A Parametric Two-Sided Model of Marriage

John Allen Logan, Peter D. Hoff, and Michael A. Newton [1]

Working Paper no. 15

Center for Statistics and the Social Sciences

University of Washington

June 26, 2001

[1] John Allen Logan is Associate Professor, Department of Sociology, University of Wisconsin, Madison, WI 53706 (E-mail: logan@ssc.wisc.edu). Peter D. Hoff is Assistant Professor, Department of Statistics, University of Washington, Seattle, WA 98195 (E-mail: hoff@stat.washington.edu). Michael A. Newton is Associate Professor, Department of Statistics and Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI 53706 (E-mail: newton@biostat.wisc.edu).

Abstract

We propose a method to simultaneously estimate men's and women's preferences for marriage partners using data on the characteristics of married couples and single individuals. Underlying our method is a parametric version of a two-sided matching model of marriage considered in game theory. There are two matching populations, men and women. Members of each population have utilities for pairing with members of the other population, which results in a stable set of matches between the two populations. The estimated preference coefficients determine the degree to which measured characteristics of individuals affect choices of marital partners. We also discuss possible uses of such estimates for projections of formations and dissolutions of marriages. Our method should be useful in many situations in which voluntary pairings have arisen through some complex process whose details have not been recorded. Besides marriage and cohabitation, data on employment, college attendence, and coresidence of elderly parents with adult children often have this character, as do some biological data on non-human mating. Only cross-sectional information, rather than observations over time, is required. By assuming each individual chooses freely and knowledgeably from the set of potential partners he or she finds available, we estimate preferences without having either to observe these sets or to specify any details of the matching process. This makes our method robust to unknown features of the process. Unlike a two-sided logit model previously introduced for employment data, our method does not require categorization of alternatives on one side of the marriage market, but instead uses individual-level, quantitative characteristics symmetrically on both sides. We test our method on simulated marriage data and illustrate its use with marriage and cohabitation data from the 1993 Panel Study of Income Dynamics.

KEY WORDS: Two-sided matching; Multinomial choice; Two-sided probit; Two-sided logit; Data augmentation; Metropolis algorithm; Gibbs sampler.

1 Introduction

Social scientists are often interested in voluntary pair formation and its implications. Marriage is a prominent example. Researchers are interested in the mating process in order to predict the future characteristics of households in demographic projections and to understand the factors leading to the formation, continuation, and dissolution of marriages more generally.

The ability of particular men or women to find desirable marriage partners depends on the level of competition from other men or women, as well as on strategy and luck during the search for a mate. Both men and women exercise choice. An abstract, theoretical understanding of pair formation which takes into account such mutual choice is available in a branch of game theory called two-sided matching (Roth and Sotomayor 1990). Two-sided models are distinguished from certain one-sided models in which one or the other gender plays a completely passive role. Two-sided matching models of marriage are made up of idealized men and women who have fixed, relative preferences for mating with different members of the opposite sex. Using a small set of axioms, a large number of theorems about what types of marriages should arise have been proved. One important attraction of the theory, especially for non-economists, is that it need not assume forward-looking rationality (i.e., rational expectations formation) on the part of actors, but can also accommodate a more realistic view of individuals whose behavior is importantly shaped by social factors, rules of thumb, and lack of foresight. A second attraction is that constraints on individuals are represented without resorting to supra-individual concepts (cf. Coleman 1990). Though this theory of two-sided matching has been used informally as a standard of comparison for real marriage systems, it has so far not been the basis of a statistical procedure that can be used when analyzing typical marriage data sets. Nor has it been of any practical use in making projections of household formation in real populations, since no means of estimating preferences in such populations has been developed.

We propose a statistical estimation procedure based on a two-sided matching model of marriage. We will estimate the preferences of men and women for each others' characteristics using the most commonly-available type of data on marriage. This type of data contains selected marriage-relevant characteristics of both single people and married partners, such as education and age. Unfortunately, the data sets commonly available for large populations contain no information either on the processes that led to the formation of the observed marriages or on the sets of alternative partners from which particular actors could have chosen mates. Either kind of information would greatly simplify estimation. In particular, if the choice sets of alternative partners available to each actor were known, standard, one-sided conditional logit or probit models could be used to infer the preferences of both men and women. But without knowledge of these choice sets the application of the standard models to the commonly-available type of data produces biased preference estimates (e.g., Ben-Akiva and Lerman 1995; for a demonstration see Logan 1996a). Obtaining data on the actual sets of available partners for each actor in a large data set seems impracticable, as is attested by its absence from all such data sets known to us. Our method allows us to estimate preferences without observing the choice sets, making it useful for data sets that contain information only on the characteristics of matched and single individuals.

Besides estimating the preferences of men and women from available data on the characteristics of married couples and single people, we also intend eventually to make predictions of marriages that would be observed under different conditions, assuming preferences remained constant. For example, we intend to project the distributions of marriages that would occur if one or the other sex decreased or increased in numbers (due perhaps to war or migration) or changed in its distribution of marriage-relevant characteristics (perhaps because of changes in employment rates or levels of education). Such predictions could be used for demographic projections as well as for validating particular models.

Some clarification of the distinctive approach represented in our statistical procedures is needed. Figure 1 presents scatterplots of married couples' ages and educations from a typical data set that we will analyze in section 6 of this paper. Each plotted point in the interior of the top panel shows the husband's and wife's ages in a particular married couple (points have been jittered to make multiple observations with identical values discernable). The bottom panel shows husbands' and wives' years of education. Along the axes of the two scatterplots we have also plotted the ages and educations of the unmarried men and women in the data set. Previous analyses of marriage have focused on the joint distributions of characteristics of husbands and wives, data corresponding to the interior points in the scatterplots of Figure 1, omitting the unmarried persons plotted along the axes. Standard regression models have been used to estimate the distributions of husbands' characteristics conditional on wives', and vice versa. Such regressions are of limited use for the goals we have established. The regression coefficients cannot differentiate the effects of wives' and husbands' preferences for spousal characteristics from the effects of the distributions of desired characteristics in the population. Neither can they distinguish the relative contributions of wives' versus husbands' preferences in the formation of marriages. Because the coefficients represent only net effects of wives' and husbands' preferences under a particular distribution of partners' characteristics, it is not possible to use them to make plausible predictions of the effects of changes in any of the components. Our inferential method will allow us to do all of these things.

Considering a population of men and women with fixed characteristics, scatterplots like those in Figure 1 suggest an interpretation different from those underlying regression analyses. In our interpretation the data do not represent draws from a joint probability distribution of men's and

Figure 1: PSID Data

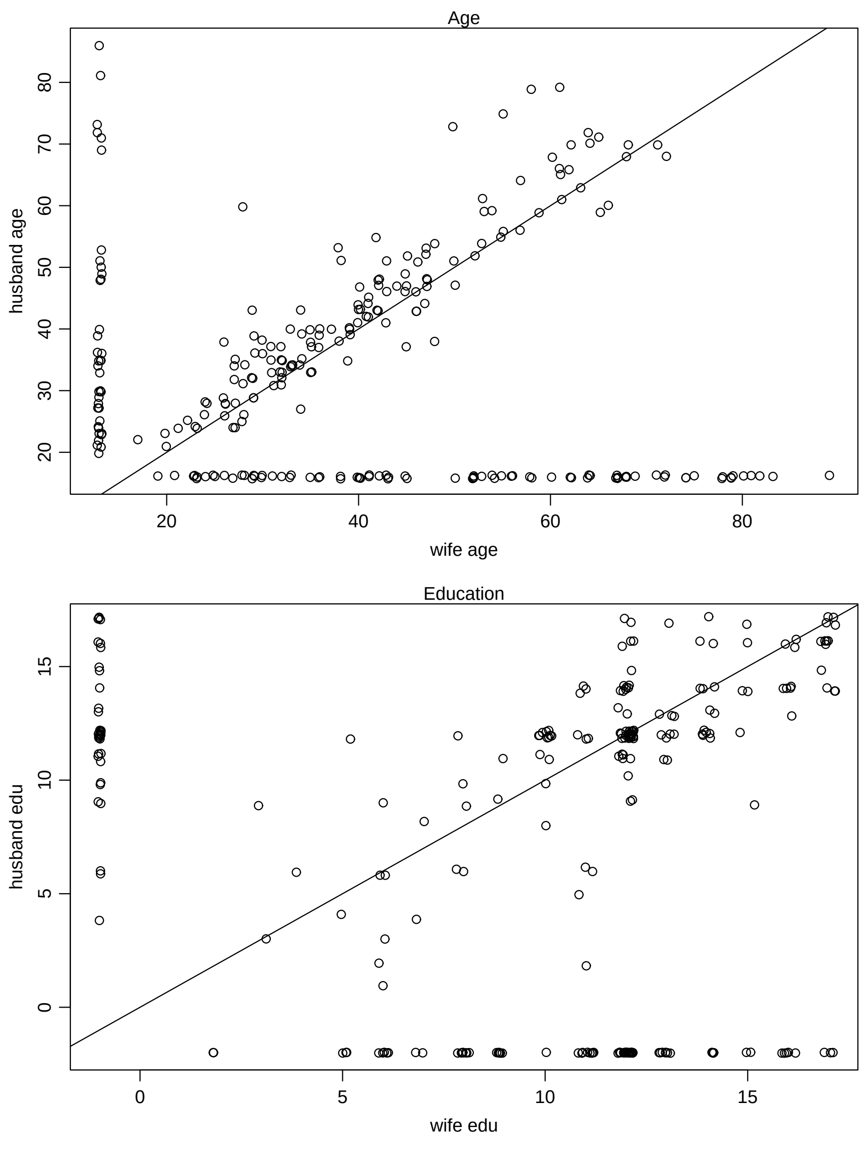
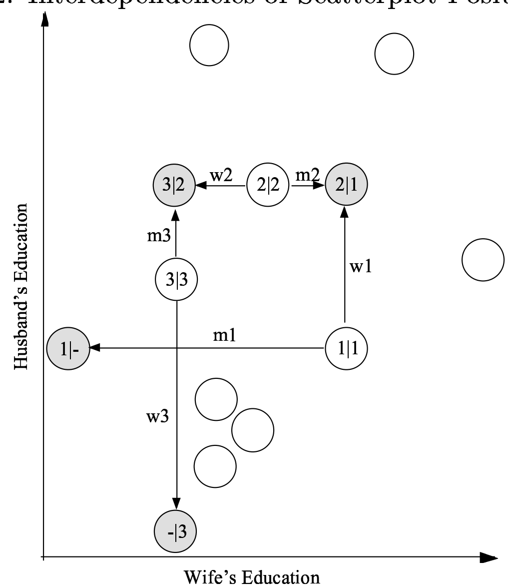


