

Structural Estimation of Marriage Models

Y. Linda Wong*

Revised: February, 2002

Abstract

This paper uses a structural approach to examine who matches with whom. A two-sided matching model that allows for marital sorting in response to marriage market flexibility and agents' preferences is utilized. Estimation is based on imbedding the numerical solution of a matching model within a maximum-likelihood procedure. Results using the Panel Study of Income Dynamics (PSID, 1968-1993) indicate that wage is a more desirable trait than education in predicting marriageability for white men, while education is more desirable for black men. The marriage market for white men is also more flexible. The desirability of wage and marriage market flexibility both decrease with age for white men. The effects of age for black men are mixed.

Keywords: Positive Assortative Mating, Marriage Classes, Classification Errors

JEL Classification: C61, C63, J41

*I have benefited from helpful suggestions of an anonymous referee, Michael Keane, Glenn McDonald, George Neumann, and seminar participants in the 7th International Conference of the Society for Computational Economics at Yale 2001 and the Search and Matching Conference at the University of Iowa 2000. Unfortunately, all errors are mine.

1 Introduction

Studies of who matches with whom have been applied to various markets such as legal market, labor market, sports market, and marriage market.¹ How individuals sort themselves into marriage, in particular, has important implications for the distribution of income, labor supply, and fertility (Pencavel (1996), Boulier and Rosensweig (1984), and Becker (1973)). Most empirical works on who matches with whom in the marriage market have been descriptive in nature (Keller et al. (1996), Spuhler (1982), and Vandenberg (1972)). These studies are largely based on a reduced-form approach using spousal correlation indices for various traits. But such an approach is *ad hoc* and inadequate in explaining individuals' sorting outcomes. For example, evidence shows a low spousal correlation for wages, as compared with other traits such as age and education.² Does this imply that wage is not an important trait in individuals' preferences? Or does it reflect that the marriage market is inflexible, so agents scramble for partners of various wage attributes instead of being selective? Observations of who matches with whom contain more information than a simple association of spousal traits. They contain valuable information on how efficiently the marriage market functions and on agents' preferences in terms of mate choice.³ But a reduced-form approach cannot address market efficiency or agents' preferences that affect the value of marriage. A better understanding of who matches with whom calls for a structural investigation.

Dynamic two-sided matching models provide an appealing framework for analyzing partnership formation among heterogeneous individuals whose utility is non-transferable.⁴ They also offer a structure that characterizes the stochastic processes governing partners' arrival/separation and the choice of partners (Burdett and Coles (1999, 1997), Bloch and Ryder (1998), and Collins and McNamara (1990)). Given these frictions, equilibrium matching outcomes are driven by the underlying assumption of match utility/production, which affects individuals' outside options. Because two-sided matching models are relatively new, little is known about structural estimation in this context.

This paper examines who matches with whom using a structural approach based on the framework of two-sided matching models. Because the model is capable of an-

¹For example, Spurr (1987) studies positive assortative mating between the size of law firms (legal claims) and lawyers' quality. Oi (1983) studies the matching between the number of entrepreneur and workers' productivity. Rosen (1981) shows that technological change leads gifted athletes to match with larger audience. Becker (1973) proves that men and women are positively sorted in complementary traits.

²Census evidence on spousal correlation in terms of wage, age, and education for newlyweds is provided in the appendix.

³Hereafter I use flexibility and efficiency interchangeably.

⁴Dynamic two-sided models are used in this paper instead of static and frictionless two-sided models such as that of Gale and Shapley (1962) because the theory of stable matching is difficult to reconcile with facts such as the rising divorce rate. Moreover, in static models agents' preferences are idiosyncratic and consequently the link between equilibrium allocation and preferences is unclear.

swering how market efficiency and individuals' preferences affect who marries whom, the results give more insights into individuals' behavior than those using the simple method of spousal correlation for traits. Using the PSID (1968-1993), results indicate that the marriage market for white men is more flexible than that for black men. Across various age groups, the flexibility of the marriage market decreases with age for white men. Wage is a more desirable trait for marriageability in younger than older white men. Education is a more marriageable trait for black men, but the effects of age for black men are mixed.

The proposed estimation method enables me to include more than one trait as individuals' types and to account for classification errors.⁵ These two measures are less likely to be implementable in a simple correlation index. Results reject the use of one trait as individuals' overall type. More important, results indicate that inattention to classification errors leads to false predictions of marriage market inflexibility and the inaccurate conclusion that wage is undesirable. Once classification errors are accounted for, the marriage market is more flexible, and wage is found to be a more desirable measure of marriageability than education for white men. As temporal income (wage) is a more desirable trait for white men, white men are more impatient than black men, whose marriageability relies more on the permanent income of education. This result is consistent with the finding that the marriage market is more flexible for white men.

Empirical results are based on a two-sided matching framework in the spirit of Burdett and Coles (1997) and Collins and MaNamara (1990), where individuals' match utility is nontransferable and matching is random. In the model, agents are *ex ante* heterogeneous, with each represented by a marriage index. Finding potential matches is a time-consuming process, characterized by scarcity of information. Given such a costly matching process generated by search friction, a range of acceptable potential partners is possible. Agents choose an acceptable set of partners to maximize expected wealth. The equilibrium of who matches with whom is affected by the structural parameters that characterize the distribution of the marriage index and the frequency of agents' marriage market transitions.

Estimation of the structural parameters of a matching model is done where observed couples are considered as an equilibrium outcome. The estimation method developed in this paper improves on empirical works describing who matches with whom in the following ways. First, the method of ranking agents based on multiple measured attributes such as wage and education is more general than using a single trait as agents' types in standard literature (Montgomery and Sulak (1986), and Boulier and Rosensweig (1984)). Second, to account for omitted variables and unobserved traits (by the econometrician), a flexible classification error model is developed. Third, the method of estimation is straight-forward. Given a distribution of agents' marriage indices and the frequency of marriage market contacts and separations, the

⁵Because some traits are excluded (either due to data unavailability or unobservability by the econometrician), classification errors are used in the estimation.

acceptable pool of partners for each type of agent can be solved numerically following an optimal reservation-match policy. The numerical solution is then nested within a maximum likelihood procedure to estimate structural parameters. The estimation technique is applied to samples of the PSID. Since the probability of who matches with whom and the spell length of looking for partners are endogenously determined by the optimal strategy, I show that observations of the age at first marriage and the characteristics of matched partners provide information on identifying structural parameters.

Fourth, individual marriage history is used rather than census data, which record a stock of individuals at a given point in time. The advantage of using marriage history data is two-fold. Agents' marriageable traits are drawn at the time of marriage because a pool of single agents is followed until they get married. This avoids the problems in most studies that use post-marriage characteristics from a census.⁶ These problems include division of labor within a family that may affect choices of occupation and the labor supply, so post-marriage wage may be correlated with marital choice. Another advantage of using panel data is that the structure of data allows a direct application to the dynamic nature of the model, thereby allowing direct testings of the model's assumptions, and hence, testing of its usefulness.

