WHO MARRIES WHOM AND WHY*

Eugene Choo & Aloysius Siow

University of Toronto

January 14 2003

Abstract

Using a transferable utility model of the marriage market, the paper derives a statistic to measure the systematic welfare gain to any two types marrying each other relative to them not marrying. This statistic also defines a non-parametric marriage matching function which can be used to (1) decompose the distribution of marital matches into the share attributable to preferences and that attributable to the availability of different types of participants and (2), to predict changes in who marries whom as the distributions of types of participants change. This marriage matching function is estimated with data from the 1970 US Census and Vital Statistics. The estimated model is able to largely predict marital behavior by age and education in 1980. It over predicts the marriage rate of young adults for all education groups. Our analysis suggests that the gains to marriage for young adults fell substantially from 1970 to 1980.

* We thank Pierre Andre-Chiappori, Angelo Melino and Adonis Yatchew, and seminar participants at Boston University, Northwestern University, 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 Introduction

The marriage market is a bilateral matching market with heterogenous participants. After controlling for the availability of the types of participants, some marital matches occur with higher frequencies than others. For example, positive assortative matching by age occur more frequently than negative assortative matching by age. Do the relative frequencies of different marital matches in a society have a normative interpretation? Can the distribution of matches be decomposed into that attributable to preferences and that attributable to the availability of different types of participants? These questions are related to the larger objective of finding ways to empirically characterize bilateral matching markets with heterogenous participants.[1]

This paper answers both questions in the affirmative. We propose a new statistic to measure the systematic welfare gain to any two types marrying each other relative to them not marrying. This statistic can be used to decompose the distribution of matches into the share attributable to preferences and that attributable to the availability of different type of participants. It can also be used to predict changes in who marries whom as the distributions of types of participants change. We will derive the statistic and its normative interpretation from an explicit model of the marriage market.

The statistic is defined as follows. Consider a society with I types of men and J types

[1] Other markets include the labor market, the market for roomates, the market for real estate contractors and in general, markets where demanders care about the identities of suppliers and vice versa.

of women. The type of a person may be multidimensional, including the person's age, education, ethnicity, etc.. Let m_i be the number of available type i men, and f_j be the number of available type j women at a point in time. Let μ_{ij} be the number of type i men who marries type j women at that time. The statistic, Π_{ij}, is given by:

$$\Pi_{ij} = \frac{\mu_{ij}}{\sqrt{(m_i - \sum_k \mu_{ik})(f_j - \sum_l \mu_{lj})}} \tag{1}$$

In words, Π_{ij} is equal to the ratio of the number of i, j marriages to the geometric average of the number of type i men and number of type j women who chose not to marry.

This statistic is derived from an explicit model of the marriage market. The model provides a normative interpretation to Π_{ij}. Our 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 *idiosyncratic* payoffs 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. Throughout this paper, we use the words payoff, gains and returns interchangeably.

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. How we scale is also important. Consider an alternative statistic κ_{ij} of the systematic gains given by:

$$\kappa_{ij} = \frac{\mu_{ij}}{g(m_i, f_j)}$$

In this case, the number of i, j marriages is scaled by the function $g(m_i, f_j)$. Reinterpreted as a marriage matching function, κ_{ij} is commonly used in the demographic literature (Qian 1998; Pollard 1997; Pollak 1990a). Unlike Π_{ij}, this alternative statistic imposes behaviorally implausible restrictions on marriage behavior. In general, the number of i, j marriages will depend on the availability of the different types of market participants. But κ_{ij} assumes that the systematic gains to an i, j marriage can be estimated without considering the supplies of other types of men and women. This behaviorally implausible restriction is referred to as the zero spillover hypothesis in the marriage matching function literature (Pollak 1990a).

Equation (1) can be reinterpreted as a marriage matching function. Consider the $I \times J$ matrices, μ and Π, whose i, j elements are μ_{ij} and Π_{ij} respectively. Let M be the population vector of available men whose i'th element is m_i, and F be the population vector of available women whose j'th element is f_j. Equation (1) implicitly defines a marriage matching function, where μ is a function of M, F and Π. The i, j element of μ is:

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

The marriage function is homogenous of degree one in M and F. Doubling the population vectors will double the number of marriages. So the marriage function has no scale effect in population vectors[2] .

The statistic as defined in Equation (1) allows us to estimate the systematic returns to any i, j marriage, Π_{ij}, using data from one point in time. The associated marriage matching function is non-parametric. It fits any observed marriage distribution. Since the type space is in general multidimensional and not necessarily ordered, it is important that our marriage matching function does not impose apriori structure on the marriage distribution.

Using ages as the only types for males and females, the second part of the paper estimates Π using data from the 1970 *US Census* and *Vital Statistics*. We use our marriage matching function, our estimate of the systematic returns to marriage, Π, and 1980 population vectors, M^{80} and F^{80}, to forecast the marriage distribution for 1980. 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 1970 and 1980 due to changes in population vectors between the two periods. However we over predict the number of marriages among young adults in 1980. When we expanded the type space to include educational attainment, the over prediction, although mitigated, was still present. This suggests that the gains to marriage fell between 1970 and 1980 for young adults of all educational groups.[3]

This paper uses results from several well known literatures. Our derivation of Π is based on Becker's seminal model of the marriage market, summarized in his 1991 book. A primary insight of that model is that transfers between spouses are used to clear the marriage

[2] Pollak (1990b) argues that no scale effect is a reasonable requirement for marriage matching functions.
[3] Demographers already noticed this decline in the marriage rate of young adults (E.g. Qian).

market. His model has had enormous influence on the empirical study of marriage markets. Researchers have used it to study the relationships between the sex ratio, the ratio of available men to women, 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 2002, South and Lloyd 1992, South and Trent 1988). None of these studies estimate the bivariate distribution of spousal ages of marriage. Our paper complements these empirical studies by developing and implementing an analytically and empirically tractable transferable utility model of marriage with heterogenous participants. We use McFadden's (1974) extreme value random utility model to generate demand and supply functions for different types of marriages.

Marriage matching functions are well known in demography (Pollak 1990a, 1990b; Pollard). This paper builds upon the demographers' use of marriage market concerns to restrict the marriage matching function (E.g. Pollard and Höhn (1993/94)).

An important antecedent to our work is Dagsvik (2000) who anticipated our methodology of using an explicit model of the marriage market to construct marriage matching functions.[4] A comparison with his work is provided in Section 5.

The statistic, Π and its associated matching function have applications beyond the marriage market. It may be used to empirically characterize finite type competitive bilateral matching markets with heterogenous participants. In fact, applications involving bilateral markets where transfers or transaction prices are observed allow more parameters of the

[4] Also see Johansen and Dagsvik 1999; Dagsvik, and et. al. 2001.

market to be identified. Without transaction prices, only the systematic gains to a match relative to not matching can be identified. As to be discussed in Section 2.1 of the paper, with transaction prices or transfers, the party specific systematic gains can be separately identified. The non-parametric nature of our empirical model allows us to fit any finite type observed competitive bilateral matching distribution and pricing function.[5] Our model is also a competitive hedonic market model. Unlike the usual competitive hedonic model (E.g. Bayer, et. al. 2002; Eckland, et.al. forthcoming; Rosen 1974), our market is one of bilateral matching and so our equilibrium pricing function depends explicitly on the types of the buyers and sellers. In the usual case, the pricing function depends on the characteristics of the good being sold, and not the characteristics of buyers and sellers.

2 The model

We begin by describing Becker's 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. At a point in time, each type of individual considers matching with each type of the opposite gender.

[5] It cannot deal with thin cells, except in an ad hoc way, but thin cells also violate the competitive assumption.

Let the utility of male g of type i who marries a female of type j be:

$$V_{ijg} = \widetilde{\alpha}_{ij} - \tau_{ij} + \varepsilon_{ijg}, \qquad \text{where} \tag{2}$$

$\widetilde{\alpha}_{ij}$: 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.[6]

Equation (2) 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 females, f_j, of type 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.[7] The payoff to g from remaining unmarried, denoted by $j = 0$, is:

$$V_{i0g} = \widetilde{\alpha}_{i0} + \varepsilon_{i0g} \tag{3}$$

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

Individual g will choose according to:

$$V_{ig} = \max_j [V_{i0g}, .., V_{ijg}, .., V_{iJg}] \tag{4}$$

[6] The random variable $\varepsilon_{ijg} \sim EV(0, 1)$, with the cumulative distribution given by $F(\varepsilon) = e^{-e^{-\varepsilon}}$.
[7] Dagsvik makes an alternative assumption as discussed in Section 5.

We assume that the numbers of men and women of each type is large. Let s_{ij} be the share of type i men married to type j women, and s_{i0} is the share of type i men that are unmarried. Then McFadden showed[8] that

$$\ln s_{ij} - \ln s_{i0} = \widetilde{\alpha}_{ij} - \widetilde{\alpha}_{i0} - \tau_{ij} \tag{5}$$

$$= \alpha_{ij} - \tau_{ij} \tag{6}$$

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. Let μ_{ij} is the number of ij marriages and μ_{i0} is the number of unmarried i type men. The share equation can be written as:

$$\ln s_{ij} - \ln s_{i0} = (\ln \mu_{ij} - \ln m_i) - (\ln \mu_{i0} - \ln m_i)$$

$$= \ln \mu_{ij} - \ln \mu_{i0} \tag{7}$$

Then Equations (5) and (7) implies

$$\ln \mu_{ij} = \ln \mu_{i0} + \alpha_{ij} - \tau_{ij} \tag{8}$$

The above is a quasi-demand equation by type i men for type j women. [9] 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 $\ln \mu_{i0}$.

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

[8] The result is also derived in many econometrics textbook (e.g. p. 780, Rudd 2000).
[9] It is not a demand curve because $\mu_{i0} = m_i - \sum_j \mu_{ij}$.

