$EV_{ig} = \Gamma + \ln(\sum_k \exp(\widetilde{\alpha}_{ik} - \tau_{ik})) = \Gamma + \widetilde{\alpha}_{i0} + \ln m_i - \ln \mu_{i0} \tag{36}$$

which is (9).

12 Appendix B

To apply the implicit function theorem to the system (21) and (22), we need to show that the Jacobian of the system is non-singular. The Jacobian is:

$$\begin{bmatrix} D_J & B \\ C & D_I \end{bmatrix}$$

where D_J is a $J \times J$ diagonal matrix where the jj element is $-1 - \sum_{i=1}^{I} \frac{\mu_{ij}}{2\mu_{0j}}$, and the off diagonal elements are zero. D_I is an $I \times I$ diagonal matrix where the ii element is $-1 - \sum_{j=1}^{J} \frac{\mu_{ij}}{2\mu_{i0}}$ and the off diagonal elements are zero. B is a $J \times I$ matrix whose ji element is $-\frac{\mu_{ij}}{2\mu_{i0}}$. C is an $I \times J$ matrix whose ij element is $-\frac{\mu_{ij}}{2\mu_{0j}}$.

As long as $\mu_{i0} \neq 0$ and $\mu_{0j} \neq 0$, we know D_I^{-1} and D_J^{-1} exist. Then using the formula for a partition inverse, the Jacobian is non-singular as long as

$$-\left[I_J - BD_I^{-1}CD_J^{-1}\right]^{-1} D_J^{-1}$$

exists.

Let $A = BD_I^{-1}CD_J^{-1}$, then $(I_J - A)$ is invertible if there is a matrix norm $\| \bullet \|$ such that $\|A\| < 1$. Consider the *maximum column sum matrix norm* defined by, $\|A\| = \max_j \sum_{i=1}^{n} |a_{ij}|$. Then:

$$\begin{aligned}
\|CD_J^{-1}\| &= \max_j \; \frac{\sum_i \mu_{ij}}{2\mu_{0j} + \sum_i \mu_{ij}} \; < \; 1 \\
\|BD_I^{-1}\| &= \max_i \; \frac{\sum_j \mu_{ij}}{2\mu_{i0} + \sum_j \mu_{ij}} \; < \; 1.
\end{aligned}$$

By definition of a matrix norm, $\|BD_I^{-1}CD_J^{-1}\| \;\leq\; \|BD_I^{-1}\| \cdot \|CD_J^{-1}\| < 1$, and hence $(I_J - A)^{-1}$ exists. $\square$

13 Appendix C: Data

Data used were extracted from the Integrated Public-Use Microdata (IPUMS henceforth) Files of the US Census. The samples used were the 5% state samples for 1980, and the 1% Form 1 and Form 2 samples for 1970. The 1970 datasets were appropriately scaled to be comparable with the 1980 files.[38]

To maintain consistency between states reporting marriages to the *Vital Statistics* and the data collected from the respective *US Census*, some states to be excluded. This result in the data from the following states being used: Alabama; Alaska; California

[38]State of residence in the 1970 census files can only be identified in the state samples (Form 1 and Form 2 samples, both of which are 1% samples). This is the reason that the other samples were not used for 1970 calculations. Further, the age of marriage variable is only available in Form 1 samples in 1970 which meant that only one sample, the Form 1 state sample, was used for calculations involving married couples in the 1970 census.

Connecticut; Delaware; District of Columbia; Florida; Georgia; Hawaii; Idaho; Illinois; Indiana; Kansas; Kentucky; Louisiana; Maine; Maryland; Massachusetts; Michigan; Mississippi; Missouri; Montana; Nebraska; New Hampshire; New Jersey; New York State; North Carolina; Ohio; Oregon; Pennsylvania; Rhode Island; South Dakota; Tennessee; Utah; Vermont; Virginia; West Virginia; Wisconsin and Wyoming.[39]

The age range studied was 16 to 75 years of age. Education level was identified using the "`higradeg`" variable in the 1970 and 1980 samples. This variable allowed us to assign each person one of the following schooling types: less than highschool, highschool graduate, and college degree or more.

We use the "`marst`" variable in the census to identify a person as either: never married, currently married (spouse present), or previously married (divorced or widowed). Further, the "`marrno`" variable (in 1970 and 1980 datasets) allows us to distinguish between married individuals in their first marriage and individuals in their second or later marriage.