This paper is similar to Burdett and Coles (1999, 1997), Block and Ryder (1998), and Collins and McNamara (1990) that study partnership formation among heterogeneous agents. But unlike those and as in Smith (1997), this paper utilizes productive aspects of marriage. In this paper, agents are productive while single and while married. Further, unlike Bloch and Ryder (1998), who assume cloning, I assume exogenous separation to achieve steady state singlehood distribution. I choose exogenous separation because it is more realistic than cloning, and it gets rid of the second infinity problem.⁷

This paper is similar in technology to Mortensen and Neumann (1988) in the sense that the reservation function is computed as a fixed point of a contraction map. Mortensen and Neumann perform Monte Carlo evidence on the properties of structural estimation of a homogeneous search model when solving for exact reservation wage. This paper is also similar to typical applications in search models and equilibrium search models such as that of Bontemps et al. (1999), Bunzel et al. (1997), and Kiefer and Neumann (1993) in two ways. First, these studies consider steady-state reservation solution and the estimation of structural parameters in a continuous time and infinite horizon model. Second, these applications take workers' labor market history and the accepted wage at employment to identify the intensity parameters in the stochastic processes and to determine the endogenous productivity distribu-

⁶See for example, Suen and Lui (1999), Bergstrom and Schoeni (1996), Montgomery and Sulak (1986), and Boulier and Rosensweig (1984).

⁷This assumption is also less complicated than exogenous inflow and endogenous entry. Endogenous entry is used in standard search-matching literature in the labor market to determine the number of vacancies (Pissarides, 1990).

tion or wage distribution. In this paper, identification requires knowledge of agents' marriage history and their characteristics at the time of the formation of matches. Such formulation allows one to test the restrictive assumptions on the memoryless properties of partners' arrival and separation rates made in the model.

The matching model is outlined in section 2. Section 3 contains a discussion of the empirical implementation of the model. The structural model is estimated by maximum likelihood. I derive the likelihood function of the matching model in subsection 3.1; then in subsection 3.2, I propose a discrete index as an attempt to rank agents' measured attributes. I present a matching algorithm in subsection 3.3. I close the section by deriving estimation strategies that take into account classification errors in estimation. Section 4 contains a description of the data. Estimation results are given in section 5. I present estimates of partners' arrival rate, separation rate, and wage elasticity for distinct age categories. Section 6 concludes.

2 The Matching Model

Time is continuous, and there are two groups of infinite-lived agents: men and women. At each point in time, all agents are in one of two states: single or married. Only single agents search for marriage partners. Let the arrival rate of partners for single agents be λ , which is governed by a Poisson process. Given an arrival of partners, agents decide whether to form a match. If a match is formed, it dissolves exogenously at rate δ , and agents flow back to the single pool. If no match is formed, then an individual searches again. Thus, the market is described by the structural parameters that describe market transitions $\langle \lambda, \delta \rangle$.

Agents are *ex ante* heterogeneous with respect to their types x_i . Let there be J types for men and women, where J is a positive finite integer bounded away from ∞ . The range of types is $[\underline{x}_i, \bar{x}_i]$, where \underline{x}_i and \bar{x}_i indicate the infimum and supremum of its support, and $\underline{x}_i \geq 2$.⁸ Assume for illustration that the distribution of types is the same for men and women.⁹ Let γ^i denote the distribution of types among single men who will propose to a type j woman if they meet, i.e., $\gamma^i = \sum_{k=1}^i \gamma_k = \Pr(X < x_{i+1} | x_j)$, where γ_i denotes the corresponding probability density function. Types are distributed according to γ^i , independently of spells of being single.

While an agent is single, the instantaneous utility is the real value of the agent's type. When an agent is married, instantaneous utility is assumed to be an equal split of match production $x_i x_j$, where i represents the type of men and j represents the type of women. The types of two potential partners are revealed upon meeting and each agent must decide whether or not to accept a proposal.

⁸The infimum \underline{x}_i (\underline{x}_j) must be at least as large as 2 to satisfy the incentive constraint for marriage: that the match utility is at least as large as the utility while single, $\frac{x_i x_j}{2} \geq x_i$ ($\frac{x_i x_j}{2} \geq x_j$). Should this constraint fail to hold, a fraction of each sex will not be married.

⁹This assumption will be relaxed in the empirical estimation.

A type i agent chooses a range of acceptable types of potential partners with the objective of maximizing his expected discounted value in the future utility stream. A single agent has instantaneous utility x_i , and the expected benefit of marriage following an optimal policy if partner type Y is realized, given that a partner has arrived. If agents discount future income at rate β , the value of being single is

$$V(x_i) = \frac{x_i + \lambda E \max[W(x_i, Y), V(x_i)]}{(\beta + \lambda)}, \quad (1)$$

where $W(x_i, Y)$ is the expected discounted value of marriage with a random partner of type Y . The *ex post* value of marriage is made up of the match utility given by the equal split of the realized match production and the value of remaining single due to an exponential random separation,

$$W(x_i, x_j) = \frac{x_i x_j}{2(\beta + \delta)} + \frac{\delta V(x_i)}{(\beta + \delta)}. \quad (2)$$

A marriage proposal is acceptable if the potential partner is above the value of the reservation type R_i of type i . Let r represent the ranking of R , so $R_i = x_{r_i}$. Not all potential partners whose type is above r accept marriage proposals offered by given agents in equilibrium. Some potential partners are simply unattainable because they may prefer not to match with lower type agents. This feature is the equilibrium outcome of the model and not a restriction. Denote M_i be the value of the maximum-attainable type and m be the ranking of M , so $M_i = x_{m_i}$. The acceptance set of a type i agent is $A_i = \{j | r_i \leq j \leq m_i\}$. The probability of acceptance is $\Pr(A_i) = \gamma^{m_i} - \gamma^{r_i}$.

The optimal policy is a reservation-match policy given by: $W(x_i, R_i) = V(x_i)$. If a marriage offer falls within the agent's acceptance set, the agent will accept the match proposal following the optimal policy; otherwise, the offer will be rejected. Combining equations (1), (2), and the optimal policy, the reservation type is the solution to the following equation (see appendix for derivation)

$$R_i = 2 + \frac{\lambda}{\beta + \delta} \sum_{j \in A_i} (x_j - R_i) \gamma_j. \quad (3)$$

The solution is unique because the left-hand side of equation (3) is increasing in R_i and the right-hand side decreasing. Since the situation is symmetric between men and women, the reservation type for women satisfies

$$R_j = 2 + \frac{\lambda}{\beta + \delta} \sum_{i \in A_j} (x_i - R_j) \gamma_i. \quad (4)$$

A Nash equilibrium of matching requires that (i) each agent find a potential partner acceptable, and (ii) every single agent select his(her) own partner type to

maximize the expected net benefit flow attributable to the choice of partner following (3) (or (4)), given the optimal choices made by all other single agents. In other words, two sets of joint matching strategies for men and women: $\{R_i, M_i\}$, $\{R_j, M_j\}$, for all $i, j = 1, \dots, J$, that solve (3) and (4) are Nash equilibrium solutions to the non-cooperative stationary game of matching.¹⁰ The equilibrium solution is a two-dimensional graph.

Positive assortative matching, i.e., a positive relation between reservation types and agents' own types, is the predicted equilibrium outcome. This result is driven by the underlying specifications of agents' output, which is agents' type when single and the product of each partner's type when married. The match output encompasses increasing returns to scale as an important gain from marriage. This property is necessary to create incentives to trade with higher type agents to exploit match rents when narcissism is allowed. When utility of being single is agents' own types, waiting for higher type partners are less costly, and agents who were previously accepting lower type partners will prefer to wait for higher type partners. Agents' outside options must be sufficiently high to induce positive assortative matching. If a constant returns to scale match output such as $\sqrt{x_i x_j}$ or $x_i + x_j$ is assumed, negative assortative matching arises.¹¹ When agents do not value their singlehood utility or when search friction is absent, increasing returns to scale is not necessary to induce positive assortative matching.¹²

In this model an increase in reservation type can be explained by either a rise in the arrival rate of partners, a fall in the separation rate, or a combination of both. The ratio $\frac{\lambda}{r+\delta}$ can be thought of as a measure of the inverse of search friction. This is a ratio of partners' arrival rate to the sum of the discount rate and separation rate, and so it measures the relative speed of offers. As $\lambda \rightarrow \infty$, partners arrive instantaneously. Friction disappears, each type of agent matches with his/her most desirable type, and strict positive assortative matching results, as in a centralized market. When chances to meet partners are low, $\lambda \rightarrow 0$, all agents are least selective. For values of λ between zero and infinity, a range of partner types is acceptable.