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. Let S_{ij} be the share of type j women who marry type i men, S_{0j} be the share of type j women that are unmarried, and μ_{0j} be the number of unmarried j type women. The corresponding share equations for a j type women is given by:

$$\ln S_{ij} - \ln S_{0j} = \ln \mu_{ij} - \ln \mu_{0j} \tag{9}$$

$$= \gamma_{ij} + \tau_{ij} \tag{10}$$

The quasi-supply equation of type j women who marry type i men is be given by

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

Again, the transfers for all the other types of men other than i is embodied in $\ln \mu_{0j}$. 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.[10]

When the competitive marriage market for all i, j pair clears, we can equate demand and supply given by Equations (8) and (11) to get:

$$\tau_{ij} = \frac{\ln \mu_{i0} - \ln \mu_{0j} + \alpha_{ij} - \gamma_{ij}}{2} \tag{12}$$

Substituting (12) into (11), we get:

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

[10]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.

If we let $\pi_{ij} = \ln \Pi_{ij} = \frac{\alpha_{ij}+\gamma_{ij}}{2}$, we can rewrite Equation (13) as:

$$\Pi_{ij} = \frac{\mu_{ij}}{\sqrt{\mu_{i0}\mu_{0j}}} \tag{14}$$

which is the statistic define in Equation (1) in the introduction to this paper. It also defines an implicit marriage matching function.

Equation (14) has an intuitive interpretation. The right hand side of (14) is the ratio of the number of i, j marriages to the geometric average of those types who are 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.[11]

The statistic as defined in Equation (14) also imposes strong restrictions on the data. It is homogeneous of degree zero in population vectors and the number of marriages. That is, doubling m_i, f_j, and μ_{ij} for all i, j's leave the $\pi'_{ij}s$ in (14) unchanged. 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

[11]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 $2\pi_{ij}$.

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 (14) 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 (14) 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 (12). In other words, knowing the systematic gains to a match is not sufficient to determine whether men pay positive or negative transfers to women in equilibrium. On the other hand, in applications where τ_{ij} is also observed, then Equations (12) and (14) would allow us to identify α_{ij} and γ_{ij}.[12] In these cases, Equations (12) and (14) 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 function 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

[12] 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 (2000)). See Edlund (2000) for an example of the problems that arise when this caveat is ignored.

is:

$$\frac{\mu_{ij}}{\mu_{i0}} = e^{\alpha_{ij}-\tau_{ij}} = n_{ij} \tag{15}$$

$$\frac{\mu_{ij}}{\mu_{0j}} = e^{\gamma_{ij}+\tau_{ij}} = N_{ij} \tag{16}$$

We will refer to n_{ij} as the systematic "*net gain*" to marriage for a type i male in an i, j marriage relative to not marrying, and N_{ij} as the systematic "*net gain*" to marriage for a type j female in an i, j marriage relative to not marrying. The net gain to marriage for an individual depends on both preferences and equilibrium transfers. If $\frac{\mu_{ij}}{\mu_{i0}} > \frac{\mu_{ij'}}{\mu_{i0}}$, the net gain to a type i male from marriage to a type j spouse is higher than to a type j' spouse. Furthermore,

$$\sum_j n_{ij} = \sum_j \frac{\mu_{ij}}{\mu_{i0}} = \frac{m_i}{\mu_{i0}} - 1 \tag{17}$$

$$\sum_i N_{ij} = \sum_i \frac{\mu_{ij}}{\mu_{0j}} = \frac{f_i}{\mu_{0j}} - 1 \tag{18}$$

The sum of the net gains to marriage for a type i over all the different types of spouses is proportional to the marriage rate of his type, $\frac{m_i - \mu_{i0}}{m_i}$. Likewise for a female of type j. The net gains results are well known properties of McFadden's random utility model.

3 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 proposed statistic may be rewritten as

$$\mu^t_{ij} = \Pi_{ij}\sqrt{\mu^t_{i0} \times \mu^t_{0j}} \tag{19}$$

$$= \Pi_{ij}\sqrt{(m^t_i - \sum_{k=1}^{J}\mu^t_{ik})(f^t_j - \sum_{g=1}^{I}\mu^t_{gj})} \tag{20}$$

If we take Π_{ij}, M and F 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^t_{i0} and μ^t_{0j}, as unknowns. This reduced system is defined by Equations (21) and (22) below. If we can solve for μ^t_{i0} and μ^t_{0j}, then the μ^t_{ij}'s are fully determined by Equation (19). To derive this system of equations, we sum Equation (19) over all i's to get:

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

$$f^t_j - \mu^t_{0j} = \sum_{i=1}^{I}\Pi_{ij}\sqrt{\mu^t_{i0} \times \mu^t_{0j}} \tag{21}$$

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

$$m^t_i - \mu^t_{i0} = \sum_{j=1}^{J}\Pi_{ij}\sqrt{\mu^t_{i0} \times \mu^t_{0j}} \tag{22}$$

Given population quantities M, F, μ and Π as defined in Equation (14), 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.

4 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.[13] In our model, the competitive market clearing conditions place further restrictions on the returns and transfers from marriage. To illustrate, consider the ratio of spousal net gains from an i, j marriage, which from Equations (15) and (16), is given by $\frac{\mu_{j0}}{\mu_{i0}} = \frac{n_{ij}}{N_{ij}}$. If we compare this ratio across two different types of males, say i and i', while holding the type of the female spouse constant, we get the following result:

$$\frac{\frac{n_{ij}}{N_{ij}}}{\frac{n_{i'j}}{N_{i'j}}} = \frac{\mu_{i0}}{\mu_{i'0}}, \ \forall \ i, i', j \tag{23}$$

Equation (23) says that the ratio of spousal net gains in an i, j marriage divided by the ratio of the spousal net gains in an i', j marriage is the same for all j. Equation (23) imposes strong restrictions on equilibrium transfers and by implication, how marriages respond to changes in population vectors. At present, we do not have a test of how plausible these restrictions are. 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

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

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. Since we want to use the model to make forecasts, we will maintain the assumption that Π_{ij}^t is uncorrelated with population vectors at time t. A discussion of the results from these specification tests is given in Section 6.4 of the paper.

5 Dagsvik's model

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

$$\theta_{ij} = \frac{\mu_{ij}}{\mu_{i0}\mu_{0j}} \tag{24}$$

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 a i, j match. ε_{ijgk} denotes his idiosyncratic returns from a match between him and the j type female individual k.[14] 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 accep-

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

tance 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 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.

6 Empirical evidence - Overview of results

We begin with an overview of our empirical methodology and the results. The data used in this paper comes from the 1970 and 1980 *US Census* and *Vital Statistics*. The data we seek to explain are the bivariate age distributions of marriages in 1970 and 1980. At each date, 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 the Appendix A.

Initially, the type space only consists of the ages of individuals. Using the observed marriages in 1970, μ^{70}, and population vectors, M^{70} and F^{70}, we estimate the preference parameters of the model, $\widehat{\Pi}^{70}$. We test the estimated model in two ways. First, our model implies that $\widehat{\Pi}^{70}$ is unrelated to population vectors at time t, M^{70} and F^{70}. We test this implication.

Second, we use the estimated 1970 parameters, $\widehat{\Pi}^{70}$, and observed population vectors in 1980, M^{80} and F^{80}, to predict 1980 marriages, $\widehat{\mu}^{80}$. We compare how well the predicted marital distribution $\widehat{\mu}^{80}$ fits the actual μ^{80} marital distribution.

Anticipating the results, we find that $\widehat{\mu}^{80}$ is able to explain much of the variation in μ^{80}. However it over predicts the number of marriages among young adults. We deal with this prediction error in two ways.

First, we expand the type space to include levels of education. The fraction of young adults who have attended college increased substantially between 1970 and 1980, while the fraction with less than high school decreased over the decade. Besides significant difference in changes in available men and women across education type, the marriage behavior across education type also differs. College educated individuals marry later than non-college educated individuals, suggesting that the preference for marriage from these types are significantly different. Accounting for the change in educational attainment between the decade and the heterogeneity in marriage behavior across individuals with differing education attainment may improve the fit of $\widehat{\mu}^{80}$.

Introducing educational attainment as an additional type significantly improved the fit of $\widehat{\mu}^{80}$. Nonetheless, we are still unable to account for all the over prediction of marriages among young adults. We attribute the over prediction to changes in preferences, π^t, between the decade. To study this change in preferences, we also estimate π^{80}. We conclude the empirical study with a discussion of the changes in the gains to marriage over the decade.

6.1 Data Summary

Population vectors in 1970 and 1980 are obtained from the respective *US Census.*[15] The number of marriages in 1970 and 1980 are obtained from *Vital Statistics.* A state has to report the number of marriages to *Vital Statistics* in 1970 and 1980 to be in the sample. This requirement eliminated 12 states.[16] In our sample of states, there were 16.8 million and 20.9 million available men and women respectively between the ages of 16 to 75 in 1970. From this population, there were 1,625,789 marriages in 1970. There were 24.2 million and 28.6 million available men and women respectively in 1980. Although the available population increased by more than 35% over the decade, there were only 1,698,579 marriages in 1980, an increase of 4.5%. A summary of the data set is in Table 1 below.

Figure 1a and 1b show the bivariate age distribution of the marrieds in 1970 and 1980 respectively. In both years, most marriages occur between young adults and there is strong positive assortative matching by age. Figures 1a and 1b cannot fully capture the magnitudes of the changes in the marriage market and population vectors between the decades.

[15] We use the sampling weights in the Census to obtain population vectors.
[16] The states of Arizona, Arkansas, Nevada, New Mexico, North Dakota, Oklahoma, Texas, and Washington never reported. Compared with 1970, three more states reported in 1980: Minnesota, South Carolina, Colorado, while Iowa ceased to report.

Table 1

	1970	1980	Δ
Number of Available Males (M^t)	16.8 mil	24.2 mil	44%
Number of Available Females (F^t)	20.9 mil	28.6 mil	37%
Number of Marrieds (μ^t)	1.63 mil	1.70 mil	4.5%
% Males with college & above education (21+)	18%	26.8%	49.7%
% Females with college & above education (21+)	13%	19.6%	49.1%
Average age of Available Males	30.7	29.8	
Average age of Available Females	39.4	37.4	
Average age Married Males	27.1	28.7	
Average age Married Females	24.6	26.1	

Figure 2a shows the 1970 age distributions of the population vectors and the marrieds.[17] For each gender, the area under the married distribution is almost equal to the height of the available individuals at 16, reflecting the fact that most individuals will eventually marry. The figure also shows that at each age, most available individuals do not marry.