To calculate the number of unmarried individuals of each type, we simply collapse the census data into counts by type. The process is straightforward and we don't loose any observations along the way.

References

[1] Akerlof, George, Janet Yellin, and Michael Katz. "An Analysis of Our-of-Wedlock Childbearing in the United States." *Quarterly Journal of Economics* 112, no. 2 (1996): 277-317.

[2] Angrist, Joshua. D. and William. N. Evans . Schooling and Labor Market Consequences of the 1970 State Abortion Reforms. *Research in labor economics 18* (1999): 75-113.

[3] Angrist, Josh. "How Do Sex Ratios Affect Marriage and Labor Markets? Evidence from America's Second Generation." *Quarterly Journal of Economics* CXVII, no. 3 (2002): 997-1038.

[39]In other words the excluded states (cities) are: Arizona, Arkansas, Colorado, Iowa, Minnesota, Nevada, New Mexico, New York City, North Dakota, Oklahoma, South Carolina, Texas and Washington.

[4] Aiyagari, S. Rao, Jeremy Greenwood, and Nezih Guner, 2000. "On the State of the Union." *Journal of Political Economy*, 109(2), 213-244.

[5] Baker, Michael, Emily Hanna, et al.. The Married Widow: Marriage Penalties Matter!, NBER WP 9782, 2003.

[6] Becker, Gary S. "A Theory of Marriage: Part I." *Journal of Political Economy* 81, no. 4 (1973): 813-46.

[7] Becker, Gary S. "A Theory of Marriage: Part II." *Journal of Political Economy* 82, no. 2, Part II (1974): S11-S26.

[8] Becker, Gary S. *A Treatise on the Family.* Cambridge, Mass.: Harvard University Press, 1991.

[9] Bergstrom, Theodore. "A Survey of Theories of the Family." In *Handbook of Population and Family Economics,* edited by Mark R. Rosenzweig and Oded Stark. Amsterdam, Elsevier Science B. V. 1A: 21-74, 1997.

[10] Bergstrom, Theodore and David Lam. "The Effects of Cohort Size on Marriage Markets in Twentieth Century Sweden", in *The Family, the Market, and the State in Ageing Societies*, edited by John Ermisch and Naohiro Ogawa, Oxford University Press, 1994, pp 46-63.

[11] Botticini, Maristella, and Aloysius Siow. "Why Dowries?" *American Economic Review*, 93, no. 4, 2003: 1385-1398.

[12] Casper, Lynne. M. and Suzanne. M. Bianchi. *Continuity and Change in the American Family.* Thousand Oaks, Calfornia, Sage Publications, 2002.

[13] Chiappori, Pierre-Andre, Bernard Fortin, and Guy Lacroix. "Marriage Market, Divorce Legislation, and Household Labor Supply." *Journal of Political Economy* 110, no. 1 (2002): 37-72.

[14] Dagsvik, John K. "Aggregation in Matching Markets." *International Economic Review* 41, no. 1 (2000): 27-57.

[15] Dagsvik, John K., and et. al. "A Behavioral Two-Sex Marriage Model." *Mathematical Population Studies* 9, no. 2 (2001): 97-121.

[16] Edlund, Lena. "Marriage Squeeze Interpretation of Dowry Inflation: A Comment." *Journal of Political Economy* 108, no. 6 (2000): 1327-33.

[17] Fernandez, Raquel, Nezih Guner, and John Knowles, "Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality", NBER working paper no. 8580, (2001).

[18] Goldin, Claudia, and Lawrence Katz. "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions." *Journal of Political Economy* 110, no. 4 (2002): 730-70.

[19] Grossbard-Shectman, Shoshana. *On the Economics of Marriage: A Theory of Marriage, Labor and Divorce.* Boulder: Westview Press, 1993.

[20] Gruber, Jonathan. Is Making Divorce Easier Bad for Children?, NBER WP 7968, 2000.

[21] Hamilton, Gillian and Aloysius Siow. "Class, Gender and Marriage", University of Toronto manuscript available at "www.economics.utoronto.ca/siow", 2000.

[22] Johansen, Rune, and John K. Dagsvik. "The Dynamics of a Behavioral Two-Sex Demographic Model." Statistics Norway, Research Department, 1999.