3 Estimation Strategy

The object of interest is to estimate the likelihood of a type i agent marrying a type j agent. The method used is maximum likelihood, for two reasons. First, the outcome of the matching model endogenously generates an acceptance set for each type of agent, which allows the derivation of the probability of who matches with whom.

¹⁰For a prove of the existence of the equilibrium, see Smith (1997).

¹¹See Wong (1998) for discussions.

¹²Burdett and Coles (1999, 1997) characterize positive assortative matching when narcissism is ruled out. Under a frictionless framework, Becker (1973) shows that monotonic preferences are sufficient for positive assortative matching.

Second, the assumptions underlying the model enable derivations of the distributions of duration data.

3.1 The Likelihood Function

Identification of all parameters requires knowledge of couples' characteristics at first marriage as well as an individual's marriage history. All agents are single at the first interview, whence information on how long they have been single can be obtained. Information on the duration of singlehood is obtained by following single agents after the first interview. Analogous to the renewal theory, these durations are backward T_{ob} and forward T_{of} recurrence times respectively. Therefore, the duration of singlehood is $T_0 = T_{ob} + T_{of}$. The data consist of a panel where some individuals are single with duration T_0 , married with duration T_1 , and a given individual's and spouse's logarithm of wage and education at first marriage are w and e respectively. Assume that the parameters of males and females are the same, the structural parameters to be estimated are λ and δ .

Consider a type i man who is single at first interview. Let T_{ob} and T_{of} be i.i.d. and have an exponential distribution with parameter $\lambda(\gamma^m - \gamma^r)$. Let D_{ob} (D_{of}) denote a binary variable that equals one, if it is known that the elapsed (residual) duration exceeds a certain value, i.e., left-censored (right-censored), and zero otherwise. Conditioned on being type i , the individual contribution of singlehood duration until and including the time of exit into marriage or censoring is

$$L_{0i} = [\lambda(\gamma^m - \gamma^r)]^{1-D_{ob}+1-D_{of}} \exp[-\lambda(\gamma^m - \gamma^r)(T_{ob} + T_{of})], \quad (5)$$

where $T_{0f} > 0$ and $T_{0b} > 0$.

Events occurring after exit from being singlehood are independent of the events up to exit. Therefore, their probability is independent of the likelihood of being singlehood. The event immediately following type i 's singlehood duration is the realization of whom to match with. This event is given by the density of accepted type, $f(x_j|x_i) = f(x_j|j \in A_i)$. Let N_j be the number of type j agents and $\sum_j N_j I(j \in A_i)$ be the number of potential partners acceptable to a type i man, where $I(j \in A_i)$ is an indicator function equaling one if a type j woman is acceptable to a type i man. The acceptance criterion of a type i man is endogenously determined by solving (3) and (4). Given a type i man, the probability that the type i man matches with a type j woman is the number of type j women out of all types of women acceptable to a type i man,

$$f(x_j|x_i) = \frac{N_j I(j \in A_i)}{\sum_{j=1}^J N_j I(j \in A_i)}. \quad (6)$$

Conditional on the realized partner type, marriage duration T_1 has an exponential

distribution with parameter δ . If $D_{0f} = 1$, I do not follow the individual any longer. Let $D_2 = 1$ if T_1 be right-censored, and equal zero otherwise. If $D_{0f} = 0$, then a type i individual's likelihood contribution to events between entering marriage and separation equals

$$L_{1ij} = f(x_j|x_i)\delta^{1-D_2} \exp(-\delta T_1), \quad (7)$$

where $T_1 > 0$. The total type i individual likelihood contribution for a respondent who is single at the time of the first interview equals the product of (5) and (7) that describes the odds of each type i man who initially is single, matching with a type j partner with a marriage offer, followed by the marriage dissolving exogenously:

$$L_{ij} = L_{0i}L_{1ij}^{(1-D_{0f})}. \quad (8)$$

Since observations of each type of men are independent, if n denotes the n -th observation of men, the likelihood function of the benchmark model is

$$L_b = \prod_{n \in (i,j)} L_{ij}, \quad (9)$$

where $i = 1, \dots, J$, and $j = ., 1, 2, \dots, J$.

Before launching the likelihood estimation, two operations must be performed. The first is to rank an individual's type based on observed data. The second is to obtain the analytical acceptance set for each type of individual based on the model. Therefore, the likelihood function must nest within it a matching algorithm to solve for the acceptance set of each type of individual based on (3) and (4).

3.2 The Marriage Index

Types are marriage indices of individuals. To estimate the model in a manageable way, I assume types are generated by two measured attributes: the individual's logarithm of wage w and education e .¹³ Then I rank individuals into discrete categories according to the following steps:¹⁴

- (a) Generate z as a function of w and e

$$z = \exp[\alpha w + (1 - \alpha)e], \quad (10)$$

where α is a scalar parameter. α measures the sensitivity of spousal demand of a change in wages. If spousal demand is sensitive to a change in wage, then wage reflects a more important marriageable characteristic relative to education.

- (b) Take the range of the corresponding order statistics of z and discretize it into J equal partitions, where J is a positive finite integer. For example, if $J = 10$ and

¹³Classification errors are introduced in subsection 3.4.

¹⁴Since the procedures of constructing "type" to each sex is the same, I drop the gender subscript.

the sample size is N , $z_{(1)} \leq \dots \leq z_{(N)}$ is partitioned into decile. Within the i -th interval, z is bounded by $z_{Li} < z \leq z_{Hi}$, where z_{Li} indicates the lowest z that makes a type i individual, and z_{Hi} indicates the highest z that makes a type i individual.

(c) The set of z within each i -th interval is mapped to x_i following the rule:

$$x_i = \text{median} [z_{Li} < z \leq z_{Hi}]. \quad (11)$$

The real-valued x_i represents a type i individual, which is a piece-wise constant within the corresponding i -th interval of $z_{(1)} \leq \dots \leq z_{(N)}$. Thus, given α , w , and e , $x_i(\alpha)$, $x_j(\alpha)$, and the corresponding empirical type distributions are generated.

3.3 The Matching Algorithm

Equilibrium matching sets are endogenously determined in the model. All parameters $(\lambda, \delta, \alpha)$ are assumed to be the same between men and women, but type distributions need not be the same because of gender difference in observed attributes. Given a vector of parameter values $(\lambda, \delta, \alpha)$, a matching algorithm is proposed to compute the acceptable pool of partners for each type of individual for the matching model. The difficulty in solving for the reservation type and the highest attainable type for each type of individual involves constantly updating the acceptance criteria from other individuals. The acceptable pool of partners can be identified by solving the equilibrium acceptance set backward, starting from the highest type from each gender side.

The Matching Algorithm

Step 1 For the highest type of men and women, set the maximum-attainable type to J .

Step 2 Use equations (3) and (4) to solve for R for the highest type of men and women respectively. $\{R_{i=J}, M_{i=J}\} \times \{R_{j=J}, M_{j=J}\}$ defines the first acceptance area.