Figure 2b shows the age distributions of the population vectors and the marrieds in 1980. Comparing Figures 2a and 2b, we see that the baby boom generation came of marriageble age between the decade, substantially increasing the population of the availables in 1980. The average age of available men fell from 30.7 in 1970 to 29.8 in 1980 and that of available women fell from 39.4 in 1970 to 37.4 in 1980. However, the number of young marrieds in

[17]The average age of available men and women were 30.7 and 39.4 respectively. This gender difference reflected the larger fraction of available older women. The average age of the married men and women were 27.1 and 24.6 respectively, reflecting the usual gender difference in ages of marriage.

1980 barely increased. Any marriage matching function with no scale effect, estimated with 1970 data will predict a substantial increase in the number of marriages in 1980 due to the substantial increase in available young adults over the decade.

6.2 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 1970 estimates for type i males, given by $\widehat{n_{ij}^{70}} = \frac{\mu_{ij}^{70}}{\mu_{i0}^{70}}$, and j type females, given by $\widehat{N_{ij}^{70}} = \frac{\mu_{ij}^{70}}{\mu_{0j}^{70}}$ are compared in Figure 3. Figure 3a plots $\widehat{n_{ij}^{70}}$ and $\widehat{N_{ij}^{70}}$ 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 $\widehat{N_{i,20}^{70}}$ is right skewed, with the 20 years old female receiving the largest systematic net gain when she marries a slightly older male. In contrast, the distribution of $\widehat{n_{20,j}^{70}}$ is more symmetric and concentrated, with the 20 years old male receiving the largest systematic net gain when he marries a slightly younger female.

According to Equations (17) and (18) stated earlier, the area below these plots is proportional to the type specific marriage rates. The smaller area under $\widehat{n_{20,j}^{70}}$ relative to that of $\widehat{N_{i,20}^{70}}$ suggests that the marriage rate of 20 year old females is larger than that of 20 year old males. Comparing the distribution of systematic net gain for a 30 years old female, $\widehat{N_{i,30}^{70}}$, with her 20 years old counterpart, we find the distribution for a 30 years old female to be more dispersed and the marriage rate to also be significantly lower. Again she receives the largest net gain when she marries someone slightly older. If we consider the distribution for 30 year old males, $\widehat{n_{30,i}^{70}}$, 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. Comparing the areas under the respective distributions, his marriage rate is higher than his 30 year old female counterpart.

Figure 3b compares the systematic net gains to marriage for 40 and 50 years old males and females. The difference in the scale of the vertical axis in Figures 3b and 3a reflects the fact 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.

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.

6.3 Estimating Π - age as only type

We will now briefly describe our strategy for estimating the systematic gains to marriage, Π^{70}.[18] Given observed vectors of marriages, and the number of unmarried males and females in 1970, denoted by $\{ \mu_{ij}^{70}, \mu_{i0}^{70}, \mu_{0j}^{70} \}$ respectively, the marriage matching function defined in Equation (14) provides an estimate for the systematic gains to marriage in 1970, $\widehat{\Pi}_{ij}^{70}$.

[18]Some details are left to Appendix C.

Assuming that preferences are independent of population vectors and doesn't change, we can make predictions of marriages in 1980 given observed population vectors in 1980. The system of nonlinear equations that we solve to make predictions, Equations (22) and (21) again are as follows:

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

We numerically solve for the unique vector of predicted unmarrieds in 1980, $\{\ \widehat{\mu}_{i0}^{80},\ \widehat{\mu}_{0j}^{80}\ \}$ given observed population vectors in 1980, $\{m_i^{80}, f_j^{80}\}$ and the estimated parameters, $\widehat{\Pi}_{ij}^{70}$.[19] The predicted marriages in 1980, $\widehat{\mu}_{ij}^{80}$, can be solved as a function of the unmarrieds according to Equation (19), that is, $\widehat{\mu}_{ij}^{80} = \widehat{\Pi}_{ij}^{70} \sqrt{\widehat{\mu}_{i0}^{80} \cdot \widehat{\mu}_{0j}^{80}}$

In the 1970 and 1980 data, there were many i, j pairs which had no marriage. This is a common problem in empirical discrete choice applications. We dealt with thin cells in two ways.

In our first approach, the estimated $\widehat{\Pi}_{ij}^{70}$ is set to zero for any i, j pair that has no marriage in 1970. This is equivalent to estimating an extremely negative value to the systematic gains to marriage, that is, $\widehat{\pi}_{ij}^{70} = -\infty$. This approach only estimated $\widehat{\Pi}_{ij}^{70}$ when the observed number of i, j marriages in 1970 was strictly positive. The main shortcoming is that the corresponding predicted marriages in 1980 for that i, j pair would also be restricted to zero. There would be as many zero predicted marriages in 1980 as there are zero marriages in the 1970 data.

[19]Unique solution exist as per Proposition 2.

In the second approach, we first fit a bivariate log normal distribution over the observed 1970 marriage data. The estimated distribution is then used to predict the number of marriages for each i, j cell that had less than a chosen cutoff of observed marriages.[20] The other cells are left as they were. The exact details are outlined in Appendix C. This second approach overcomes part of the thin cell concerns and some of the limitations of the first approach.

On the whole, the predicted marriage distributions based on the two different methods are essentially identical. Differences in predictions arose only in thin cells but cells with few marriages have little influence on the predicted cells with more observations. Given the similarity between the two methods, we will focus our discussion on the results from the second approach. We will also periodically refer to these estimates as our smoothed estimates.

Figure 4a presents a plot of the distribution of $\widehat{\pi}^{70}$ using the unsmoothed estimates.[21] These estimates of $\widehat{\pi}_{ij}^{70}$ has been truncated for thin cells that typically occur at extreme differences in ages of marriages (for example, marriages between young males and old females). The surface of $\widehat{\pi}^{70}$ is also smoother among young adults reflecting the fact that there were more marriages in those cells. As cells become thinner, estimates of $\widehat{\pi}_{ij}^{70}$ become more variable. All the values in $\widehat{\pi}^{70}$ are negative reflecting the fact that the systematic gains to marriage is smaller than not marrying. The parametric specification of our model predicts

[20]The cutoff is set to 0.002% of the total number of observed marriages in 1970. We repeated the estimation for a number of sensible values, without any significant change on our results.

[21]The smoothed estimates predict large negative values of $\widehat{\pi}_{ij}$ for thin cells. Plotting these large negative estimates (where there is barely any data) masks the features of the the $\widehat{\pi}$ distribution where there is a lot of data.

a match to occur only if the idiosyncratic draws is large. The negative values should also not be surprising since at any age, most available individuals do not marry (as suggested by Figure 2a). The estimated systematic gains to marriage fell as individuals aged. It also systematically fall away from the diagonal (where $i = j$). Thus there are large gains to positive assortative matching by age. Although somewhat more difficult to see, there is also a ridge along the diagonal. The systematic gains fell faster in the southwest direction away from the ridge than in the northeast direction. In other words, the systematic gains to marriage between older women and younger men were less than that between older men and younger women.

Comparing Figure 4(a) which is plots $\widehat{\pi}^{70}$ with Figure 1(a) which plots μ^{70}, illustrates the importance of scaling the number of marriages to obtain estimates of the systematic gains to marriage. While the estimate of the systematic gains to marriage is highest between young adults, the decline for older adults and between young and old adults is relatively gradual. On the other hand, Figure 1(a) shows that the number of marriages is largest between young adults and falls rapidly elsewhere. Thus one would not want to use the distribution of μ to estimate the systematic gains to marriages.

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.

6.4 Testing model mispecification

A strong implication of our model, as given in Equation (14), is that π^t only reflect preference parameters and is independent of population vectors. To the extent that dynamic

TABLE 2

	OLS	MED	IV
Dependent var.	$\widehat{\pi}_{ij}^{70}$	$\widehat{\pi}_{ij}^{70}$	$\widehat{\pi}_{ij}^{70}$
$\ln(M_i^{70})$	2.1778 (0.3707)	1.7970 (0.0565)	7.3138 (1.3917)
$\ln(F_j^{70})$	2.3693 (0.3247)	2.4146 (0.0579)	3.6382 (0.8229)
Age polynomials	Y	Y	Y
Observations	2456	2456	2456
R^2	0.73		0.52
Instruments			$\ln(M_i^{80})$, $\ln(F_j^{80})$

Robust standard errors in parentheses.

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 $\widehat{\pi}_{ij}^{70}$ on demographics and 1970 population vectors.[22]

We present results from OLS, median and IV regressions. IV regression was carried out because $\widehat{\pi}_{ij}^{70}$ is constructed with 1970 population vectors. So if there is measurement error in our measure of the population vectors, this may induce a correlation between $\widehat{\pi}_{ij}^{70}$ 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 cubic polynomials of ages, i and j.

[22]We do not use the observations where the actual number of marriage in those i, j cells is zero because our smoothed estimates of $\widehat{\pi}_{ij}^{70}$ are relatively large negative numbers in those cases and including them change some point estimates substantially. But the qualitative conclusion remains unchanged.

All three methods of estimation give the same results. Even after controlling for demographics (using a cubic polynomial in ages), the 1970 population vectors can still explain variations in $\widehat{\pi}_{ij}^{70}$.[23] The estimated coefficients for the population vectors are statistically significant at the 1% level in all cases. This result strongly suggests that π_{ij}^{70} is correlated with the population vectors in 1970, a violation of our model. As discussed in Section **??**, this violation suggests that dynamic considerations, scale effects in the marriage market and/or other forms of misspecifications are important. We plan to address these concerns in our future research.