Figure 2: Interdependencies of Scatterplot Positions



women's characteristics with fixed mean and covariance, nor conditional draws from one sex's characteristics given the other's. Instead, the data in the scatterplots–here including the characteristics of the unmarried individuals along the axes–satisfy fixed marginals on both axes. These fixed marginals are the pre-existing age or education distributions of the men and women making up the population. The reasonable requirement that the marginal distributions of characteristics should not be changed by the matching process means the plotted points are not independent observations. Figure 2 illustrates this implicit interdependence using a small, hypothetical scatterplot of men's and women's education levels. The plotted circles labeled 1 | 1, 2 | 2, and 3 | 3 represent marriages of men 1 through 3 with women 1 through 3, respectively. Any different arrangment of points in the scatterplot would entail coordinated shifts among men and women. For example, if woman 1 were to be matched with man 2 instead of man 1, a new point would form in the location that preserves their two levels of education, as indicated by the shaded circle marked 2 | 1. The circles labeled 1 | 1 and 2 | 2 would disappear. This matching of woman 1 with man 2 would leave man 1 and woman 2 unattached. The figure shows one possible resolution of this situation in which woman 2 joins with man 3 to form circle 3 | 2, and man 1 and woman 3 move to the distributions of unattached individuals along the axes, indicated by circles 1 | − and − | 3. None of the three original labeled circles remains, while the four new circles preserve the appropriate education levels of their men and women. The set of possible scatterplots is evidently finite and systematically constrained, once the marginal distributions of characteristics are given. Our statistical analysis will take the marginal distributions of quantitative characteristics of men and women as fixed, and will take the observed joint distribution of characteristics, given the marginals, to reflect the free mutual choices of men and women according to their preferences for those characteristics. This is a new way of looking at scatterplots of matching data. Rather than considering the data points to be independent draws from suitable distributions, as is done in standard regressions, we view them as highly interdependent observations and base our estimations on a model of the interdependence.

The key to our analysis of typical marriage data is the recognition that the data contain no direct information on the processes that lead to marriage, but only on the resulting state, the observed marriages. We therefore model the state rather than the process. On the assumptions that marital partners remain together voluntarily, and are knowledgeable about possible alternatives, we will estimate the preferences that their unions reveal. These preferences could also have influenced the unobserved and unmodeled processes by which the marriages formed. The strategy of modeling the state of observed matches rather than the process of match formation was also used in an employer-employee context by Logan (1996b, 1998), who recognized the implied assumption of stability (in a special sense) that we will introduce below (Logan 1996c). Logan called his statistical model two-sided logit, since the choices made on either side of the market follow a standard logit model once the preferences of the other side are taken into account, and in recognition of the close connection to the two-sided matching models of game theory. Logan's method required that one side of the market be reduced to categories of actors for computational reasons, while our procedure will treat all actors as individuals. The models that we consider include two-sided probit and logit models that represent an extension of Logan's approach to completely individual-level data, as well as other models.

Potential applications of our estimation method include any situations in which voluntary pairings of actors are observed but detailed information on the process leading to the pairings is not. In the employment context, the method would allow detailed job characteristics to replace the general occupational characteristics used by Logan (1996b). The mutual choices of colleges and would-be students may be studied in more detail than has been possible with the earlier method. More general household formation questions, such as the decisions of children and their aged parents to co-reside, are also possible applications. Finally, the method may have application in non-human mating populations when only data on observed matches are available (see Bergstrom and Real [2000] on game-theoretic modeling of non-human mate choice). In all potential applications, combining our preference estimates with specific behavioral assumptions will make possible counterfactual predictions analogous to the demographic household projections mentioned earlier.

2 The Marriage Model and Stability

As described by Roth and Sotomayor (1990), a marriage market involves a population of men and women in which each individual has a preference ordering over the members of the opposite sex. In this paper, we assume the preference orderings are derived from subjective utilities that individuals have for forming pairs with one another and for being single: Man i's utility for pairing with woman j is denoted as $U_{i,j}$, and man i's utility for being single is $U_{i,0}$. Similarly, $V_{j,i}$ and $V_{j,0}$ denote woman j's utilities for man i and for being single, respectively.

In the context of the marriage market, a matching m is a set of pairs of men and women, such that each individual is a member of at most one pair. A matching m determines a husband function h, where $h(j)$ gives the index of the man married to woman j, and $h(j) = 0$ if woman j is single. Similarly, the matching can be written in terms of a wife function w, mapping the indices of the men to those of the women.

We assume each actor in the system chooses a partner or remains single so as to maximize his or her utility over the set of available opportunities: Each man forms a pair with the highest-utility woman among those available to him, or remains single if his utility for singleness exceeds that for his highest-rated available woman. Women do the same, with the roles reversed. Of course, not every actor has the same opportunities: The set of available partners varies from actor to actor, and depends in a complicated way on the utilities of the other actors in the system. For a given set of utilities and a given matching, the available pairing states for man i and woman j are given by their opportunity (or choice) sets

$$\begin{aligned} O(i) &= \{j : V_{j,i} \geq V_{j,h(j)}\} \cup \{0\} \\ O(j) &= \{i : U_{i,j} \geq U_{i,w(i)}\} \cup \{0\}, \end{aligned}$$

where the state of being single (denoted "0") is available to each actor. Thus each man has the opportunity to be single or to pair with any woman who prefers him to her match. Opportunities of the women are similar, with the roles reversed. We assume there are no ties among the utilities, so that preferences are strict.

The assumption that the matches are voluntarily maintained helps motivate the further assumption that an observed matching is stable in a game-theoretic sense, that is, the matching is such that each actor in the system prefers his or her state over all states available. Using the above notation, stability implies

$$U_{i,w(i)} \geq U_{i,j} \ \forall j \in O(i) \tag{1}$$

$$V_{j,h(j)} \geq V_{j,i} \ \forall i \in O(j).$$

In other words, in a stable matching no individual can make a move that will improve his or her situation. Stable matchings are of central interest in the application of two-sided matching models. Since a matching that is not stable could be improved on by the voluntary actions of system members, it is reasonable to expect that real systems will approximate stable matchings. In addition it is known that there is at least one (and typically many) stable match for any given set of utilities, and that a random flow of information about potential partners will lead to a stable matching with probability one in the long run (Roth and Vande Vate 1990). Arguments such as these make the analysis of stable matchings relevant to understanding real systems, even when stability cannot directly be confirmed.

The stability of the matchings should be thought of as inherently transitory, subject to change as persons enter or leave the system, or as their characteristics or preferences change over time through aging processes or for other reasons. The meaning is therefore quite different from what is meant by a stable marriage in ordinary conversation. Our assumption is that system members are generally knowledgeable about the characteristics of other actors, and that their observed matches are entered and maintained voluntarily. Though existing matches may for legal, child custody, or other reasons entail large potential costs if they were to be broken, this does not argue against the voluntary nature of the relationships nor suggest that an unstable matching is somehow enforced against the wills of the partners, when all relevant factors are considered. Stability in the sense adopted here means mutually voluntarily maintained, rather than unchanging as circumstances change.