Step 3 For any i -th type individuals, where $i < J$, two cases can occur:

Case 1. M_i is identifiable. This occurs when $\{j | R_j \leq x_i\}$ is not empty for some $j > j'$. In this case, i is accepted by some $j > j'$. We set $M_i = \max_{j > j'} \{j | R_j \leq x_i\}$ and solve for R_i . Repeat for the women's side.

Case 2. M_i is unidentifiable. This occurs when $\{j | R_j \leq x_i\}$ is empty for a $j > j'$. In this case, we reverse the role of i and j and solve for $\{R_j, M_j\}$ aiming at determining the acceptance sets of additional $j = j - 1, j - 2, \dots$, until the first woman accepts the i -th man, i.e., $R_j \leq x_i$. However, to solve for $\{R_j, M_j\}$, we need to be able to determine $M_j = \max_{i > i'} \{i | R_i \leq x_j\}$, which may not be possible if the set is empty. If we hit an empty set of $\{i | R_i \leq x_j\}$ before obtaining $R_j \leq x_i$, reset M_j to x_i and M_i to x_j at the point of failure and repeat step 2 again.

Stopping Criterion: We have solved for all acceptable sets of partners for type i and j . If there are remaining types of i or j that have not been matched, they will

be assigned a null acceptance area (no matching possible.)

3.4 Classification Errors

Wage and education alone may not be the only traits that matter in mate choice. Other traits, whether observed or not by the econometrician, may be important. One way to handle this issue is to get more information, but it can be costly. Alternatively, I use a classification error model.¹⁵

The classification error model proposed here is simple and flexible. Let k and l denote the true type for i and j respectively. Let $q_{ik} = f(x_i|x_k)$ be the density of classification error for a type i agent, and $q_{jl} = f(x_j|x_l)$ be the density of classification error for a type j agent. Because classification is assumed to be exhaustive and mutually exclusive, the sum of the probabilities for a given observed category is 1, i.e., $\sum_k q_{ik} = 1$. Let \mathbf{Q} denote a $J \times J$ matrix of the probability of all possible classification errors, with elements q_{ik} . The exhaustive assumption forms a total of J restrictions on possible values for q_i . Thus, \mathbf{Q} has a total of $J(J-1)$ free parameters.

To reduce the parameter space, I consider the problem of classification errors in terms of the distance between true and observed type in the following way. To illustrate, let $d(k, i) = |k - i|$ be the distance between the true type k and the observed type i , and $q(d)$ denotes the classification error with distance equals to d . I make two assumptions.

Assumption 1. The probability of misclassifying an individual is the same for any i and k with the same distance, $q(d) = q(d')$ for any $|k - i| = |k' - i'|$.

Assumption 2. The error distributions are the same for both sexes and are independent of sex.

Assumption 2 indicates that the density of misclassifying the true types of a pair of individuals (k, l) as (i, j) can be written as $f(x_i, x_j|x_k, x_l) = f(x_i|x_k)f(x_j|x_l) = q_{ik}q_{jl}$.

In what follows, I propose a flexible model to estimate the distribution of classification errors q_{ik} and q_{jl} , and then present the likelihood function of the censored

¹⁵From a technical standpoint, a classification error model is needed. The reason is that (6) indicates estimates must accommodate the selection constraint $I(j \in A_i) = 0$, otherwise, the contribution to the likelihood function (9) is zero. Imposing such a condition distorts parameter estimates, and estimation from the model may yield misleading conclusions. Consider a marriage market in which most people match with partners of similar types, except a few serious outliers, e.g., the highest type matches with the lowest type. Then, parameter estimates must accommodate the matching of extreme types, and the matching model implies that a sufficiently low partners arrival rate λ is necessary to sustain all couplings. Thus, instead of obtaining estimates that support sorting of partners of similar types, estimation results would yield estimates that produce acceptance probability equal to one, search friction would be erroneously high, and assortative matching would be erroneously low.

model.¹⁶

Consider the number of types being ten, $J = 10$. The following table represents \mathbf{Q} , in which the columns represent observed type i , the rows represent true type k , and the entries represent $q(d)$

		type i									
		1	2	3	4	5	6	7	8	9	10
type k	1	$q(0)$	$q(1)$	$q(2)$	$q(3)$	$q(4)$	$q(5)$	$q(6)$	$q(7)$	$q(8)$	$q(9)$
	2	$q(1)$	$q(0)$	$q(1)$	$q(2)$	$q(3)$					
	3	$q(2)$	$q(1)$	$q(0)$							
	4	$q(3)$	$q(2)$.						
	5	$q(4)$	$q(3)$.					
	6	$q(5)$.				
	7	$q(6)$.			
	8	$q(7)$.		
	9	$q(8)$.	
	10	$q(9)$									$q(0)$

If types are observed without errors, \mathbf{Q} is a diagonal matrix with elements equal one, i.e., $q(0) = 1$ and $q(d) = 0$, for $d > 0$. Since there are at most 10 errors: $q(0)$, $q(1)$, ..., $q(9)$, the parameter space is reduced to 9. In general, this restriction reduces the number of free parameters from $J(J - 1)$ to $(J - 1)$. Note that the exhaustive condition gives rise to the following equality,

$$\mathbf{A}\mathbf{q} = \mathbf{1}, \quad (12)$$

where \mathbf{A} represents a 10×10 matrix containing the number of occurrence of error probabilities associated with each element of a 10×1 vector $\mathbf{q} = (q(0), q(1), \dots, q(9))^T$, and $\mathbf{1}$ is a 10×1 unit vector. In other words, (12) can be written as

$$\mathbf{A}\mathbf{q} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 2 & 2 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 2 & 2 & 2 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 2 & 2 & 2 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 2 & 2 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \times \begin{pmatrix} q(0) \\ q(1) \\ q(2) \\ q(3) \\ q(4) \\ q(5) \\ q(6) \\ q(7) \\ q(8) \\ q(9) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

¹⁶The error structure is assumed to be symmetric between men and women, and so parameters of the classification errors are identical for men and women. The forthcoming discussions on q_{ik} apply to q_{jl} .

Clearly, rank \mathbf{A} is $5 < J = 10$. The solution to \mathbf{q} is:

$$\mathbf{q} = \begin{pmatrix} 1 - 2e_1 - 2e_2 - 2e_3 - 2e_4 - e_5 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_4 \\ e_3 \\ e_2 \\ e_1 \end{pmatrix} \quad (13)$$

However, not any solution of e_1, \dots, e_5 is admissable. The solutions for e_1, \dots, e_5 have to be chosen such that the following two conditions must be satisfied: (a) $e_1, e_2, e_3, e_4, e_5 > 0$, and (b) $2e_1 + 2e_2 + 2e_3 + 2e_4 + e_5 \leq 1$.

Formally, let the classification errors for type i and j agents be denoted as v_1 and v_2 respectively such that $i = k + v_1$ and $j = l + v_2$. Further, let the density of classification error for type i and j agents be $f(x_i|x_k) = q(|v_1|)$ and $f(x_j|x_l) = q(|v_2|)$ respectively. For all sample of men, the likelihood function adjusted for classification errors, satisfying (13), (a), and (b), is

$$L_c = \prod_{n \in (i,j)} \prod_{v_1=i-M_j}^{i-R_j} \prod_{v_2=j-M_{(i-v_1)}}^{j-R_{(i-v_1)}} L_{0(i-v_1)} \left(L_{1(i-v_1)(j-v_2)} q(|v_2|) \right)^{(1-D_{0f})} q(|v_1|), \quad (14)$$

where $i = 1, \dots, 10$, $j = 1, 2, \dots, 10$, $M_i = 10$, and $R_i = 1$.