6.5 Predictive power

In this section, we provide some tests of the predictive power of our model. We use our estimates of $\widehat{\pi}^{70}$ and the 1980 population vectors, M^{80} and F^{80}, to predict the 1980 marriage distribution, $\widehat{\mu}^{80}$. Let $\Delta \ln \mu_{ij} = \ln \mu_{ij}^{80} - \ln \mu_{ij}^{70}$ and $\widehat{\Delta \ln \mu}_{ij} = \ln \widehat{\mu}_{ij}^{80} - \ln \mu_{ij}^{70}$. We can decompose the growth rate in the number of i, j marriages into a constant plus an idiosyncratic shock

$$\Delta \ln \mu_{ij} = \lambda + \varepsilon_{ij}$$

Similarly, let

$$\widehat{\Delta \ln \mu}_{ij} = \eta + \nu_{ij}$$

Then consider a regression of $\Delta \ln \mu_{ij}$ against $\widehat{\Delta \ln \mu}_{ij}$. The estimated coefficient on $\widehat{\Delta \ln \mu}_{ij}$ is equal to $\frac{cov(\varepsilon_{ij},\nu_{ij})}{var(\nu_{ij})}$ and the R^2 from the regression measures the proportion of the variation in $\Delta \ln \mu_{ij}$ explained by variation in $\widehat{\Delta \ln \mu}_{ij}$.

[23]The 1970 population vectors remain statistically significant when we used fifth order polynomials in age.

Column 1 in Table 3 presents the OLS regression of $\Delta \ln \mu_{ij}$ against $\widehat{\Delta \ln \mu}_{ij}$.[24] The estimated coefficient on $\widehat{\Delta \ln \mu}_{ij}$ is 0.82 and it is statistically significant at the 1% level. The R^2 suggests that $\widehat{\Delta \ln \mu}_{ij}$ is able to explain 73% of the variation in the growth rate of the number of i, j marriages between 1970 and 1980.

Column 2 in Table 3 adds $\ln F_j^{70}$, $\ln F_j^{80}$, $\ln M_i^{70}$ and $\ln M_i^{80}$ as regressors. The estimated coefficient on $\widehat{\Delta \ln \mu}_{ij}$ becomes 0.93 and remains statistically significant at the 1% level. Since the population vectors are present in this regression, the continued predictive power of $\widehat{\Delta \ln \mu}$ is due to non-linear adjustments of $\widehat{\Delta \ln \mu}$ to changes in the population vectors. In summary, the results in Table 3 show that $\widehat{\Delta \ln \mu}$ have significant predictive power.

Although successful, Table 3 masks a particular failure with our predictive distribution. Figure 5a and 5b plots the observed and predicted marriage distribution in 1980 respectively. Figure 5b shows that we over predicted the number of marriages among young adults. A plot of the prediction error is in Figure 5c. We over predict marriages among young adults and there is a slight under prediction for adults in their late twenties and early thirties. The over prediction should be expected. As shown earlier in Figure 2a and 2b, the young adult population grew substantially over the decade, while marriages did not. Most marriage functions with only age as the type space, without scale effect, would predict substantial increase in marriages in 1980 given the change in the population vectors.

Note however that most models of bilateral matching predict increasing or constant returns to scale in population vectors.[25] Models with increasing returns will do an even worse

[24]Observations of i, j pair with zero marriage in either 1970 or 1980 were omitted. There are 1917 observations in the regression.
[25]Increasing returns come from liquidity and market thickness concerns.

TABLE 3

	OLS	OLS
Dependent var.	$\Delta \ln \mu_{ij}$	$\Delta \ln \mu_{ij}$
$\widehat{\Delta \ln \mu_{ij}} = \ln \hat{\mu}_{ij}^{80} - \ln \mu_{ij}^{70}$	0.8236 (30.68)	0.9262 (27.18)
$\ln(F_j^{70})$		1.3417 (6.87)
$\ln(F_j^{80})$		-0.9907 (4.62)
$\ln(M_i^{70})$		1.5455 (4.68)
$\ln(M_i^{80})$		-1.6709 (6.17)
Observations	1917	1917
R^2	0.73	0.75

Robust t-statistics in parentheses.

job of predicting the 1980 marriage distribution.

6.6 Including education as an additional type

The over prediction of the number of marriages in 1980 may be attributed to true changes in preferences which our approach do not model. Alternatively, it may also be partly due to model mispecification in the form of ignored individual heterogeneity. In reality, individuals match in more dimensions than just age. Educational attainment of the individuals would be another dimension in which we see positive assortative matching. We expect the returns to marriage to differ across matches of differing education as well as age. Expanding the model to include educational attainment will allow responses in marital behavior to differ across age as well as education attainment (as population vectors change).

Figure 6a and 6c show the ages of marriage in 1970 by college attainment for men and women respectively. For both sexes, the marriage distribution for college educated individuals are shifted to the right relative to the non-college educated individuals. The fact that college educated individuals on average delay marriage relative to the non-college educated group suggest that there is significant difference in age preference for marriage across these two groups. Including educational attainment as an additional type will account for this heterogeneity in preferences.

Figure 6b show that, for every age past 25, the proportion of available males with college education is higher in 1980 than in 1970. The increase in the proportions for female began at age 22 as shown in Figure 6d. Since there are proportionately more available college educated young adults in 1980 and college adults delay marriage, the change in college attainment may help decrease the over prediction in our first pass model especially amongst the young adults.

We expanded the type space of our model to include three levels of educational attainment: (1) less than high school, (2) high school graduate, and (3) college graduate and beyond. Thus our type space for an individual is now characterized by his or her age and educational attainment. Let k, l index the education attainment of the male and female individuals respectively. The same notational convention carries through, for example, μ_{i0k0} denotes the male population of age i and education level k that remained unmarried, and m_{ik} denotes the available males of age i and education level k. Our marriage matching function for this expanded type space is given by ,

$$\mu^t_{ijkl} = \Pi_{ijkl}\sqrt{(m^t_{ik} - \sum_h \sum_{g=1}^{J} \mu^t_{igkh}) \cdot (f^t_{jl} - \sum_p \sum_{s=1}^{I} \mu^t_{sjpl})}. \tag{25}$$

We estimate the preference parameters by age and education, denoted by $\widetilde{\Pi}^{70}_{ijkl}$. Using this and the 1980 population vectors, adjusted by education, we predict another marriage distribution for 1980, denoted by $\widetilde{\mu}^{80}_{ij} = \sum_k \sum_l \widetilde{\mu}^{80}_{ijkl}$.

The marriage data from the *Vital Statistics* do not record information on the education attainment of married couples while information on educational attainment and marital status are available in the US Censuses. To overcome the data constraint of the *Vital Statistics*, data from the US Censuses is used to obtain at an estimate of the number of marriages by the level of education attainment. Details of the estimation and prediction procedures are outlined in Appendix C.

Figure 7a graphs the predicted marriage distribution by age, $\widetilde{\mu}^{80}$. Compared with Figure 5b, $\widetilde{\mu}^{80}$ do not predict as many marriages among young adults as when we ignored educational attainment. However compared with the observed marriage distribution graphed in Figure 5a, we still over predict the number of marriages by young adults even after controlling for educational attainment. Figure 7b plots the prediction error for $\widetilde{\mu}^{80}$. Compared with the previous prediction errors graphed in Figure 5c, the education adjusted model does better.

Figure 7c shows the actual and predicted marriage distributions for marriages between adults of the same age. This slice of the distribution is informative because there are relatively many marriages between adults of the same age due to positive assortative matching by age. The observed number of marriages among young adults in 1980 fell substantially compared with 1970. Figure 7c further emphasize that including educational attainment as a type improves the predicted number of marriages among young adults in 1980. The plots also suggest that the predicted number of marriages for adults over age 25 is close to the observed in 1980, and the largest discrepancy between the observed and predicted is among young adults below the age of 25. The double peak in the marriage distributions is due to the difference in modal ages of marriage between college and non-college educated individuals.

The next set of graphs provide a decomposition of where the improvements in predictions come from. Figure 8 shows the observed marriage distributions in 1970 and 1980, and the predicted distributions in 1980 before and after accounting for education attainment. These distributions are graphed by gender, educational attainment and age.

Consider Figures 8a and 8d that plot the marriage distributions of males and females respectively, with less than high school education. Comparing the observed distribution in 1970 and 1980, we find a substantial fall in the number of marriages over this decade. This fall is observed over all ages. The difference between the predicted distributions before and after accounting for education attainment quantifies the improvement in prediction attributable to accounting for heterogeneity in preferences across types. This improvement suggest that the the marriage behavior of indivduals in this education group is substantially different from the population as a whole.

Even after accounting for the difference in marriage behavior for this education group, our predictions still predict an increase in the number of observed marriages from 1970, contrary to the observed distribution in 1980. This difference between the observed and predicted distribution after accounting for individual heterogeneity suggest a significant change in preferences for marriage within this less than high school group. In Figure 8d, we observe that the prediction ignoring education attainment for females also incorrectly predict the modal age of marriage.

A similar story arise when we compare the distributions for individuals with up to high school education in Figures 8b and 8e. Unlike the previous group, the decrease in the observed number of marriages occur only for young adults below the age of 23, while an increase occur amongs adults age 25 and older. The significant over prediction after accounting for education attainment suggest again that there is significant change in marriage behavior of this education group.

Consider Figure 8c which plots the distributions for college educated males. We observe a shift to the right in the actual distribution over the decade. The predicted distribution in 1980 over predicts the number of young married males before the age of 30, and under predicts the distribution of male marriages over age 30. The same is true for college educated females as shown in Figure 8f. One way to think about our pattern of forecast errors is that we underpredict the delay in marriage among college educated individuals.