The marriage model is composed of individuals but has a systemic quality. Each actor observes the set of partners available to him or her, and then chooses the one that offers the highest utility. The choices of actors on either side of the market are constrained by the preferences of members of the opposite sex, but also by the willingness (or not) of more desirable members of their own sex to match with the partners they themselves would prefer. Both men's and women's preferences weight the actual characteristics of possible partners, meaning that opportunity is in part a function of the distribution of these characteristics and not just of preferences as such. The marriage model is therefore a system model of the interwoven effects of the preferences and characteristics of all men and women.

3 Inference

Given a stable matching for a population of men and women, our goal is to estimate the extent to which certain observed characteristics of an individual influence the unobserved utilities that other actors have for pairing with that individual. We assume that a set of unknown preference parameters θ along with observed female characteristics X and male characteristics Y give rise to utilities U, V that men and women have for each other, and that the utilities in turn determine preference rankings that men and women have for the members of the opposite sex. The particular rankings vary across the rankers due to unobserved components of utility. Individuals form pairs with one another according to their individual preference rankings of those particular members of the opposite sex available to each of them. This in turn gives rise to the data we observe.

Under the assumption that the system has reached stability, the data tell us of the occurence of two events:

- $M = m$: The matching m was realized by the system of actors.

- $m \in \mathcal{S}$: The matching m is in the set $\mathcal{S}$ of stable matchings for the system.

As mentioned earlier, game theorists have shown that there is at least one stable matching, and possibly more, for every set of utilities. These observations motivate the following form for a full probability model of the data, given a set of preference parameters:

$$p(M = m, m \in \mathcal{S}|\theta) = p(M = m|m \in \mathcal{S}, \theta)p(m \in \mathcal{S}|\theta)$$

The first component of the probability, $p(M = m|m \in \mathcal{S}, \theta)$, depends on the particular mechanism or process by which actors formed pairs with one another, which is likely to have involved some strategic behavior that was not a simple reflection of their preferences (see Roth and Sotomayor 1990). Without knowledge of this mechanism, specification of the first component is difficult. However, the difficulty can be avoided if we base our inference only the second component, the marginal probability of the stability of the observed matching. Stability itself is a function only of the preferences of the actors (Roth and Sotomayor 1990), so its maintenance does not depend on strategic considerations. Although some information may be lost by basing inference only on the marginal likelihood, we emphasize that as a result, our method of estimation does not depend on the particular proposal mechanism that generates the matching.

Explicitly representing the unobserved utilities in the marginal probability of stability gives

$$p(m \in \mathcal{S}|\theta) = \int p(m \in \mathcal{S}|U, V, \theta)p(U, V|\theta)\ dUdV, \tag{2}$$

where dependence on the characteristics X, Y remains implicit. Note that the probability that matching m is stable is either 0 or 1, given U, V, so that the first term in the integral is simply a zero-one indicator of the event that the utilities lie in the constrained space required by the stability of the observed matching. In this paper, we focus primarily on models for $p(U, V|\theta)$ that are linear in the preference parameters (although more general parametric models are discussed in Section

4.3). For example, we might model man i's and woman j's utilities for each other and for being single as

$$
\begin{aligned}
U_{i,j} &= \alpha' x_j + \epsilon_{i,j}, \qquad & U_{i,0} &= a' x_{i,0} + \epsilon_{i,0}, \\
V_{j,i} &= \beta' y_i + \gamma_{j,i}, \qquad & V_{j,0} &= b' y_{j,0} + \gamma_{j,0},
\end{aligned} \tag{3}
$$

where α is a vector of shared male preferences for the characteristics x of women, a is a vector of shared male preferences for characteristics x_0 of being single, and β, b, y, and y_0 are defined similarly. The error terms ϵ, γ in the linear models can be thought of as individual deviations from mean utilities. A convenient model for error is that the ϵ's and γ's are i.i.d. from a known distribution centered around zero. With the parametrization above, the estimand of interest is the set of preference parameters $\theta = (\alpha, \beta, a, b)$.

4 Estimation Methods

Our goal is to estimate the marital preferences θ using the information that the observed matching is stable. Maximum likelihood estimation amounts to finding those values of θ which maximize (2). The maximum likelihood estimation methods of Logan (1998) would be applicable to the current problem with minor alterations, except the estimation time associated with those methods makes them impractical here. The difficulty lies in the calculation of $p(m \in \mathcal{S}|\theta)$, which is a complicated integral over a high-dimensional space.

Instead of trying to maximize (2) in θ, we take a Bayesian approach to the estimation: Our inference is based on the posterior distribution $\pi(\theta|m \in \mathcal{S}) \propto \pi(\theta) p(m \in \mathcal{S}|\theta)$, where $\pi(\theta)$ is a probability distribution representing our prior uncertainty about θ. Although the posterior involves the same intractable integral $p(m \in \mathcal{S}|\theta)$, Bayesian inference is made feasible by approximating the posterior via samples of θ from a Markov chain.

4.1 MCMC

In contrast to many applications, in our model the contributions from different individuals to the likelihood $p(m \in \mathcal{S}|\theta)$ are not multiplicative. Constraints induced by the assumption of stability connect the matches in a complicated way: Given the values of U, V, the probability of stability is the indicator function

$$p(m \in \mathcal{S}|U,V) = \begin{cases} 1 \quad \text{if} & V_{j,h(j)} \geq V_{j,i} \ \forall i \in O(j), \forall j \text{ and} \\ & U_{i,w(i)} \geq U_{i,j} \ \forall j \in O(i), \forall i, \\ 0 \quad \text{else.} & \end{cases}$$

where h and w are the husband and wife functions determined by the matching m. As none of the utilities is observed, the above must be integrated over U and V in order to calculate the posterior of θ. Markov chain Monte Carlo provides a feasible method for approximating such integrals: While calculation of $\pi(\theta|m \in \mathcal{S})$ or direct sampling from $\pi(\theta|m \in \mathcal{S})$ is impractical, sampling values of θ, U, V from a distribution approximating $\pi(\theta, U, V|m \in \mathcal{S})$ is feasible. The resulting sampled θ values can be examined marginally to estimate $\pi(\theta|m \in \mathcal{S})$. This kind of data augmentation is often used to simplify posterior sampling (Tanner and Wong, 1987).

Let $\phi = (\theta, U, V)$ represent a state of the unknowns in our model, which includes both the unobserved utilities as well as the unknown parameters. We use the Metropolis-Hastings algorithm to iteratively construct sequences of ϕ from which we can approximate $\pi(\theta|m \in \mathcal{S})$. Given the most recent state ϕ^b, a new value ϕ^{b+1} is constructed by separately updating each unobserved utility, as well as separately updating each component of the preference parameters $\theta = (\alpha, \beta, a, b)$. Each update consists of sampling a proposal value for the utility or preference parameter being updated, then accepting or rejecting the proposal with the appropriate probability. The generation of ϕ^{b+1} from ϕ^b constitutes one scan of the algorithm. As $b \to \infty$, the sampling distribution of ϕ^b

approaches the desired posterior $\pi(\phi|m \in \mathcal{S})$, and so the empirical distribution of the samples of ϕ from the Markov chain can be used to approximate $\pi(\phi|m \in \mathcal{S})$ (Tierney 1994).

Specific methods of constructing such Markov chains are given below. We first address the simple case involving conjugate linear models for utilities, and then discuss a method for constructing chains for some non-conjugate models.

4.2 Conjugate Models: Two-Sided Probit

In this section, we assume the utilities follow the linear model given by (2), the error terms are independent standard normal deviates, and the preferences α, β, a, b are a priori independent and normally distributed: $\alpha \sim \mathcal{N}(\mu_\alpha, \Sigma_\alpha)$, $\beta \sim \mathcal{N}(\mu_\beta, \Sigma_\beta)$, $a \sim \mathcal{N}(\mu_a, \Sigma_a)$, $b \sim \mathcal{N}(\mu_b, \Sigma_b)$. This specification constitutes a two-sided probit model: The probability of each man's match, given his unobserved opportunity set, is a probit model, as is the probability of each woman's match, given her opportunity set.

Utilities: We first consider updating the matching utilities of males. For a particular male i and female j, we sample a proposed utility $U^*_{i,j}$ from its full conditional distribution, which is the distribution of the utility conditional upon the data and the current state of all other utilities and parameters. Sampling from a full conditional distribution is Gibbs sampling, and the acceptance probability of such a sample is always unity.

For normal error terms, the full conditional distribution of $U^*_{i,j}$ given all other unknowns is simply a normal distribution constrained by the other utilities and the stability of the observed matching (cf. Albert and Chib 1993). For example, if man i is matched to woman $w(i)$, then he must have a higher utility for matching with $w(i)$ than for being single or matching with any other woman available to him. Similarly, if man i is single then $U_{i,0}$ must be higher than the utility of being matched to any woman available to him. More precisely, the constrained sampling is as follows:

- if $j \neq w(i)$

  1. if $j = 0$ or $V_{j,i} > V_{j,h(j)}$ (i.e., j is available to i), sample $U^*_{i,j}$ from the normal distribution with mean $\alpha' x_j$ if $j \neq 0$ and mean $a' x_{i,0}$ if $j = 0$, conditional upon $U^*_{i,j} < U_{i,w(i)}$.
  2. if $V_{j,i} < V_{j,h(j)}$ (i.e., j is not available to i), sample $U^*_{i,j}$ from the normal distribution with mean $\alpha' x_j$ (unconstrained).