4 Data

4.1 The Sample

To estimate the matching model, panel data with marriage history is used. I consider individuals who transited from being single to married between 1968 and 1993, and follow their marriage histories thereafter. Specifically, I use data on the age at first marriage, the couple's wages and education at first marriage, and the duration of marriage. Although longer marriage histories are available, I use only the first two spells.

I based my findings on PSID 1968-1993 family and individual samples.¹⁷ The family files do not contain marriage history of the respondents and the individual file

¹⁷ Respondents have been interviewed annually since 1968. Although the PSID also has data for

does not contain detailed demographic and employment data. I first use the individual file to create an eligible sample population with well-defined marriage history, and then link the sample to family files to obtain for each household head, and spouse if married, their corresponding demographic and employment data.

The 1968-93 individual file contains 53013 respondents ever interviewed with their corresponding marriage histories, including information on when first marriages started and ended if this happened. I exclude individuals with inconsistent marriage histories, e.g., those with uncertain first marriage years or marriage termination years, or those who ended the first marriage before their first marriage started. Imposing this restriction, only 27741 respondents are left.

I keep samples whose first marriages began in or after 1968 and lasted until or after 1985. For those individuals whose first marriages began before 1968, there was no information on the first age at marriage or the corresponding marriageable characteristics. The race of household wives were not available in the PSID before 1985. Thus, for couples whose first marriages dissolved before 1985, there was no race information. This selection leaves me with 20441 observations. The PSID oversamples Hispanics. After excluding the Latino sample, 15193 observations are left.

A substantial fraction of the original sample is not present after 25 years of survey.¹⁸ Attrition occurred as a result of loss of contact as families moved, maturing children could not be traced, or respondents refused to continue to be interviewed. Observations are not used for single respondents if their marriage histories were last updated before 1993. For married couples, observations are not used when marriages were not updated in 1993 but first marriages lasted after 1993. Imposing this restriction, the sample population is reduced to 11141. This sample contains individuals from the main family and the subfamily of each household.

To create data linking members of a couple, I use information on the household head and spouse (or cohabitor), because each family file only contains demographic and employment data for household heads (and spouses or cohabitators,) not respondents from the subfamilies. Thus, I further restrict the marriage sample so that both household heads and spouses (or cohabitators) were present. In addition, only one fifth of the sample contains the marriage histories of single female heads who transitioned to marriage, and so I focus on men's marriage histories. This leaves me a total of 2855 households, of which 1517 are single and 1338 married couples.

Search duration can only be partially observed because the elapsed singlehood du-

1994 to 1996 in early release versions, an individual file cannot be accessed in a cross-year way as in the 1968-1993 file, and thus individuals from year to year cannot be followed.

¹⁸By 1993, only 122 out of 1181 households came from the 1968 sample, 20 percent of which were married. Due to the small sample size, I choose the 1968-1993 sample that included those born after 1968 who become 15 or older in 1993. Duncan et al. (1991) documents the representativeness of the PSID after 17 years from 1968. They find that there is a serious problem of attrition and most of the original households are not represented by respondents in 1968. However, the samples still have comparable mean characteristics to these in 1968.

ration T_{0b} is unknown.¹⁹ In what follows, I use 15 as an index for spousal search starting time.²⁰ Therefore, if T^* is the stopping time of being single and C is the censoring time, the completed spell of search duration is $T_0 = T_{ob} + T_{of} = \min\{T^*, C\} - 15$. The duration of marriage is defined as the number of years a couple stays married before or until the censored time, whichever comes first. I do not use samples with uncertain race and age records, nor when respondents have ever been institutionalized. This leaves a total of 633 married couples and 1264 single individuals.

Data concerning wages and education of respondents (and their spouses if married) are taken as of the year of their marriages.²¹ I assume that these characteristics are time invariant. I cross-check all time and pay rate responses against the upper and lower bounds collected for men and women of the same education from the Current Population Survey for the period.^{22,23} Those whose wages do not fall within the admissible ranges are treated as missing. This gives me the final total of 443 married couples and 696 single individuals.

4.2 Data Description

There are five race groups in the sample: whites, blacks, Chinese, Japanese, and American Indians. Due to the small observation of some groups, I focus on white and black men only.²⁴ Table 1 contains sample characteristics on average durations of being singles (uncensored spells only), average wages and education at first mar-

¹⁹The initial condition problem is solved by Chamberlain (1979) using a bayesian technique, in which the random effect distribution is conditioned on forward recurrence information. Ondrich (1985) controls for heterogeneity assuming that both unemployment and employment spells have Weibull Distribution with a parametric unobserved heterogeneity. Results from an exponential model and Cox's Proportional hazard model reveal that there is a significant heterogeneity in the duration of being single in my sample. Heterogeneity in my model is captured by the acceptance selection of each individual type, assuming that $\lambda > 0$.

²⁰Age 15 is used because it is the official Census definition for the marriageable age (see Statistical Abstract 1996 for details.) Moreover, my sample does not contain respondents who married at ages younger than 15, and only 2 cases of zero single duration spells, so the choice of 15 as starting age does not seem unreasonable.

²¹I decode the interval data of education in 1968-1974 and 1985-1990, using auxillary relations with the 1980 Annual Demographic March CPS data.

²²I use the 5th and the 95th percentiles from the set of hourly wages for paid hourly workers from the March outgoing rotation groups for each year.

²³Because 27 percent of married women in the sample did not work, I compute potential wages for them using Heckman's two-step procedure. To correct for the selectivity bias, I estimate a participation probit using the standard Heckman procedure. The probit equation contains all variables in the wage equation and the number of children. The wage equation is controlled for by husbands' and wives' ages, ages squared, education, region dummies, city populations, and husbands' wages. The wage equation and the probit equation are estimated separately by race.

²⁴Whites and blacks represent over 98.25 percent of the sample population for men and over 98.42 percent of the sample population for women. However, I use the entire sample for the purpose of generating the marriage index distribution, which depends on the characteristics of the sample population.

riage, average durations of marriage (including censored spells), and the proportion of complete singlehood spells and marriage spells stratified by race. The number of observations N corresponds to uncensored single data, in which all individuals transit into marriage.

	Mean	S. D.	Min	Max
white men (N*=301)				
singlehood duration*	11.45	6.05	2	41
marriage duration	7.88	5.72	1	25
proportion of complete singlehood spells	.6044	-	-	-
proportion of complete marriage spells	.1196	-	-	-
wage	461.1	234.4	106.0	1334.0
education	13.1	2.03	7	18
wife's wage	315.8	160.7	80.1	1095.9
wife's education	12.9	2.16	3	18
wage correlation	.500	-	-	-
education correlation	.427	-	-	-
black men (N*=133)				
singlehood duration*	12.56	5.80	2	34
marriage duration	6.90	5.41	1	21
proportion of complete singlehood spells	.5195	-	-	-
proportion of complete marriage spells	.0752	-	-	-
wage	432.5	219.5	113.1	1210.1
education	12.4	1.96	7.32	18
wife's wage	325.4	143.5	67.5	861.0
wife's education	12.5	1.84	7	18
wage correlation	.453	-	-	-
education correlation	.389	-	-	-

Note: * denotes uncensored spells only.

Table 1. The Sample Means of the PSID, 1968-1993.

The data reveal that on average, white men take a shorter time (1.11 years) to search for spouses than black men. The duration of marriages for white men is longer. The proportion of the complete (or interrupted) marriage spell for white men is 11.2 percent, which is 3.7 percent higher than that for black men. Wage and education data tell the conventional story: blacks earn less and have lower education levels than whites. Matchings of whites are more assortative in wage and education than those of blacks.²⁵

²⁵Logarithms of wages are used to compute the wage correlation to adjust for skewedness. When the imputed wage for women is not used, the correlation is 0.472 and 0.376 for white men and black men respectively.