The empirical results suggest that there is substantial changes in the systematic gains to marriage between 1970 and 1980 particularly among individuals with less than high school and up to high school education. To study these changes in preferences, we estimate $\widetilde{\pi}^{80}$ and compute $\Delta\widetilde{\pi} = \widetilde{\pi}^{80} - \widetilde{\pi}^{70}$. Figures 9a-i plot $\Delta\widetilde{\pi}$ for different educational pairings for couples who marry at the same age. The scale are kept to be the same in all the figures and the zero change line is drawn in. Figure 9a plots $\Delta\widetilde{\pi}$ for less than high school men who married less than high school women of the same age. The plot shows that $\Delta\widetilde{\pi}$ is negative at all ages which implies that the systematic gain to marriage has fallen substantially for this group over the decade. Figure 9e shows that $\Delta\widetilde{\pi}$ is essentially also negative for high school graduate males who married high school graduate females of the same age up until age 35. After age 35, $\Delta\widetilde{\pi}$ turns positive. So the gains for marriage fell for young adults and increased for older adults in this educational group. For college graduates, Figure 8i shows that $\Delta\widetilde{\pi}$ is negative at the early ages and then turn positive after age 26 or so. One interpretation of this pattern is that the gains to delaying marriage for college graduates in 1980 rose relative to their 1970 counterparts. The gains to delaying is not unique to college graduates. This pattern is also observed between for couples with high school degrees or more education (Figures 8e,f,h,i).

Demographers, for example Qian, have already noticed the fall in marriage rate among young adults. There were substantial changes in available reproductive technologies over the decade. Birth control pills became widely available to single women (Goldin and Katz 2002). Legal abortion also became available. The effects of these changes led to declines in the systematic gains to marriage (Akerlof, et. al. 1996, Goldin and Katz 2002, Siow

2002). These changes will show up in our model as a change in $\pi's$ over the decade. However it is beyond the scope of this paper to quantify the effect of the changes in reproductive technologies on the changes in $\pi's$.

Finally, Figures 9a-i do not show much changes in the systematic gains for assortative matching by education. Although Figures 9c and 9g show that $\Delta\widetilde{\pi}$ has fallen substantially for matches between college graduates and less than high school graduates, the fall in $\Delta\widetilde{\pi}$ is not less than that for own matches by less than high school graduates. Likewise, there is no evidence that $\Delta\widetilde{\pi}$ has fallen more for other mix educational matches compared with own matches by the lower educational group. This conclusion contrasts with studies of assortative matching by education which do not clearly separate between changes in preferences and population vectors (E.g. Fernandez, Guner and Knowles 2001[26]).

7 Conclusion

This paper proposed a statistic to empirically characterize marital distributions. We used the associated marriage matching function to characterize the change in the marriage distributions between 1970 and 1980.

We discuss four avenues for future research. Three are methodological whereas the last is substantive.

First, we considered a particular specification of the demand for marriage partners. Following the large literature on discrete choice model, there is room for alternative specifications of the demand for marriage partners.

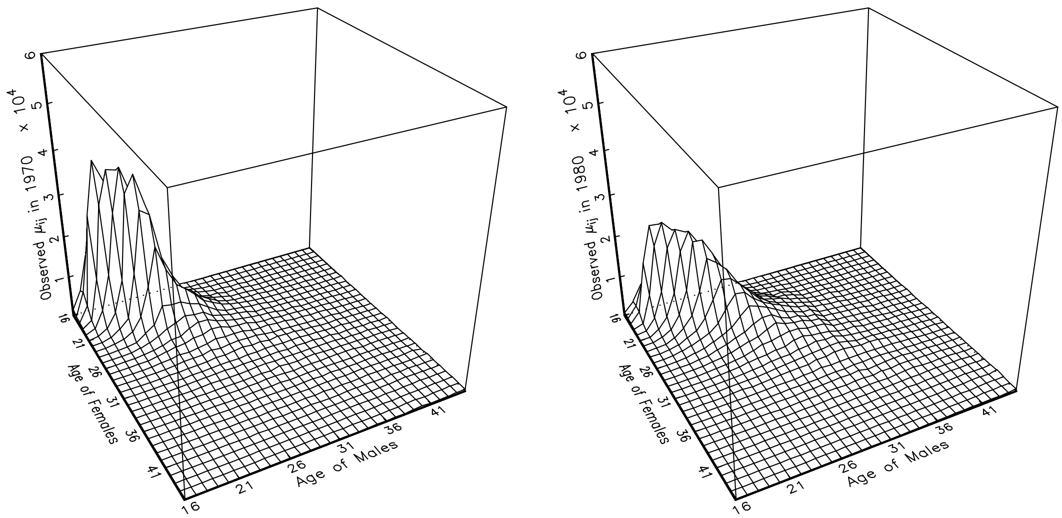
[26]Their study provides a very nice model of educational assortative matching.

Second, we considered a static model of the marriage market. A dynamic transferable utility model of the marriage market is needed. An issue that arises in modelling dynamic marriage market models with search frictions is the specification of the matching technology. To date, economists have primarily considered random matching technology (E.g. Ayagari, et. al., Seitz 1999, Hamilton and Siow 2000). Our marriage matching function provides a methodology for generating non-random matching technologies.

Third, 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 empirical analysis is because we cannot calculate the number of new entrants into cohabitation in any year from census data.

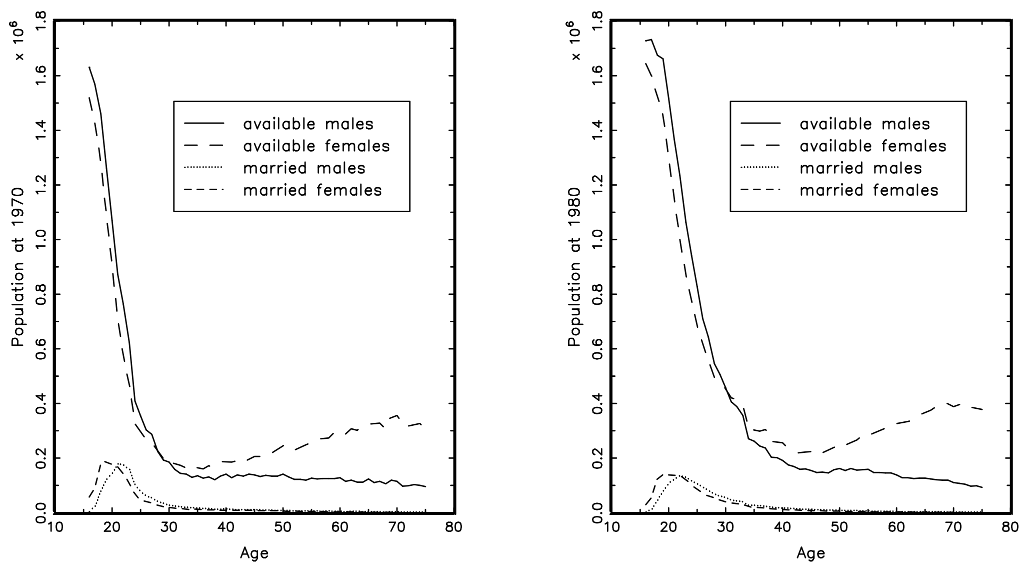
Finally, this paper provides clear evidence of a drop in the gains to marriage for young adults in all education groups between 1970 and 1980. The reason for this change and whether this trend continued into the nineties and later remain to be investigated.

Figure 1



a) Surface of observed μ_{ij} for 1970 b) Surface of observed μ_{ij} for 1980

Figure 2



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

Figure 3

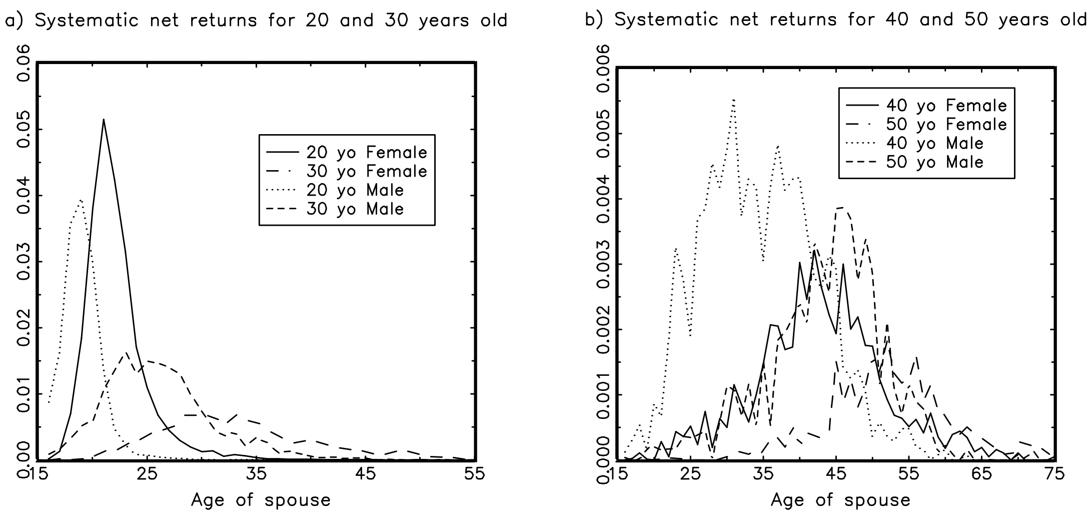
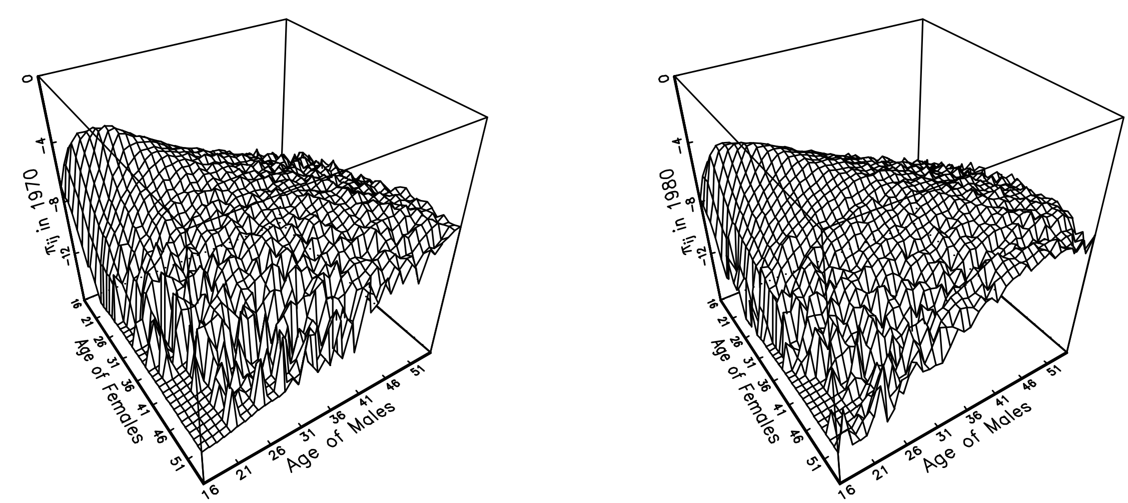
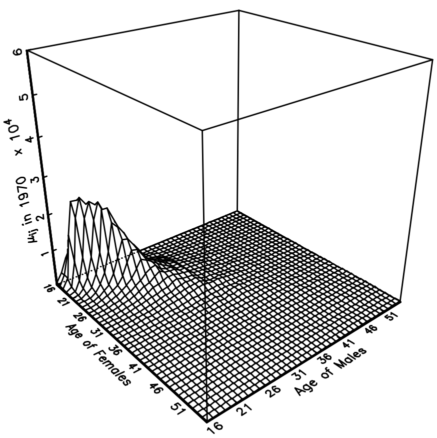