- if $j = w(i)$, sample $U^*_{i,j}$ from the normal distribution with mean $\alpha' x_j$ if $j \neq 0$ and mean $a' x_{i,0}$ if $j = 0$, conditional upon $U^*_{i,j} > U_{i,0}$ (if $j \neq 0$) and $U^*_{i,j} > U_{i,k}$ for any $k : V_{k,i} > V_{k,h(k)}$.

The utilities $V_{i,j}$ of women for men are sampled the same way, except with V, h interchanged with U, w. Note that sampling the utilities is equivalent to sampling the error terms: For example, sampling $U_{i,j}$ constrained to be less than $U_{i,w(i)}$ is equivalent to sampling $\epsilon_{i,j}$ constrained to be less than $U_{i,w(i)} - \alpha' x_{i,j}$. See Geweke (1991) for efficient methods of sampling from constrained normal distributions.

Preferences: Normal priors for regression coefficients are conjugate for linear models with normal errors, so the conditional distributions of α, β, a, b given the utilities are normal as well. A straightforward calculation shows:

- $\alpha|_{\theta_{-\alpha}, U, V, m \in \mathcal{S}} \sim \mathcal{N}(\tilde{\mu}_\alpha, \tilde{\Sigma}_\alpha)$, where

  $\tilde{\Sigma}_\alpha^{-1} = n_m X X' + \Sigma_\alpha^{-1}, \;\; \tilde{\mu}_\alpha = \tilde{\Sigma}_\alpha (X U' \mathbf{1} + \Sigma_\alpha^{-1} \mu_\alpha)$

- $\beta|_{\theta_{-\beta}, U, V, m \in \mathcal{S}} \sim \mathcal{N}(\tilde{\mu}_\beta, \tilde{\Sigma}_\beta)$, where

  $\tilde{\Sigma}_\beta^{-1} = n_f Y Y' + \Sigma_\beta^{-1}, \;\; \tilde{\mu}_\beta = \tilde{\Sigma}_\beta (Y V' \mathbf{1} + \Sigma_\beta^{-1} \mu_\beta)$

- $a|_{\theta_{-a},U,V,m\in\mathcal{S}} \sim \mathcal{N}(\tilde{\mu}_a, \tilde{\Sigma}_a)$, where

  $\tilde{\Sigma}_a^{-1} = X_0 X_0' + \Sigma_a^{-1}, \;\; \tilde{\mu}_a = \tilde{\Sigma}_a (X_0 U_0 + \Sigma_a^{-1} \mu_a)$

- $b|_{\theta_{-b},U,V,m\in\mathcal{S}} \sim \mathcal{N}(\tilde{\mu}_b, \tilde{\Sigma}_b)$, where

  $\tilde{\Sigma}_b^{-1} = Y_0 Y_0' + \Sigma_b^{-1}, \;\; \tilde{\mu}_b = \tilde{\Sigma}_b (Y_0 V_0 + \Sigma_b^{-1} \mu_b)$

in which

- n_m and n_f are the male and female population sizes,
- X and Y are matrices of female and male characteristics, with n_f and n_m columns respectively,
- X_0 and Y_0 are matrices of self matching characteristics,
- U and V are $n_m \times n_f$ and $n_f \times n_m$ matrices of utilities of men for women and women for men,
- U_0 and V_0 are vectors of self-matching utilities.

Such normal distributions are easy to sample from, and so we use them to generate our proposal values α^*, β^*, a^*, and b^*. Since these are Gibbs samples, the acceptance probability of each proposal is unity.

4.3 Non-conjugate Models: Two-Sided Logit

We now consider models of the form

$$\begin{aligned} U_{i,j} &= u(x_j, y_i|\alpha) + \epsilon_{i,j}, \qquad & U_{i,0} &= u_0(x_{i,0}|a) + \epsilon_{i,0}, \\ V_{j,i} &= v(y_i, x_j|\beta) + \gamma_{j,i}, \qquad & V_{j,0} &= v_0(y_{j,0}|b) + \gamma_{j,0}, \end{aligned} \tag{4}$$

where the functions u, u_0, v, v_0 are determined by the preference parameters α, a, β, and b, respectively, and the error terms are independent and distributed according to known distributions. One convenient choice for the error distribution is the type 1 extreme value (Gumbel) distribution, having a closed form CDF, making constrained sampling relatively easy. When specification (3) is substituted for (4), the non-conjugate model with Gumbel disturbances is an individual-level, symmetrical, two-sided logit model.

Utilities: Conditional on the preference parameters, the distribution of utilities can be resampled by sampling the errors ϵ and γ, constrained by stability. This proceeds in a manner similar to that in the case of normally distributed errors, as described above.

Preferences: Generally there will be no simple conditional distribution for a preference parameter given the utilities. Therefore, instead of generating parameter updates from a full conditional distribution, we can use a random walk proposal distribution, that is, we sample a proposal value uniformly from a box around the current value. For example, a possible proposal distribution q for α is

$$\begin{aligned} q(\alpha^*|\alpha) &= A^{-1} \quad \text{for } \alpha^* \in \alpha \pm \delta_1 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

where δ is a vector of the same dimension as α, and A is the volume of the box spanned by $\alpha \pm \delta_\alpha$. Updates for β, a and b can be generated similarly.

Because such a proposal distribution is not the full conditional, the acceptance probability is not necessarily unity, and needs to be calculated. In what follows, note that the errors ϵ, γ together with the preference parameters θ determine the utilities, and so we can base our Metropolis algorithm and acceptance probabilities on these quantities. For example, when updating the α parameter the

Metropolis-Hastings ratio, or update probability, is

$$\begin{aligned}
\frac{\pi(\theta^*, \epsilon, \gamma | m \in \mathcal{S})}{\pi(\theta, \epsilon, \gamma | m \in \mathcal{S})} \frac{q(\alpha|\alpha^*)}{q(\alpha^*|\alpha)} &= \frac{p(m \in \mathcal{S}|\theta^*, \epsilon, \gamma)\pi(\alpha^*)p(\epsilon)p(\gamma)}{p(m \in \mathcal{S}|\theta, \epsilon, \gamma)\pi(\alpha)p(\epsilon)p(\gamma)} \frac{q(\alpha|\alpha^*)}{q(\alpha^*|\alpha)} \\
&= \frac{p(m \in \mathcal{S}|\theta^*, \epsilon, \gamma)\pi(\alpha^*)}{p(m \in \mathcal{S}|\theta, \epsilon, \gamma)\pi(\alpha)} \qquad (5) \\
&= p(m \in \mathcal{S}|\theta^*, \epsilon, \gamma)\frac{\pi(\alpha^*)}{\pi(\alpha)}, \qquad (6)
\end{aligned}$$

where $\theta^* = (\alpha^*, \beta, a, b)$. Equality (5) holds because the proposal distribution is symmetric, i.e. $q(\alpha^*|\alpha) = q(\alpha|\alpha^*)$. Equality (6) holds because $p(m \in \mathcal{S}|\theta, \epsilon, \gamma)$ is simply the zero-one indicator variable of whether or not m is a stable matching under the values of θ, ϵ, γ, which must be one since θ was accepted at a previous step of the chain. Therefore, the probability of accepting α^* is $p(m \in \mathcal{S}|\theta^*, \epsilon, \gamma)\pi(\alpha^*)/\pi(\alpha)$, which is $\pi(\alpha^*)/\pi(\alpha)$ if m is a stable match under θ^*, and zero otherwise. The acceptance probabilities for proposals of other components of θ are similar.

5 Parametrizations and Identifiability

One formulation of the parametric marriage model described above involves a linear model for the latent, unobserved utilities. The linear model is composed of unknown preference parameters $\theta \in \Theta$ and observed covariates X, Y. We say a linear model is identifiable for the marriage model if for each $\theta_1, \theta_2 \in \Theta$, $\theta_1 \neq \theta_2$, there exists a state of the covariates X, Y such that $p(m \in \mathcal{S}|\theta_1) \neq p(m \in \mathcal{S}|\theta_2)$, where conditional dependence on the covariates is again implicit (this is similar to Manski's (1988) definition of identifiability for logistic regression models). Even though this notion of identifiability is fairly non-restrictive, the complexity of $p(m \in \mathcal{S}|\theta)$ makes it difficult to determine which linear models are identifiable. However, we can determine some models which are not identifiable by using the following formulation: Each individual i has a preference ordering R_i over all members of the population, and the set of all individuals' preference orderings is denoted by $R = \{R_1, \ldots, R_n\}$

(note that restricting individuals to match with the opposite sex can be achieved by assuming an individual prefers being single to matching with any member of the same sex). An observed matching m restricts the possible preference orderings to $\mathcal{R}(m) = \{R : m \text{ is stable under } R\}$. The probability the observed matching is stable can be written

$$\begin{aligned} p(m \in \mathcal{S}|\theta) &= \sum_R p(m \in \mathcal{S}|R) p(R|\theta) \\ &= \sum_{\mathcal{R}(m)} \prod_{i=1}^{n} p(R_i|\theta). \end{aligned}$$

The probability $p(R_i|\theta)$ that an individual has a particular preference ordering is the probability that the individual's utilities follow that ordering.