5 Results

Estimation results are based on categorizing agents into ten types, i.e., $J = 10$, and setting $\beta = 0.05$. My estimation strategy is applied to three specifications of α , in order to separate out the contributions of wage and education in explaining who matches with whom: $\alpha = 1$, 0 , and $(0, 1)$. Tables 2 and 3 show the estimated results for white and black men respectively.

When wage is used as the marriage index (row 6 in tables 2 and 3), white men are about 4.5 times more likely than black men to contact partners.²⁶ The separation rate is also higher for white men. The marriage market for white men is more than four times as flexible (24.24 versus 5.87) as that for black men, indicating more assortative matching for white men. This result is consistent with data showing that white men had higher correlations in wage with their spouses than black men. When using education as the marriage index (row 11 in tables 2 and 3), white and black men exhibit similar stratification patterns because of similar search efficiency.

Once classification error is accounted for, matching is more assorted using wage as the only marriageable trait rather than education, except for the age group 41-65.²⁷ For example, the marriage market flexibility for white men equals 2.4 when $\alpha = 0$ and 14.23 when $\alpha = 1$. This result contrasts with the simple wage correlation in subsection 4.2 that suggests spouses are sorted similarly in wage and education. This result and the significance of classification error probabilities indicate the empirical importance of accounting for classification errors.

When using both wage and education (row 16), a significant racial contrast in the relative importance of wage occurs: 0.816 and 0.031 for white men and black men respectively.²⁸ This indicates that wage has a greater impact on white men's desirability as marriage partners, whereas education constitutes almost all marriageability for black men. Alternatively, as wage represents temporal income and education represents permanent income, α can be interpreted as the time rate of preference. The

²⁶I am aware of the systematic difference in wage during life cycle. I reestimate the model using standardized wage (by mean age); and I also use the present value of expected income as agents' trait. I find that there is no qualitative difference in results.

²⁷Results from a benchmark model (without classification errors) show that in all α specifications, black men have lower partners' arrival and separation rates than white men. The estimates of λ for black and white men are 0.039 and 0.034 respectively when $\alpha = 1$, and 0.032 and 0.027 respectively when $\alpha = 0$. So the acceptance sets for both race groups range from the lowest to the highest type. That is to say, sample individuals scramble for partners, and the acceptance probability is one. The values of market efficiency are found to be close to zero, indicating that search friction is high.

²⁸Results on the estimates of α may be affected by using imputed wages for non-working women (27 percent in the sample). Because people may self-select to not working, assigning them positive wages leads to an over-estimation of α . So, the estimated explanatory power of wage may be too large. Even if only agents of positive wages are included, over-estimation of α would still occur and wage as a marriageable trait tends to predict more assortative mating. However, results from Monte Carlo simulations indicate that α is under-estimated. So, the net effect of over- and under-estimation of α may cancel out.

high α of white men implies that white men are more impatient than black men, which is consistent with the result of a higher $\frac{\lambda}{\beta+\delta}$ that leads to a more efficient market for black men.²⁹ Likelihood ratio tests reject specifications using only one trait.

AGE	α	λ	δ	e_0	e_1	e_2	e_3	e_4	e_5	$\log L$
15-21	1	.751 (.309)	.003 (.002)	.625 (.226)	.091 (.204)	.044 (.052)	.019 (.053)	.020 (.055)	.025 (.053)	-562.27
22-30	1	.716 (.306)	.002 (.002)	.449 (.209)	.116 (.204)	.050 (.051)	.047 (.051)	.044 (.055)	.037 (.052)	-3401.26
31-40	1	.548 (.412)	.001 (.002)	.307 (.209)	.121 (.214)	.077 (.066)	.053 (.054)	.066 (.055)	.057 (.055)	-1175.17
41-64	1	.175 (.378)	.009 (.003)	.248 (.209)	.075 (.205)	.087 (.059)	.080 (.052)	.086 (.061)	.095 (.055)	-296.42
all	1	.740 (.200)	.002 (.001)	.449 (.118)	.117 (.046)	.063 (.021)	.041 (.020)	.036 (.020)	.035 (.020)	-5454.00
15-21	0	.261 (.102)	.011 (.009)	.424 (.203)	.078 (.202)	.062 (.024)	.058 (.072)	.060 (.073)	.059 (.073)	-787.43
22-30	0	.168 (.108)	.018 (.010)	.306 (.201)	.096 (.201)	.062 (.021)	.079 (.071)	.072 (.071)	.076 (.072)	-3948.33
31-40	0	.285 (.251)	.005 (.010)	.270 (.318)	.120 (.209)	.064 (.029)	.066 (.075)	.081 (.074)	.068 (.070)	-1214.41
41-64	0	.155 (.198)	.014 (.009)	.224 (.217)	.075 (.204)	.082 (.029)	.094 (.077)	.093 (.073)	.088 (.073)	-273.11
all	0	.156 (.068)	.015 (.001)	.333 (.093)	.098 (.043)	.054 (.004)	.074 (.021)	.068 (.021)	.078 (.028)	-5991.60
15-21	.861 (.216)	1.257 (.507)	.005 (.004)	.700 (.313)	.100 (.217)	.009 (.071)	.013 (.086)	.022 (.087)	.011 (.086)	-504.80
22-30	.826 (.109)	.776 (.462)	.003 (.004)	.508 (.339)	.096 (.217)	.039 (.050)	.044 (.086)	.034 (.087)	.066 (.083)	-3138.98
31-40	.470 (.241)	.736 (.583)	.002 (.004)	.383 (.314)	.134 (.219)	.057 (.064)	.046 (.087)	.042 (.086)	.058 (.085)	-1117.37
41-64	.959 (.201)	.437 (.502)	.002 (.004)	.319 (.330)	.098 (.217)	.078 (.064)	.048 (.086)	.078 (.087)	.075 (.087)	-291.88
all	.816 (.082)	.801 (.208)	.002 (.001)	.493 (.138)	.104 (.043)	.055 (.026)	.023 (.025)	.043 (.020)	.058 (.022)	-5184.16

Notes: $e_0 = 1 - 2e_1 - 2e_2 - 2e_3 - 2e_4 - e_5$, standard error in parentheses.

Table 2. Estimates from the PSID (1968-1993), White Men

To examine whether there exists heterogeneity of search behavior across age groups, I stratify the sample by four age groups: 15-21, 22-30, 31-40, and 41-65. Within each age group, type described by both wage and education is significantly different from models when only wage or education is used. Besides, the marriage market for white men is most efficient when using both traits, while the picture for black men is less clear.

When examining across age groups for $\alpha \in (0, 1)$, wage demonstrates itself as a better representation of white men's marriageability, except for the age group 31-40.

²⁹ Comparisons of these results to those from the benchmark model reveal significant shifts of the estimate of α : from 0.005 to 0.807 for white men, and from 0.005 to 0.031 for black men. The estimates of λ become higher but those of δ smaller in the classification error model. The opposite movements of the estimates of λ and δ allow more assortative matching as described in section 2. When classification errors are incorporated in the model, the transition parameters adjust to allow for more assortative matching.

In contrast, education is a more desirable marriageable trait for black men except those between the ages of 15 and 20 (rows 12-16, table 3).