Figure 4

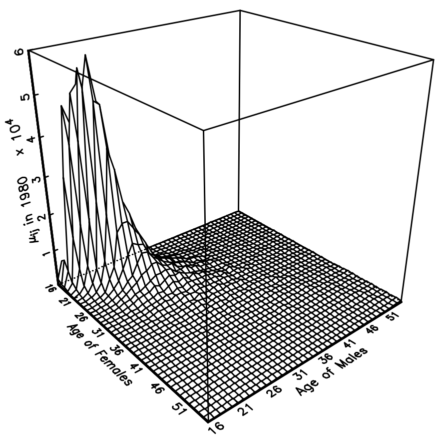


a) Observed π_{ij} at 1970 b) Observed π_{ij} at 1980

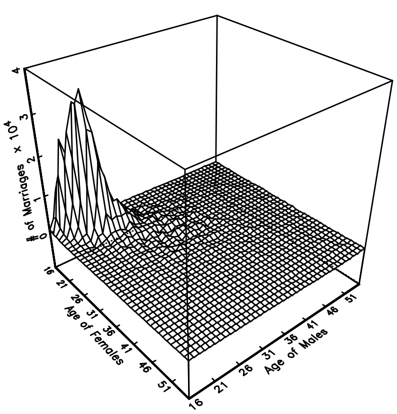
FIGURE 5



a) Observed μ_{ij} at 1980

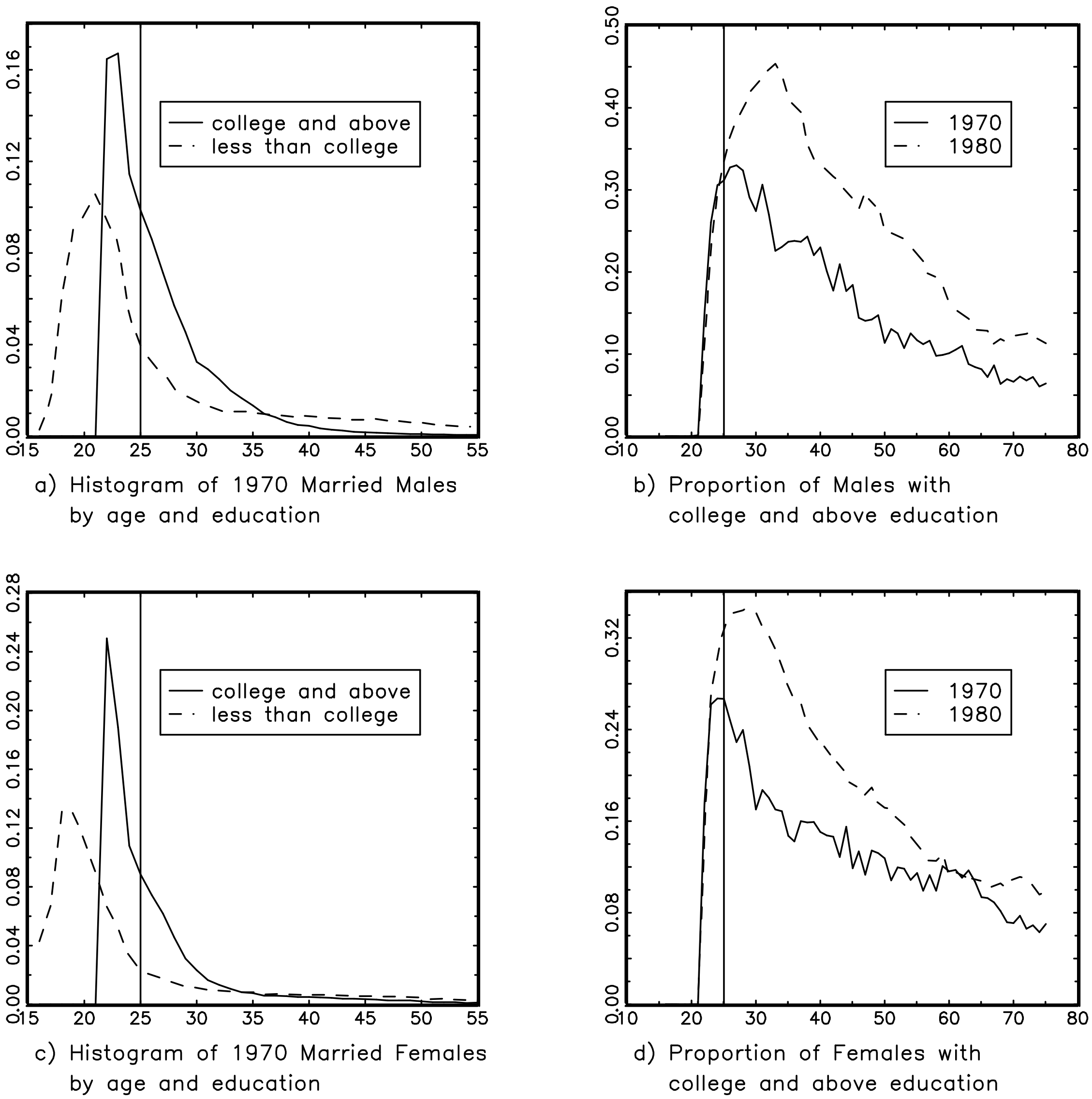


b) Predicted μ_{ij} at 1980
ignoring education



c) Residuals from prediction
without accounting for education

FIGURE 6



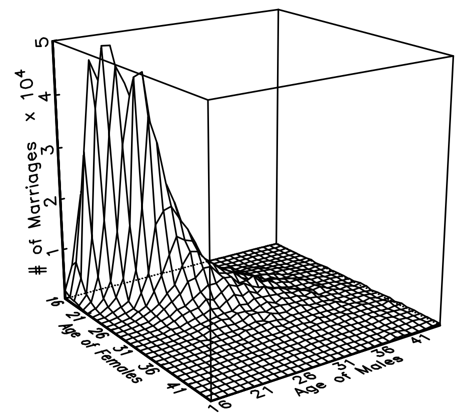
a) Histogram of 1970 Married Males by age and education

b) Proportion of Males with college and above education

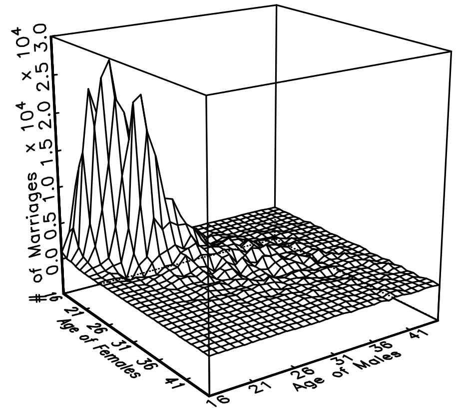
c) Histogram of 1970 Married Females by age and education