Scale: Suppose our model for utilities is $U_{i,j} = \alpha' x_j + \sigma \epsilon_{i,j}$, $U_{i,0} = a' x_{i,0} + \sigma_0 \epsilon_{i,0}$, where σ, σ_0 are scale parameters. The probability of any individual's preference ordering is written

$$p(R_i|\theta) = p(\alpha' x_{(1)} + \sigma \epsilon_{i,(1)} < \cdots < \alpha' x_{(n)} + \sigma \epsilon_{i,(n)} < a' x_{i,0} + \sigma_0 \epsilon_{i,0}),$$

where (j) denotes the index of the individual ranked jth under the ordering R_i, and without loss of generality we have assumed actor i prefers being single to matching with anyone in the population. Letting $\theta_1 = (\alpha, a, \sigma, \sigma_0)$ and $\theta_2 = (c\alpha, ca, c\sigma, c\sigma_0)$, it is easy to see that $p(R_i|\theta_1) = p(R_i|\theta_2)$ for each i regardless of the values of the x's. Therefore, $p(m \in \mathcal{S}|\theta_1) = p(m \in \mathcal{S}|\theta_2)$ for all states of the covariates, and the model is not identifiable. Note that this argument does not apply if either one of the scale parameters is set to one.

Intercept: Suppose our model is $U_{i,j} = \alpha_0 + \alpha' x_j + \epsilon_{i,j}$, $U_{i,0} = a_0 + a' x_{i,0} + \epsilon_{i,0}$. The probability of any individual's preference ordering is written

$$p(R_i|\theta) = p(\alpha_0 + \alpha' x_{(1)} + \epsilon_{i,(1)} < \cdots < \alpha_0 + \alpha' x_{(n)} + \epsilon_{i,(n)} < a_0 + a' x_{i,0} + \epsilon_{i,0}).$$

Letting $\theta_1 = (\alpha_0, \alpha, a_0, a)$, it is easy to see that $p(R_i|\theta_1) = p(R_i|\theta_2)$ for any $\theta_2 = (\alpha_0^*, \alpha, a_0*, a)$ such that $\alpha_0 - a_0 = \alpha_0^* - a_0^*$. Thus, the intercepts α_0 and a_0 cannot be simultaneously estimated, but as with the scale parameters this does not preclude us from estimating one or the other.

In addition to the above identifiability issues, there may multiple local maxima of the posterior distribution. These maxima might give rise to a variety of interpretations of the data: For example, consider a population in which women with a relatively high value for some characteristic x tend to be single. Using a univariate version of the linear model (3), this trend in the population could be explained by a male preference for low values of x ($\alpha < 0$), a female self-preference for high values of x ($b > 0$), or both. Such identifiability problem may be ameliorated somewhat by including other characteristics in the model, or by making a priori assumptions that certain characteristics are desirable (i.e. making assumptions about the signs of the preference parameters). Even though the simulation studies of the next section give us some confidence that our methods will assign higher probability to the true values of the preference parameters, we note the importance of a careful parametrization of the model.

6 Simulation Study

A simulation study on twelve datasets was performed to test the ability of the above MCMC algorithm to estimate the preference parameters. Three datasets were generated for each possible combination of two population sizes and two distributions for the observed characteristics. The population sizes used were $n = 200$ and $n = 1000$ individuals (populations having equal numbers of males and females), and characteristic data were generated according to one of two different regimes, outlined below:

Regime 1:

- $x_1, \dots, x_n, y_1, \dots, y_n \sim$ i.i.d $N((0,0)', I)$.
- $x_{1,0}, \dots, x_{n,0}, y_{1,0}, \dots, y_{n,0} \sim$ i.i.d $N((-1,-1)', I)$.

Regime 2:

- $x_1, \dots, x_n, y_1, \dots, y_n \sim$ i.i.d $N((-\frac{1}{2}, \frac{1}{2})', I)$.
- $x_{1,0}, \dots, x_{n,0}, y_{1,0}, \dots, y_{n,0} \sim$ i.i.d $N((0,0)', I)$.

For each generated set of characteristics, utilities were simulated as

$$U_{i,j} = \alpha' x_j + \epsilon_{i,j}, \qquad U_{i,0} = a' x_{i,0} + \epsilon_{i,0}$$

$$V_{j,i} = \beta' y_i + \gamma_{j,i}, \qquad V_{j,0} = b' y_{j,0} + \gamma_{j,0}$$

where the disturbances ϵ, γ are sampled independently from the standard normal distribution, and $\alpha' = a' = (1,3), \beta' = b' = (2,2)$ for each dataset. By estimating α, a, β and b under two different distributions of characteristics, we hope to show that our estimates measure the underlying preferences of the actors, and not the distributions of characteristics. By estimating the parameters with different population sizes, we hope to get some idea of the rate of information gain with increasing population.

¿From the utilities in each dataset, two (possibly identical) stable matchings were generated using the Gale-Shapley algorithm (Gale and Shapley 1962, Roth and Sotomayor 1990). The Gale-Shapley algorithm entails each member of one sex proposing to their most preferred member of the opposite sex. Each member of the opposite sex then makes a temporary engagement with their most preferred proposer, and each rejected proposer proposes to the next member on their ranked list of opposite sex members. The algorithm iterates until no further rejections or proposals are made, at which point all proposers are either engaged or have been rejected by all members of the opposite sex whom they prefer to self-matching.

This matching algorithm was run on each computer-generated dataset twice, once with men proposing and once with women proposing. In systems of utilities allowing more than one possible stable matching, the male-proposing and female-proposing matchings are polar opposites in the sense that the former is at least as good for every man as any other stable matching, and the same property holds for women in the latter (Roth and Sotomayor 1990). These two matchings are called male-optimal and female-optimal, respectively.

Using both matchings separately to estimate the preference parameters is useful in evaluating our methods: The model outlined in Section 1 did not specify which stable matching is being observed in the data. Instead, we have assumed only that *some* stable matching is observed. For this reason, we expect the estimation algorithm to be insensitive to the particular mechanism by which a stable matching has been achieved, be it a male-optimal, female-optimal, or some other stable matching mechanism. We compute separate estimates from the male- and female-optimal matchings to evaluate this property. In systems of real men and women there are potentially many possible stable matchings for a given set of utilities, so insensitivity of the estimation algorithm to the particular matching mechanism is crucial.

6.1 Small Population Study

Prior distributions for all preference parameters were taken to be independent normal distributions, each with mean zero and variance 100. A Markov chain of 500,000 scans was constructed for each stable matching, in which the initial states of the preference parameters were all zero. Posterior distributions were approximated by the empirical distribution of scans 50,001 though 500,000 from each chain, dropping the first 50,000 scans to allow each chain to approach its stationary distribution.

Figure 3 gives the posterior modes and 95% highest posterior density (hpd) regions for the preference parameters. For example, the upper-left panel gives results for α_1 and a_1, the first components of the male preference parameters. The datasets labeled 1, 2, and 3 had characteristics generated under the first regime, whereas the second three datasets had characteristics generated under the second. For each dataset, there are four lines plotted, representing from left to right the hpd regions for α_1 from the male-optimal match, a_1 from the male-optimal match, followed by the corresponding estimates from the female-optimal match. A horizontal line is plotted representing the true value of $\alpha_1 = a_1 = 1$. The other components of the preference parameters are presented in a similar fashion. Note that in datasets 4 and 6, the estimates from the male-optimal and female-optimal matchings are identical. This results from the male- and female-optimal matches being identical for these two datasets.

The results in Figure 3 give some indication of the effectiveness of our estimation procedure. The hpd regions typically cover the true values, and the estimates from the male-optimal and female-optimal matches are quite similar within datasets (we note, however, that the male-optimal and female-optimal stable matchings from the same dataset rarely differed except for changes involving a few individuals). Also note there is more posterior uncertainty in the self-matching parameters a and b than in α and β. This reflects the fact that our model assumes the existence of, for example, $n_m \times n_f$ utilities of men for women, compared to only n_m utilities for male self-matching.

6.2 Large Population Study

To measure how much information is gained as sample size increases, we repeated the simulation study with a larger population size. The simulation procedure described above was re-implemented with 500 men and 500 women in each dataset.

Figure 3: Posterior modes and hdp regions from the small population study.