AGE	α	λ	δ	e_0	e_1	e_2	e_3	e_4	e_5	$\log L$
15-21	1	.372 (.218)	.012 (.018)	.824 (.200)	.082 (.114)	.001 (.057)	.001 (.058)	.003 (.071)	.001 (.004)	-74.12
22-30	1	.865 (.103)	.004 (.008)	.500 (.199)	.152 (.098)	.043 (.055)	.023 (.054)	.028 (.072)	.008 (.002)	-1398.97
31-40	1	.620 (.220)	.008 (.018)	.286 (.224)	.109 (.104)	.085 (.054)	.069 (.053)	.059 (.074)	.072 (.004)	-1025.07
41-64	1	.030 (.220)	.002 (.017)	.161 (.212)	.068 (.103)	.110 (.054)	.095 (.053)	.101 (.074)	.092 (.004)	-164.15
all	1	.317 (.080)	.004 (.001)	.376 (.091)	.103 (.047)	.057 (.023)	.052 (.024)	.065 (.026)	.007 (.000)	-2552.22
15-21	0	.278 (.140)	.017 (.011)	.482 (.205)	.112 (.104)	.048 (.056)	.039 (.052)	.041 (.071)	.038 (.056)	-150.53
22-30	0	.283 (.132)	.005 (.003)	.464 (.207)	.118 (.106)	.061 (.056)	.038 (.051)	.033 (.070)	.036 (.056)	-1255.30
31-40	0	.170 (.140)	.012 (.012)	.327 (.232)	.075 (.102)	.064 (.056)	.083 (.052)	.077 (.071)	.077 (.055)	-678.48
41-64	0	.288 (.141)	.001 (.012)	.478 (.234)	.208 (.102)	.039 (.056)	.005 (.052)	.004 (.073)	.010 (.056)	-100.76
all	0	.205 (.092)	.004 (.001)	.430 (.108)	.099 (.043)	.053 (.025)	.060 (.020)	.051 (.028)	.043 (.026)	-2215.45
15-21	.768 (.136)	.320 (.169)	.014 (.010)	.766 (.314)	.110 (.108)	.002 (.060)	.002 (.061)	.002 (.061)	.003 (.063)	-78.03
22-30	.004 (.102)	.289 (.128)	.006 (.003)	.449 (.208)	.121 (.102)	.054 (.054)	.039 (.057)	.045 (.057)	.032 (.056)	-1258.05
31-40	.040 (.129)	.146 (.155)	.008 (.013)	.317 (.212)	.065 (.107)	.061 (.052)	.089 (.062)	.074 (.061)	.105 (.061)	-673.38
41-64	.209 (.189)	.162 (.156)	.002 (.012)	.565 (.281)	.194 (.117)	.005 (.055)	.001 (.062)	.004 (.061)	.026 (.061)	-72.56
all	.031 (.012)	.205 (.079)	.006 (.001)	.436 (.107)	.094 (.054)	.050 (.029)	.055 (.029)	.054 (.025)	.058 (.024)	-2221.07

Notes: $e_0 = 1 - 2e_1 - 2e_2 - 2e_3 - 2e_4 - e_5$, standard error in parentheses.

Table 3. Estimates from the PSID (1968-1993), Black Men

Interestingly, there is a striking racial contrast in results from the effects of classification errors and market efficiency across age groups. Results for white men show that the chance of being classified incorrectly increases with age. Further, the marriage market of younger white men is more efficient than that for older agents except for $\alpha = 0$ in table 2. However, there is no tractable effect of age for black men. Given a particular specification of α , both e_0 and $\frac{\lambda}{(\beta+\delta)}$ may increase or decrease with age. For example, using wage only as a marriage index reveals little about older black men's marriageability (e_0 is low), but it reveals considerable information about young black agents' marriageability. As education or a combination of wage and education is used as a marriage index, black men of age groups 22-30 and 31-40 are more prone to classification errors. A mixture of results is also found for the marriage efficiency of black men. The market is more efficient at both ends of the age distribution when education is the marriageable trait. But when wage only is used, the market is more efficient for age groups 22-30 and 31-40. At $\alpha \in (0, 1)$, younger black men have a more efficient market; this result resembles that of white men.

On comparing the sum of the maximum log likelihoods by age groups (rows 12 to 15) to the maximum log likelihood using the restricted model (row 16), the hypothesis of homogeneity across subsamples is clearly rejected. This result is found in all α specifications for white and black men.

The estimated acceptance sets for white and black men are shown in figures 1-6. The data are generated following (11) and are based on the estimates in rows 6, 11, and 16 of tables 2 and 3. From the specifications of match production of $x_i x_j$ and sexual symmetry in parameters α and λ , disjoint “marriage classes” in the spirit of Burdett and Coles (1999, 1997) are expected. When wage is the only determinant of the marriage index, there are three marriage classes and they are asymmetric (figure 1). The asymmetry is due to the substantial difference in wage distribution between white men and women. White men’s types who are type five or above accounts for more than five percent from the sample, while it is about 3 percent for women. Because there is more high type white men than high type women, high type women are more picky than men. Women of type five through type ten have acceptance sets: $A_{j=5} = \{i|i = 5, \dots, 10\} = A_{j=6} = \dots = A_{j=10}$, while white men of type five through type ten have acceptance sets: $A_{i=5} = \{j|j = 4, \dots, 10\} = A_{i=6} = \dots = A_{i=10}$. These acceptance sets form the first marriage class. The second marriage class is made up of type two and three women and type three and four men. The last marriage class is made up of type one women and type two men. Because women are picky, type one men remain single in this equilibrium.

Because search friction is highest when education is used as the marriage index, white men are the least sorted: type three through type ten agents share the same acceptance set, and types one and two agents are sorted with their own type (figure 2). When both wage and education are considered, there are four marriage classes for white men. Men and women of type four through type ten are in the same class, and strict sorting occurs in the rest of the three types (figure 3). Because the market is most flexible in this case, agents are most sorted.

Given a higher search friction, the marriage market for black men is less sorted. In all α specifications, similar offer arrivals lead to same acceptance sets of black men, with three marriage classes in each model (figures 4 to 6).

The matching model posts strong assumptions on the distribution of singlehood and marriage spells: both follow exponential distributions with intensity parameters $\lambda(\gamma^m - \gamma^r)$ and δ respectively.³⁰ To check how well the exponential model fits the data, I perform a formal test by fitting a Weibull model to the duration data, the purpose is to test the slope of the shape parameter ρ in the Weibull model. Under the null hypothesis of an exponential model, $\rho = 1$.

³⁰The transition rate to marriage is obtained after conditioning out unobserved heterogeneity in types.