d) Proportion of Females with college and above education

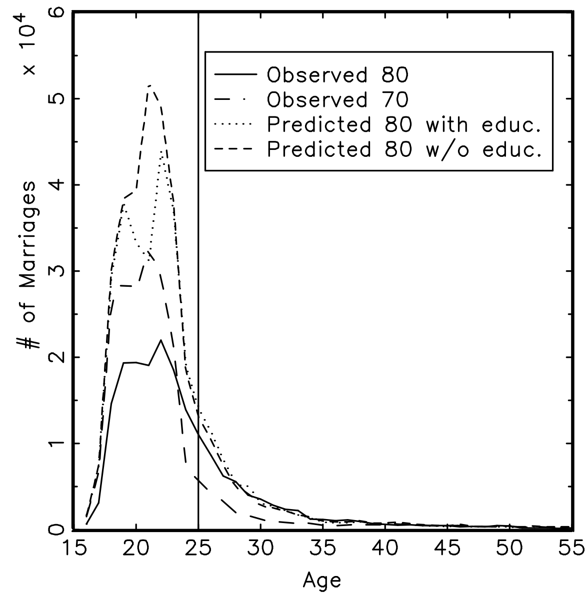
FIGURE 7



a) Predicted μ_{ij} at 1980
accounting for education

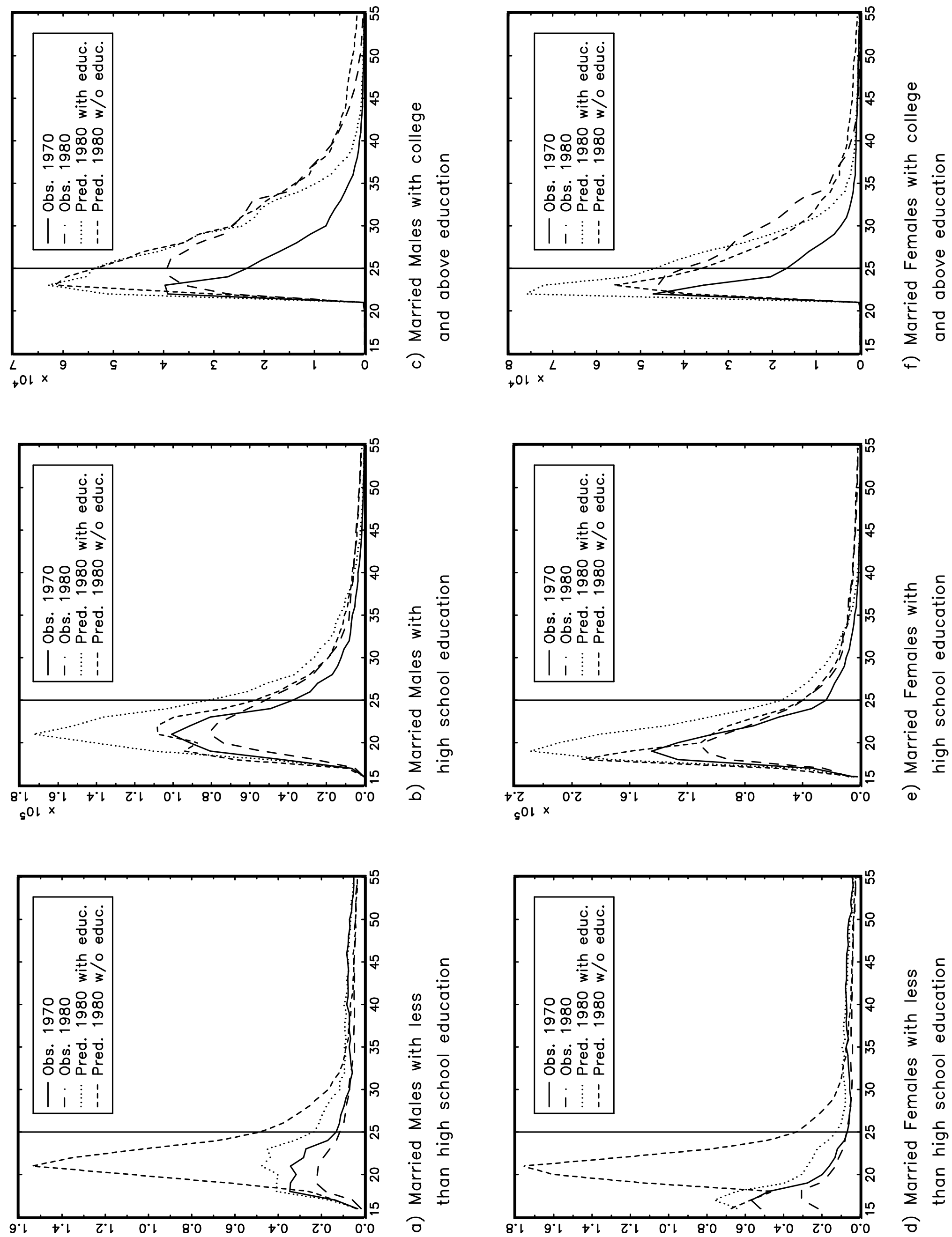


b) Residuals from prediction
accounting for education



c) Marriages between same
age individuals

Figure 8



a) Married Males with less than high school education

b) Married Males with high school education

c) Married Males with college and above education

d) Married Females with less than high school education

e) Married Females with high school education

f) Married Females with college and above education

FIGURE 9

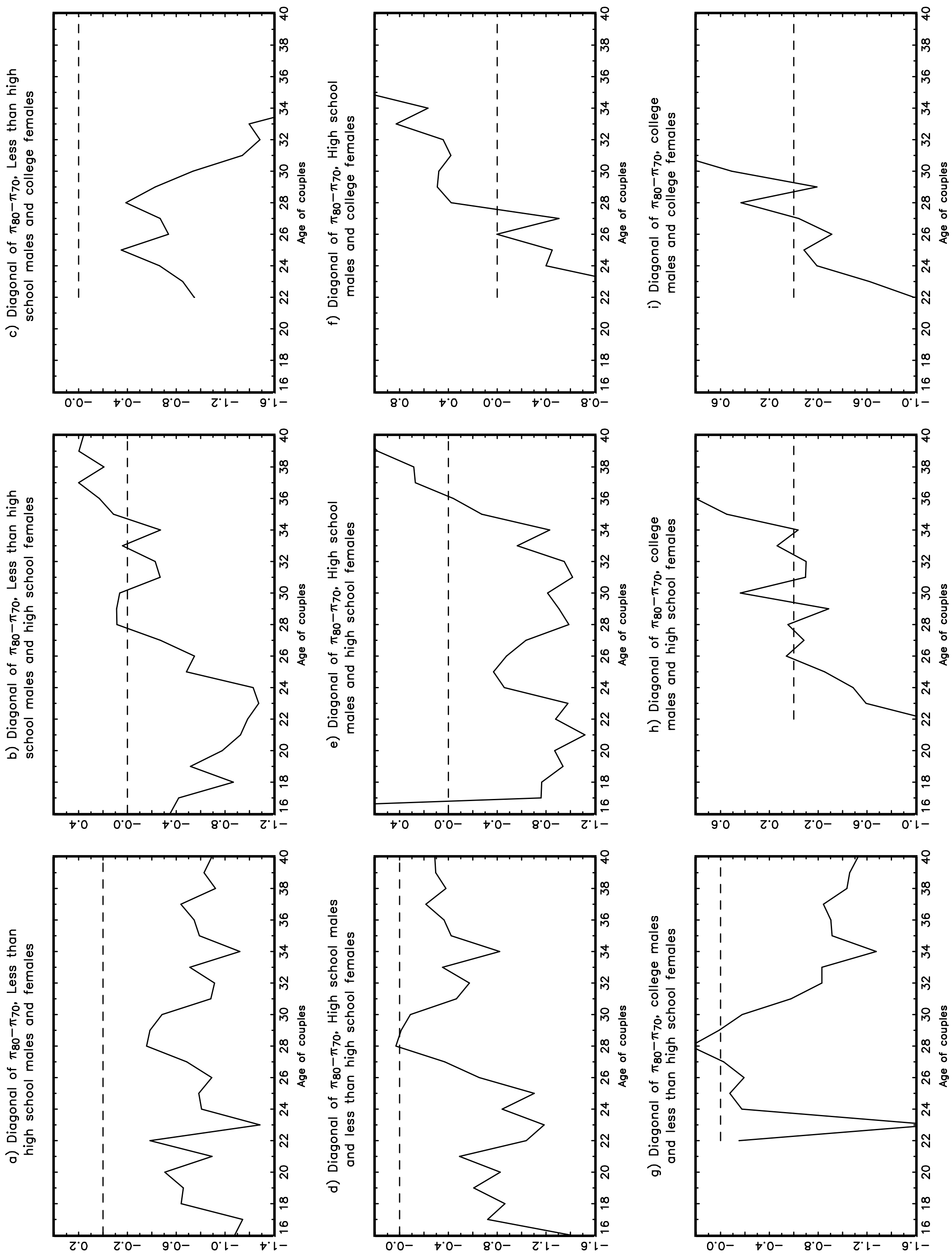
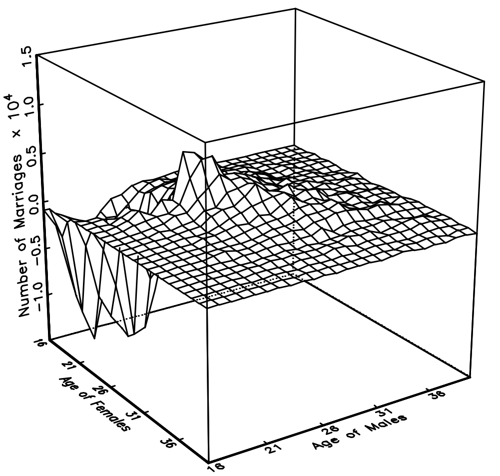
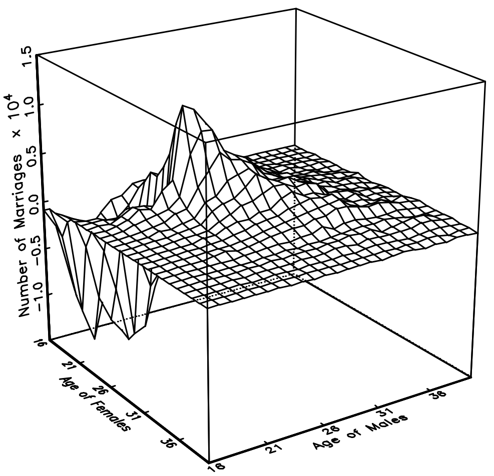

a) Diagonal of π80−π70, Less than high school males and females
b) Diagonal of π80−π70, Less than high school males and high school females
c) Diagonal of π80−π70, Less than high school males and college females
d) Diagonal of π80−π70, High school males and less than high school females
e) Diagonal of π80−π70, High school males and high school females
f) Diagonal of π80−π70, High school males and college females
g) Diagonal of π80−π70, college males and less than high school females
h) Diagonal of π80−π70, college males and high school females
i) Diagonal of π80−π70, college males and college females
Age of couples

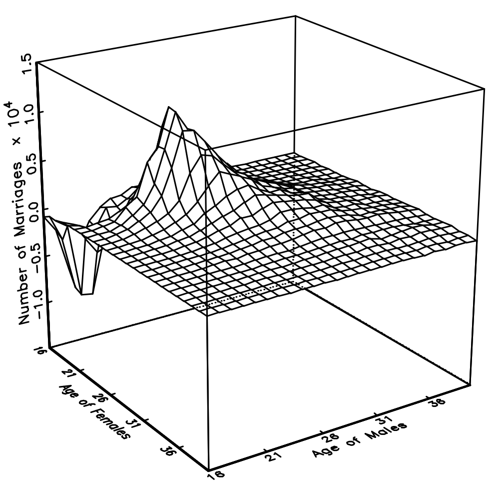
FIGURE 10



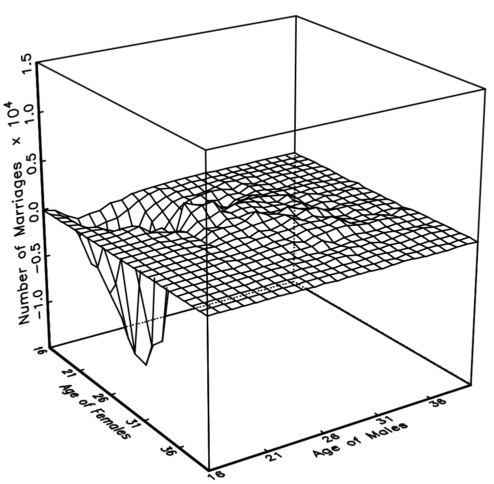
a) Changes in observed marriages from 1970 to 1980 i.e. $(\mu_{ij}^{80}-\mu_{ij}^{70})$



b) Changes in marriages accounted by the sum of the first and second order terms.



c) Changes in marriages accounted by the first order term, i.e. $\Delta\Theta' \cdot D\mu_{ij}(\Theta)$.



d) Changes in marriages accounted by the second order term, i.e. $1/2 \cdot \Delta\Theta' \cdot D^2\mu_{ij}(\Theta) \cdot \Delta\Theta$.