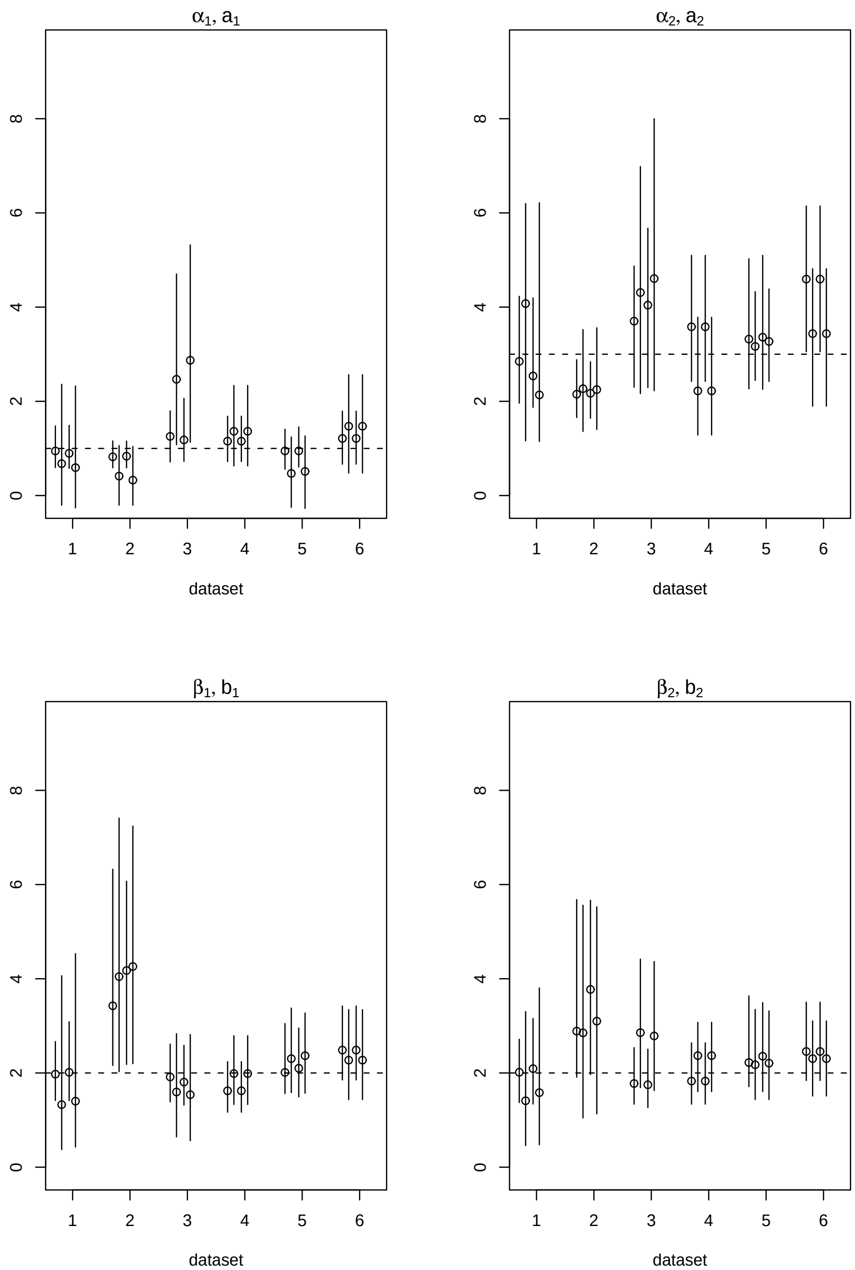


Figure 4: Posterior modes and hdp regions from the large population study

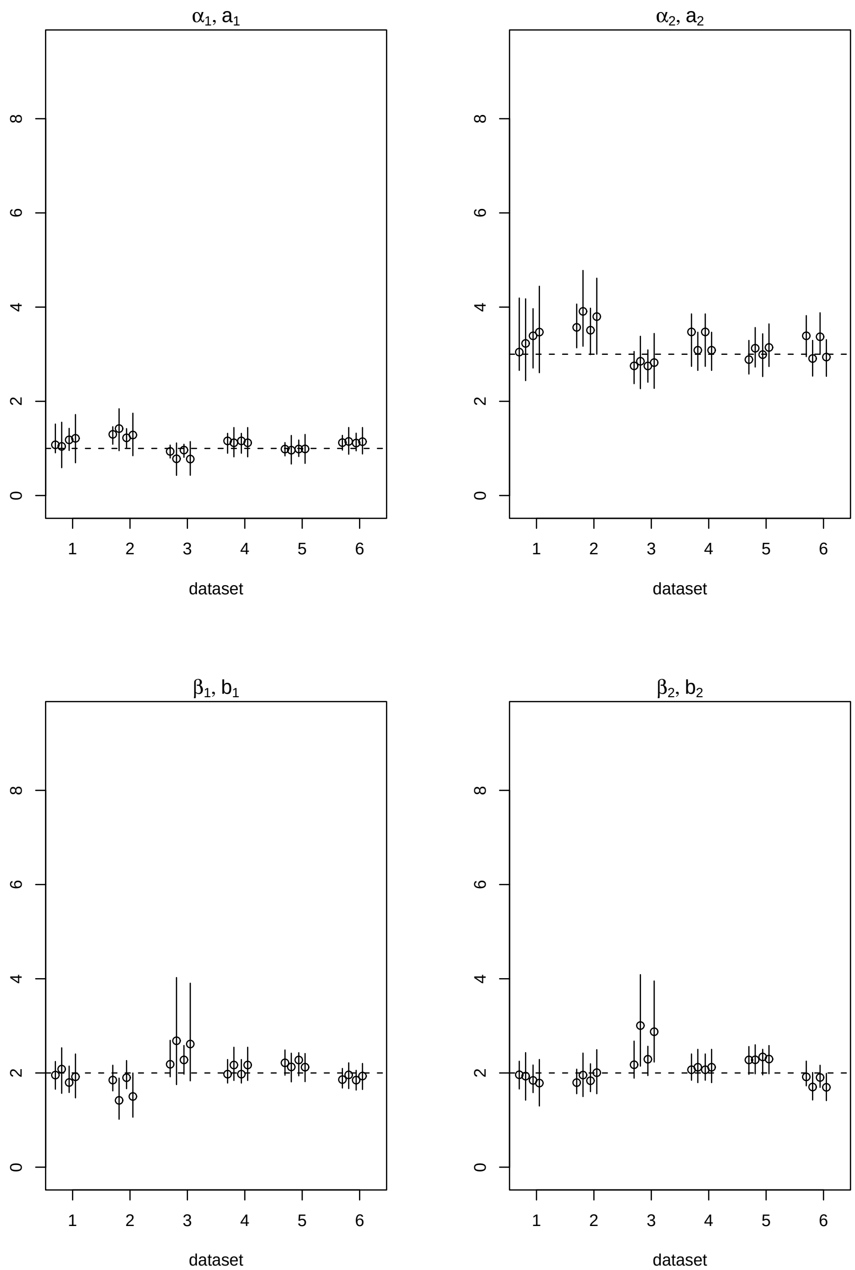


Table 1: Comparison of small and large population studies

Precision Statistic	$n_f = n_m = 100$	$n_f = n_m = 500$
average absolute bias, α and β	0.45	0.18
average absolute bias, a and b	0.65	0.25
average hpd region length, α and β	1.79	0.62
average hpd region length, a and b	2.69	0.95

As seen in Figure 4, the coverage probabilities of the hpd regions are similar to those of the small simulation study, but the estimates are more precise. Table 1 compares the small and large population studies in terms of average absolute bias of the estimates and width of the hpd regions. The average absolute bias is roughly one-third of that in the small population study. The average widths of the hpd regions from the large population study are also roughly one-third of the average widths of the small population study. Of course, running the Markov chain for the large populations involves much more computation, as the number of utilities to keep track of increases quadratically with population size if the numbers of men and women are kept roughly equal. Thus computation for a dataset with 500 men and 500 women is roughly 25 times more intensive than for a dataset of 100 men and 100 women.

7 PSID Analysis

We now analyze a dataset extracted from the 1993 Panel Study of Income Dynamics, and estimate the preferences of men and women for the age and education levels of potential matches.

The 1993 PSID includes data on 9977 households containing 15,650 adult individuals. From the data on households having non-missing levels of education and age, a marriage rate of 71.6%

was calculated (marriage defined here as meaning any form of cohabitation), and a random sample of 140 couples and 120 single individuals was selected, resulting in a population of size 400 while maintaining a marriage rate of 70%. The 120 single individuals were obtained by randomly sampling cases from the single households, giving 82 single women and 38 single men. This discrepancy between single male and female populations is characteristic of the PSID dataset, and can be explained by factors such as differing rates of mortality, incarceration, military participation, and homelessness.

Note that we do not expect actors in the dataset other than couples to actually know each other. However, it is still of interest to estimate the underlying preferences of the actors in this system, and so we approximate the distribution of potential partners' characteristics with the empirical distribution of characteristics from the sample. In essence, we view the actors in the system as surrogates for the types of individuals that any particular actor is likely to have in their pool of potential matches.