[23] Legros, Patrick and Andrew Newman. "Assortative Matching in a Non-Transferable World." University College London, Department of Economics, 2002.

[24] Logan, John A., Hoff, Peter D., and Newton, Michael A. "A Parametric Two-Sided Model of Marriage", manuscript, University of Wisconsin, Department of Sociology, 2001.

[25] McFadden, Daniel. "Conditional Logit Analysis of Qualitative Choice Behavior", in *Frontiers in Econometrics*, edited by P. Zarembka, New York: Academic Press, 1974

[26] Pollak, Robert. "Two-Sex Demographic Models." *Journal of Political Economy* 98, no. 2 (1990a): 399-420.

[27] ____________ "Two-Sex Population Models and Classical Stable Population Theory." In *Convergent Issues in Genetics and Demography*, edited by Julian Adams and et. al., 317-33. New York, Oxford: Oxford University Press, 1990b.

[28] Pollard, John H. and Charlotte Hohn. "The Interaction between the Sexes." *Zeitschrift fur Bevolkerungswissenschaft* 19, no. 2 (1993-94): 203-08.

[29] Pollard, John H. "Modelling the Interaction between the Sexes." *Mathematical and Computer Modelling* 26, no. 6 (1997): 11-24.

[30] Qian, Zhenchao and Sam Preston. "Changes in American Marriage, 1972 to 1987: Availability and Forces of Attraction by Age and Education," *American Sociological Review,* 58, August (1993): 482-495.

[31] Qian, Zhenchao. "Changes in Assortative Mating: The Impact of Age and Education, 1970-1990." *Demography* 35, no. 3 (1998): 279-92.

[32] Rao, Vijayendra. "The Rising Price of Husbands: A Hedonic Analysis of Dowry Increases in Rural India." *Journal of Political Economy* 101, no. 4 (1993): 666-77.

[33] Roth, Alvin, and Marilda A. Oliveira Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis.* New York: Cambridge University Press, 1990.

[34] Ruud, Paul. *An Introduction to Classical Econometric Theory.* New York: Oxford University Press, 2000.

[35] Schoen, Robert. "The Harmonic Mean as the Basis of a REalistic Two-Sex Marriage Model." *Demography*, 18, May (1981): 201-216.

[36] Seitz, Shannon. "Employment and the Sex Ratio in a Two-Sided Model of Marriage." Queen's University manuscript, 1999.

[37] Shimer, Robert and Lones Smith. "Assortative Matching and Search." *Econometrica*, 68, no. 2 (2000): 343-370.

[38] Siow, Aloysius. "Do Innovations in Reproductive Technologies Improve the Welfare of Women.", University of Toronto manuscript available at "www.economics.utoronto.ca/siow", 2002.

[39] South, Scott, and Katherine Trent. "Sex Ratios and Women's Roles: A Cross-National Analysis." *American Journal of Sociology* (1988): 1096-115.

[40] South, Scott, and Kim M. Lloyd. "Marriage Opportunities Amd Family Formation: Further Implications of Imbalanced Sex Ratios." *Journal of Marriage and the Family* 54 (1992): 440-51.

[41] Suen, Wing and Lui, Hon-Kwong. "A Direct Test of the Efficient Marriage Market Hypothesis." *Economic Inquiry* 37, no. 1 (1999): 29-46

[42] Waite, Linda, et. al., Ed. *The Ties that Bind: Perspectives on Marriage and Cohabitation.* Hawthorne, New York, Aldine De Gruyter, 2000.

[43] Weiss, Yoram. "The Formation and Dissolution of Families: Why Marry? Who Marries Whom? And What Happens Upon Divorce." In *Handbook of Population and Family Economics.* edited by Mark Rosenzweig and Oded Stark. Amsterdam, Elsevier Science B. V. 1A: 81-124, 1997.

[44] Wong, Linda. "Structural estimation of marriage models." *Journal of Labor Economics* 27, no. 3 (2003a): 699-728.

[45] Wong, Linda. "An Empirical Study of Darwin's Theory of Mate Selection", SUNY at Binghampton manuscript, 2003b.

[46] Yatchew, Adonis. *Semiparametric Regression for the Applied Econometrician.* Cambridge: Cambridge University Press, 2003.