8 Appendix A: 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.[27]

To maintain consistency between states reporting marriages to the *Vital Statistics* and the data collected from the *US Census*, some states to be excluded. This result in the data from the following states being used: Alabama; Alaska; California 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.[28]

The age range studied was 16-75. 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,

[27]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.

[28]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.

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 more 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.

Since the census reports both family and individual records, married couples have to be linked up. The IPUMS data include a family serial number for family records and each married individual's record includes a spouse locator. We used these variables to convert the individual records of married individuals into married couples records. Some observations are lost in this process because they can't be linked up. The married couples were then split into four categories: both spouses in their first marriage; husband in first marriage and wife in second or later marriage; wife in first marriage and husband in second or later marriage; and both in second or later marriage. Once couples are identified, we could use the same technique of collapsing the data into counts of observation of each combination of husband and wife types.

To get an estimate of the joint distribution of age of marriage and education attainment, the following approach was adopted. We chose couples who were married three years or less as of the census year. This requires us to obtain the couple's date of marriage, which is not reported in IPUMS. Instead, we use the "`agemarr`" variable (age at first marriage) to calculate the length of marriage by subtracting age at marriage variable from age. This gives two versions of the length of marriage for each couple (one from each spouse). For couples who are both in their first marriage, the 2 versions should be identical. Observations in which there was more than a two year discrepancy were deleted. For individuals who are in their second or later marriage, this calculation is invalid and we cannot identify the length of their current marital status unless their spouse is in his/her first marriage. Thus for couples who are both in their 2nd or more marriage, we can only obtain an upper bound on the length of their marriage, not an accurate measure. For couples in which at least one spouse is in his/her first marriage, we delete all observations in which the length of the marriage is more than three years. For couples in which both spouses are in their second or more marriage, we delete all observations in which the upper bound on the length of the marriage is larger than nine years.

9 Appendix B

In order 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

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

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:

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

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. □

10 Appendix C

10.1 Dealing with thin cells

The primary objective of our second estimation approach is to provide an estimate of the marriages for cells that are thin. We approximate the empirical marriage distribution by age, denoted by $\mathbb{P}\{i, j\}$, using a bivariate lognormal distribution, where the probability of observing an i, j marriage conditional on parameter vector θ is given by,

$$\mathbb{P}\{i, j \mid \theta\} \approx \int_i^{i+1} \int_j^{j+1} g(x_1, x_2 \mid \theta)\, dx_1\, dx_2 \quad \text{where } i, j \in \{16, \ldots 75\}.$$

$g(x_1, x_2 \mid \theta)$ denotes the lognormal probability density function, that is, $\log(x_1, x_2)' \sim N(\nu', \Gamma\Gamma')$. The parameters of the distribution are given by the $\nu = (\nu_1, \nu_2)'$ and $\Gamma = (\Gamma_{11}\, \Gamma_{12},\, \Gamma_{21}\, 0)'$ and the parameter vector to be estimated is $\theta = (\nu_1, \nu_2, \Gamma_{11}, \Gamma_{12}, \Gamma_{21})'$. The estimated parameters $\widehat{\theta}$ minimizes the least squares criterion,

$$Q(\theta) = \sum_{i=1}^{I} \sum_{j=1}^{J} \Big(\mathbb{P}\{i, j \mid \theta\} - \mathbb{P}\{i, j\} \Big)^2.$$

The proposed extremum estimator is inefficient and more efficient alternatives are available. The main advantage of the proposed criterion is that it allows $\mathbb{P}\{i, j \mid \hat{\theta}\}$ to be zero for certain i, j pair, which more efficient alternatives like maximum likelihood or minimum chi-square estimator doesn't allow.

Let the total number of marriages at t be denoted by n, where $n = \sum_{i=1}^{I} \sum_{j=1}^{J} \mu_{ij}$, and the total number of marriages for i, j pairs that are less than some cutoff c be denoted by n_c, i.e., $n_c = \sum_{i=1}^{I} \sum_{j=1}^{J} \mathbb{I}\big\{\mu_{ij} < c\big\}\, \mu_{ij}$. The modified marriage distribution by age, denoted by $\widetilde{\mu}_{ij}(\hat{\theta})$ is defined by

$$\widetilde{\mu}_{ij}(\hat{\theta}) = \begin{cases} \mu_{ij} & \text{if } \mu_{ij} \geq c, \\ n_c \cdot \mathbb{P}\{i, j \mid \hat{\theta}\} & \text{if } \mu_{ij} < c. \end{cases}$$

The cutof c was set to be 33, approximate 0.002% of 1970 total marriages. Our estimation methodology would proceed as described in Section 6.3 with $\widetilde{\mu}_{ij}(\hat{\theta})$ in place of μ_{ij}.

10.2 Including education as type

The marriage data from the Vital Statistics do not record information on the education attainment of married couples. The following outlines the methodology used to obtain an estimate of the number of marriages by the level of education attainment using information from the US Census.

There are three levels of educational attainment: (1) less than high school, (2) high school graduate, and (3) college graduate and beyond, which we index by k and l for male and female individuals respectively. The Census provides us with a joint empirical marriage distribution by age and education attainment, which we denote by $\mathbb{P}\{\, i, j, k, l \,\}$. ¿From this we construct $\mathbb{P}\{\, k, l \,\}$, an empirical distribution of a couple with a k, l education pairing, and $\mathbb{P}\{i, j \mid k, l \,\}$, the marriage distribution by age conditional on an education pairing.

The problem of thin cells is even more pronounced when dealing with Census data. A methodology similar to that outlined in the previous section is adopted. We approximate the marriage distribution by age conditional on an education pairing k, l, using the lognormal distribution, denoted by $\mathbb{P}\big\{i, j \mid k, l, \widehat{\theta}_{kl}\big\}$, where $\widehat{\theta}_{kl} = argmin \sum_{i=1}^{I} \sum_{j=1}^{J} (\, \mathbb{P}\{i, j \mid k, l, \theta\} - \mathbb{P}\{i, j \mid k, l\}\,)^2$. Using the fitted distribution we construct a modified distribution, $\widetilde{\mathbb{P}}\big\{i, j \mid k, l, \widehat{\theta}_{kl}\big\}$, which is defined as,

$$\widetilde{\mathbb{P}}\big\{i, j \mid k, l, \widehat{\theta}_{kl}\big\} = \begin{cases} \mathbb{P}\big\{i, j \mid k, l, \widehat{\theta}_{kl}\big\} & \text{if } \mu_{ijkl} < c_{kl}, \\ \mathbb{P}\big\{i, j \mid k, l\big\} = \mu_{ijkl} / \sum_i \sum_j \mu_{ijkl} & \text{if } \mu_{ijkl} \geq c_{kl}. \end{cases}$$

c_{kl} defines a cutoff specific to the conditional distribution. Using the estimated distribution, we can construct an estimate of number of marriage by age and education attainment as follows,

$$\widetilde{\mu}_{ijkl} = \frac{\widetilde{\mathbb{P}}\{i,j \mid k,l,\widehat{\theta}_{kl}\} \cdot \mathbb{P}\{\,k,l\,\}}{\sum_k \sum_l \widetilde{\mathbb{P}}\{i,j \mid k,l,\widehat{\theta}_{kl}\} \cdot \mathbb{P}\{\,k,l\,\}} \cdot \widetilde{\mu}_{ij}(\hat{\theta})$$

We estimate the preference parameters, $\widetilde{\Pi}_{ijkl}$ using Equation (25) for 1970. Given the population vector by age and education for 1980, m_{ik} and f_{jl}, we numerically solve the following system of nonlinear equations for the predicted vector of unmarried by age and education,

$$m^t_{ik} - \mu^t_{i0k0} - \sum_l \sum_{j=1}^{J} \Pi_{ijkl} \sqrt{\mu^t_{i0k0} \times \mu^t_{0j0l}} = 0$$

$$f^t_{jl} - \mu^t_{0j0l} - \sum_k \sum_{i=1}^{I} \Pi_{ijkl} \sqrt{\mu^t_{i0k0} \times \mu^t_{0j0l}} = 0.$$

Estimation proceeds as described in section 6.3.

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, 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.

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

[4] Bayer, Patrick, Robert McMillan and Kim Rueben. "An Equilibrium Model of Sorting in an Urban Housing Market", University of Toronto manuscript available at "www.economics.utoronto.ca/mcmillan", 2000.

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

[6] Botticini, Maristella, and Aloysius Siow. "Why Dowries?" University of Toronto manuscript available at "www.economics.utoronto.ca/siow", 2000.

[7] 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.

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

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

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

[11] Ekeland, Ivar, James J. Heckman, and Lar Nesheim. "Identification and estimation of hedonic models", *Journal of Political Economy*, forthcoming.

[12] 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).

[13] 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.

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

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

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

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

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

[19] ____________ "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, 1990a.

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

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

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

[23] 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.

[24] Rosen, Sherwin. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82, (1974): 34-55.

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

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

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

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

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

[30] 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.