The PSID data were presented graphically in Figure 1 above. The scatterplots in the figure suggest a tendency of actors to form matches with others having similar age and education levels. This could reflect preferences of actors to choose mates like themselves. However, such similarity between mates can also arise when actors share the same, absolute preferences for desirable traits in their partners, rather than preferring partners like themselves. In such cases the similarity of partners results from the competitive outcome of the matching process, a possibility that appeared in the simulations in section 5. Even though the simulated datasets there were generated under a specification in which all actors shared the same preferences for the absolute (rather than relative) characteristics of their partners, the resulting scatterplots (not presented) also show similarities between matched partners much like Figure 1. Thus it is interesting to consider whether the similarity between partners in the PSID data reflects actors' preferences for similar mates, rather than simply preferences for more attractive mates. For this reason, we consider a variant of the model presented in Section 2, and allow for the possibility that a man's utility for a woman might depend on her similarity relative to his own characteristics, as well as on her absolute characteristics, and vice-versa for women. Letting $(x_{j,1}, x_{j,2})$ and $(y_{i,1}, y_{i,2})$ be the age and education levels of woman j and man i respectively, we assume the utilities for matching can be written as follows:

$$\begin{aligned}
U_{i,j} &= \alpha_1(x_{j,1} - y_{i,1})^2 + \alpha_2 x_{j,1} + \alpha_3(x_{j,2} - y_{i,2})^2 + \alpha_4 x_{j,2} + \epsilon_{i,j} \\
V_{j,i} &= \beta_1(y_{i,1} - x_{j,1})^2 + \beta_2 y_{i,1} + \beta_3(y_{i,2} - x_{j,2})^2 + \beta_4 y_{i,2} + \gamma_{i,j} \\
U_{i,0} &= a + \epsilon_{i,0} \\
V_{j,0} &= b + \gamma_{j,0}.
\end{aligned} \tag{7}$$

The disturbance terms are all modeled as independent standard normal random variables.

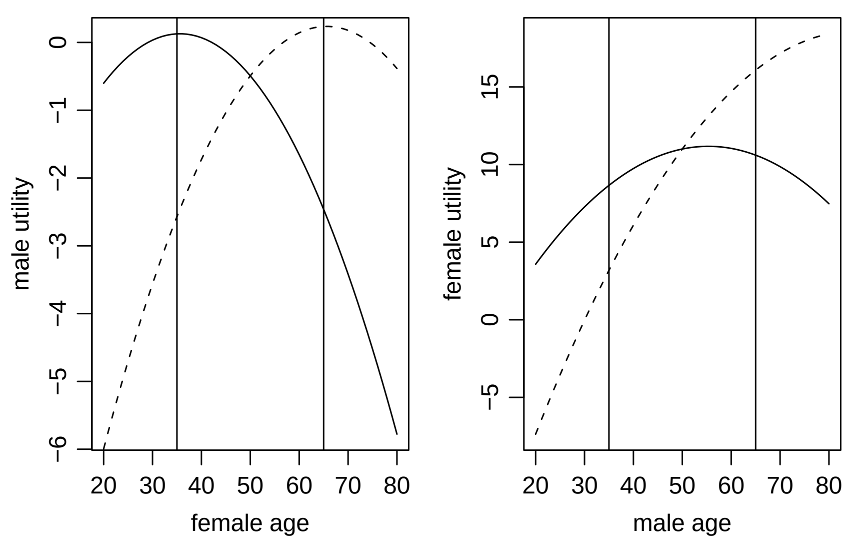
Two Markov chains of length 2×10^6 scans were constructed using the methods of Section 3.2, having different random seeds, but each having starting values of zero for all regression coefficients. The prior distributions for the regression parameters were taken as independent normal distributions, each with a mean of zero and variance of 100. Output of each Markov chain was recorded every 200 scans. The two chains indicate slow mixing, but approximate stationarity after about 500,000 scans (although in one of the chains, there is a slight indication of drift in b, the female self-matching parameter).

Means and standard deviations of preference parameters were calculated using the pooled second-halves of each Markov chain, and are given in Table 2. Because of the pooling, standard deviations reflect both within- and between-chain variability. The table also shows values of $\sqrt{\hat{R}}$ for each parameter, a measure of the ratio of between-chain to within-chain variance of the

Table 2: Posterior means, standard deviations, and $\sqrt{\hat{R}}$ of men's and women's preference parameters.

	Men's Preferences				Women's Preferences			
characteristic		mean	sd	$\sqrt{\hat{R}}$		mean	sd	$\sqrt{\hat{R}}$
squared age difference	α_1	-0.003	0.0004	1.012	β_1	-0.008	0.002	1.280
absolute age	α_2	0.001	0.008	1.122	β_2	0.242	0.088	1.344
squared education difference	α_3	-0.018	0.005	1.031	β_3	-0.054	0.030	1.000
absolute education	α_4	0.049	0.021	1.012	β_4	0.051	0.137	1.006
constant	a	1.452	0.588	1.110	b	6.276	3.403	1.112

Figure 5: Estimated mean utilities for age; fixed education levels. Solid lines indicate utilities of 35 year olds, dashed of 65 year olds.



parameters (Gelman 1996). Values close to 1.0 indicate that between-chain variability is not high, based on what would be expected from the variability within chains. The values of $\sqrt{\hat{R}}$ in the table are all below Gelman's suggested cutoff of 1.2, except for the two age preferences of women, which are about 1.3. Both men and women show a significant preference for mates with a small squared age difference from themselves. Beyond age differences, men do not show a strong preference for youth or age in potential mates, whereas women show a strong absolute preference for age. The combined absolute and relative preferences for age are shown graphically in Figure 5, which plots the contribution of the age effects to the mean utilities of men and women of ages 35 and 65. For men of both ages, women of similar age are the most preferred, while women prefer men who are more than ten years older than themselves. The latter effect could be due in part to preferences for unmeasured correlates of men's ages, such as earnings, which could be explored in further analyses. The combined effects of absolute and relative preferences for education are similar across genders: Men and women with ten years of education prefer partners with about twelve years of education, while those with sixteen years also prefer partners with slightly more education than themselves. However, the preferences of women for education are not clearly distinguishable from zero.

As a measure of the goodness of fit and applicability of this model to the PSID sample, we evaluate the model's predictive power: Do the estimated preference parameters predict relevant features of the observed dataset? To answer this question, we sampled ten stable matchings from the posterior predictive distribution as follows:

1. Obtain preference parameters α, a, β, and b from a randomly-selected scan in the second half of the Markov chain, and disturbances ϵ and γ drawn from unconstrained standard normal distributions.

2. Using the observed male and female characteristics X, Y from the sample, generate a set of

utilities using Model 7.

3. Use the Gale-Shapley algorithm to generate a stable matching between the men and women.

4. Record relevant statistics of the predicted matching, such as the marriage rate, average within-match age and education differences, correlation of within-match ages and education levels, etc.

Note that this method of evaluating goodness-of-fit demonstrates a technique by which data from one marriage market can be used to make predictions about a matching in another market having a different distribution of characteristics, provided the underlying preferences remain the same.

The summary statistics for ten such predicted matchings are given in Table 3, and the matched age and education characteristics from the last of these matchings are plotted in Figure 6. The model does a reasonably good job of capturing many of the relevant features of the actual dataset shown in Figure 1, although it tends to underestimate the mean within-couple age difference, and overestimate the variability of this difference. This suggests a more flexible model in which ages enter the utility equation as $|x_{j,i} - y_{i,j}|^{\gamma}$, where γ is a parameter to be estimated.

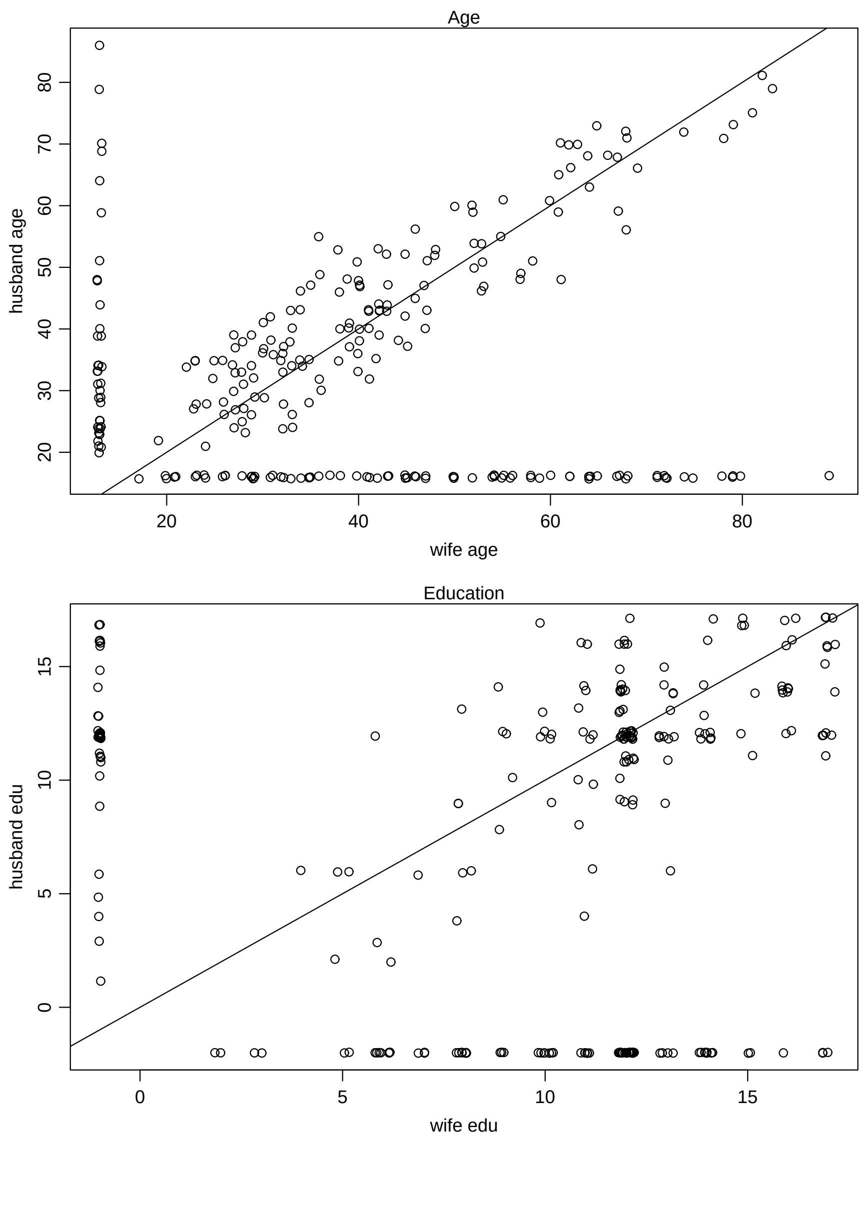
8 Discussion

Our estimation results using simulated and actual data on marriages suggest that the behavioral model of two-sided matching theory can be a useful foundation for inference about the preferences of men and women, even in the absence of detailed information about the contexts that led to the formation of observed marriages. The use of these estimates for projections of marriages that would occur under counterfactual resource distributions has yet to be explored, though the necessary methodology was demonstrated in the reconstruction of the PSID scatterplots from the posterior

Table 3: Goodness of fit statistics for observed and predicted matchings

sample	marriage rate	mean age diff	var age diff	corr age	mean edu diff	var edu diff	corr edu
observed	0.70	3.26	31.74	0.91	-0.04	5.60	0.74
1	0.72	1.53	43.10	0.90	0.12	5.22	0.70
2	0.68	1.75	38.48	0.91	0.05	5.25	0.69
3	0.66	2.00	40.88	0.91	-0.25	7.22	0.57
4	0.68	2.72	47.38	0.88	0.01	5.87	0.66
5	0.66	1.62	38.35	0.91	-0.15	5.17	0.72
6	0.69	2.36	58.35	0.85	-0.09	7.37	0.55
7	0.67	1.84	47.03	0.89	0.30	6.35	0.55
8	0.69	2.28	40.70	0.90	-0.33	5.04	0.72
9	0.64	3.16	61.82	0.84	-0.16	7.07	0.56
10	0.68	2.96	47.42	0.89	0.04	6.816	11.375

Figure 6: A stable matching, generated from PSID data and preference parameters estimated from the posterior.



distribution in section 6. More detailed behavioral models could also be used for projection, by combining the preference estimates obtained from the present method with more specific ideas of information flow and partner-seeking strategies.

Our proposal is to draw inference about preference parameters θ using the likelihood component $p(m \in \mathcal{S}|\theta)$, where $m \in \mathcal{S}$ is the event that m is one of the stable matchings of the system. We note briefly that $p(m \in \mathcal{S}|\theta)$ may also be viewed as a certain conditional likelihood function when additional model assumptions are made. The matching M to be recorded and the event $M \in \mathcal{S}$ constitute the full data (in addition to covariate information). That $M = m$ and $M \in \mathcal{S}$ imply $m \in \mathcal{S}$ means that $p(m \in \mathcal{S}|\theta)$ is a component of the full likelihood, as argued earlier in section 2. Indeed our numerical results demonstrate that this component contains a large amount of information on θ; importantly, its use frees us from the modeling of strategies by which the observed matching arises. Straightforward manipulations show that if all the utilities (U, V) are independent of M, a priori, then $p(m \in \mathcal{S}|\theta) = p(M \in \mathcal{S}|M = m, \theta)$; i.e., it is the conditional probability that the system is stable given the matching. Conditional likelihood is often used when one component of the data is difficult to model (e.g. ascertainment in pedigree analysis). The independence of utilities and matching required to draw this connection may be suspect, since of course matchings are considered to arise from the utilities by some mechanism. The independence could be viewed as a form of prior ignorance, since it is marginal to the event $M \in \mathcal{S}$. On the other hand, marginal likelihood is also an approach that is used when one component of the data is difficult to model. In our case, the marginal likelihood approach described in section 2 has the advantage of not requiring a suspect assumption analogous to the one just described.

The marriage model in this paper can be thought of as a special case of a more general matching problem: Consider a population of n individuals or objects, each member potentially forming a match with any other member of the population. Utilities are then represented by an $n \times n$ matrix U, and stability implies $U_{i,m(i)} \geq U_{i,j} \forall j : U_{j,i} \geq U_{j,m(j)}$, where $m(i)$ is the index of the partner of i. The marriage model can be represented as a special case of this more general model by parametrizing the utilities such that $U_{i,j}$ is $-\infty$ if i and j are two different individuals of the same type (e.g. of the same sex in a marriage market, or either both workers or both employers in the labor market). This general model and modifications to it could prove useful in a variety of matching or choice problems in which the set of matches available to an actor is potentially limited by the choices of other actors.

One modification we are investigating is to allow for correlations among the disturbances in the utility functions across alternative potential partners. Such correlations are commonly advocated when important attributes of alternatives may have been omitted from a choice model (but see Horowitz [1991]). We expect the basic approach of the present paper can accommodate this extension with appropriate restrictions on the covariances.

The MCMC methods described here show the promise of making the marriage model applicable to data sets of moderate to large size. However, as we have seen in the simulation study, it takes a fairly large dataset to make precise inference on the preference parameters. The difficulty with much larger datasets is that the number of latent utilities increases quadratically with the population size. However, if we were to assume that each individual only has information about or access to a small subset of the population (which is probably more realistic than the current assumption of system-wide stability), we could drastically reduce the number of utilities in our model. Access as such could perhaps be modeled based on existing data: For example, we could assume actors have information only on spouses or jobs within a geographic region, within an economic class, or within a social network. Incorporating models of limited access within the framework of the marriage model is a subject of current research of the authors.

References

Albert, J. H., and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679.

Ben-Akiva, M., and Lerman, S.R. (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge, MA: MIT Press.

Bergstrom, C. T., and Real, L.A. (2000), "Towards a Theory of Mutual Mate Choice: Lessons from Two-Sided Matching," *Evolutionary Ecology Research.* 2, 493–508.

Coleman, J.S. (1990), *Foundations of Social Theory.* Cambridge, MA: Harvard University Press.

Gale, D., and Shapley L.S. (1962), "College Admissions and the Stability of Marriage," *American Mathematical Monthly.* 69, 9–15.

Gelman, A. (1996), "Inference and Monitoring Convergence," in W.R. Gilks, S. Richardson and D.J. Spiegelhalter, eds., *Markov Chain Monte Carlo in Practice*, New York: Chapman and Hall, 131–143.

Geweke, J. (1991), "Efficient Simulation From the Multivariate Normal and Student-t Distributions Subject to Linear Constraints." in *Computing Science and Statistics. Proceedings of the 23rd Symposium on the Interface*, 571–578. Fairfax Station, VA: Interface Foundation of North America.

Gilks, W.R., Richardson, S. and Speigelhalter N.G. ,eds. (1996), *Markov Chain Monte Carlo in Practice*, New York: Chapman and Hall.

Horowitz, J. L. (1991), "Reconsidering the Multinomial Probit Model." *Transportation Research B*, 25, 433–438.

Logan, J.A. (1996a), "Rules of Access and Shifts in Demand," *Social Science Research*, 25, 174–199.

Logan, J. A. (1996b), "Opportunity and Choice in Socially Structured Labor Markets," *American Journal of Sociology*, 102 ,114–160.

Logan, J. A. (1996c), "Rational Choice and the TSL Model of Occupational Opportunity," *Rationality and Society*, 8, 207–230.

Logan, J. A. (1998), "Estimating Two-Sided Logit Models," *Sociological Methodology*, 28, 139–173.

Manski, C. (1988), "Identification of Binary Response Models," *Journal of the American Statistical Association*, 83, 729–738.

Roth, A. E., and Sotomayor, M. A. (1990), *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Cambridge: Cambridge University Press.

Roth, A. E., and Vande Vate, J.H. (1990), "Random Paths to Stability in Two-Sided Matching," *Econometrica*, 58, 1475–80.

Tanner, M. A., and Wong, W.H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association* 82,528–540.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions" (with discussion), *The Annals of Statistics*, 22, 1701–1762.