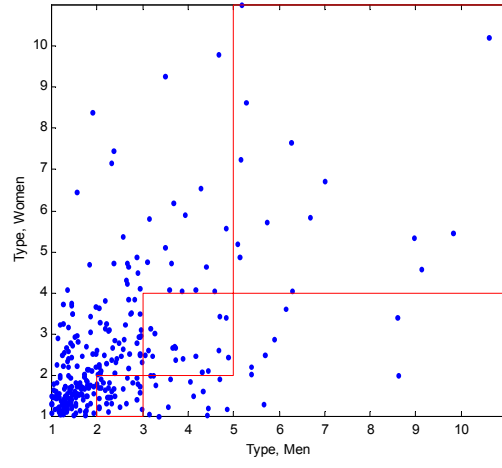


Figure 1: White Men's Acceptance Sets, $\alpha = 1$

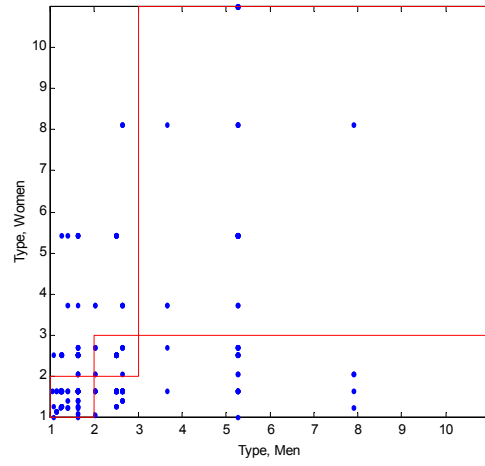


Figure 2: White Men's Acceptance Sets, $\alpha = 0$

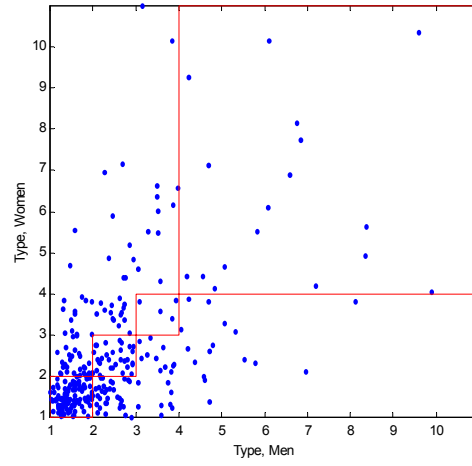


Figure 3: White Men's Acceptance Sets, $\alpha = 0.816$

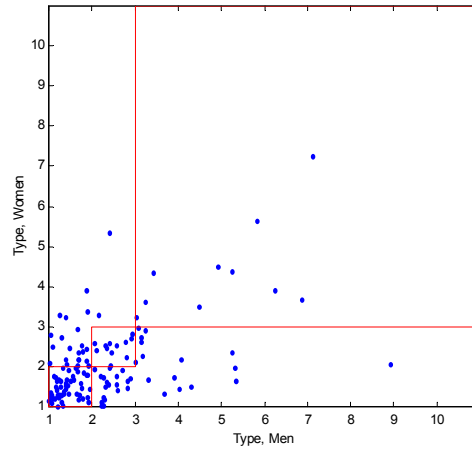


Figure 4: Black Men's Acceptance Sets, $\alpha = 1$

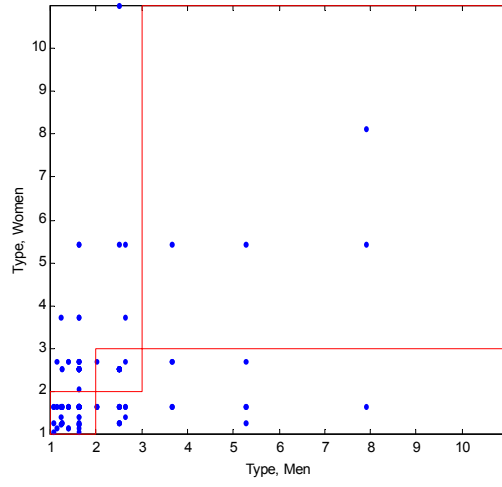


Figure 5: Black Men's Acceptance Sets, $\alpha = 0$

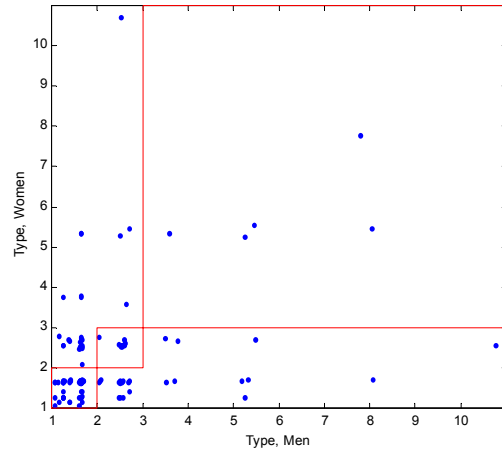


Figure 6: Black Men's Acceptance Sets, $\alpha = .031$

	White Men			Black Men		
	ρ	95% LL	95% UL	ρ	95% LL	95% UL
Singlehood Spells						
$\alpha = 1$	0.372	0.332	0.412	0.718	0.630	0.806
$\alpha = 0$	1.791	1.641	1.941	1.145	0.995	1.295
$\alpha = (0, 1)$	0.354	0.314	0.394	1.145	0.995	1.295
Marriage Spells						
$\alpha = 1$	0.578	0.492	0.664	0.780	0.590	0.970
$\alpha = 0$	1.053	0.893	1.213	0.780	0.590	0.970
$\alpha = (0, 1)$	0.578	0.492	0.664	0.874	0.658	1.090

Note: LL represents lower limit, UL represents upper limit.

Table 4. Specification Tests For Exponential Search Times, PSID (1968-1993)

Results are shown in table 4. The exponential model fits the singlehood spells of black men quite well (columns 4 to 6). The shape parameter in singlehood spells is fairly close to one and the asymptotic confidence interval straddles 1.0 for $\alpha < 1$. However, the model has difficulty in fitting black men’s marriage spells and white men’s singlehood and marriage spells data. Results for black men’s marriage spells show that the standard errors of ρ are enormous. These results raise suspicion of the model fitness even though the asymptotic confidence interval covers one when $\alpha = (0, 1)$ and it is right around one to two decimal points at the corner cases (columns 4 to 6). The estimates for singlehood and marriage spells for white men are around 0.4 and 0.6 respectively for $0 < \alpha$, exhibiting decreasing singlehood and marriage hazards (columns 1 to 3). These results reflect that singlehood (marriage) tenure is negatively associated with the marriage (separation) hazard. Such duration dependence may be spurious, and unobserved heterogeneity may be required to improve the fit of the model. Another way to improve the fit to marriage spells data is to discard the assumption of the exogenous separation as the only match termination mechanism. Exogenous separation can underestimate the transition rate from marriage to separation. To obtain a reasonable estimate of the transition rate of separation, one may need to extend the model to introduce endogenous separation.

6 Conclusion

This paper is the first step in an attempt to examine empirical issues involved in structural estimation of a marriage model. The approach requires estimation of agents’ types and their reservation types, as a means of determining the deeper behavioral parameters. Classification errors will affect these estimates. My approach in this paper retains the behavioral matching model while adding classification errors. An application to data from the PSID shows that the model can be interpreted straightforwardly. Results reject the use of a single trait in determining agents’ types, and favor the use of a mixture of measured traits and classification errors. White men

are found to be more impatient and their marriage market more flexible than black men. Despite that wage is found to be a more desirable marriageable trait than education for white men, its desirability decreases with age. The results indicate no tractable effects of age for black men. These results show that the improvement in understanding who marries whom is gained by adopting structural estimation.

The advantage of the proposed method to account for unobserved heterogeneity is that it is flexible. An alternative way is to consider stochastic match production. But this might introduce a substantial computation burden in solving the dynamic programming problem. Although this paper provides an intuitively appealing method to estimate a matching model, it is clear that the separation process assumed by the model is not realistic. The exogenous match destruction does not incorporate shocks that lead to destruction. More complex models to incorporate endogenous match destruction may be required for future research.

References

- Becker, Gary. "A Theory of Marriage: Part I." *Journal of Political Economy*, 81 (1973): 813-846.
- Bergstrom, Ted and Robert Schoeni. "Income Prospects and Age-At-Marriage." *Journal of Population Economics*, 9 (1996): 115-130.
- Bontemps, Christian, Jean-Marc Robin, and Gerard van den Berg. "An Empirical Equilibrium Job-Search Model with Search on the Job and Heterogeneous Workers and Firms." *International Economics Review*, 40 (1999):1039-1075.
- Boulier, Bryan and Mark Rosensweig. "Schooling, Search, and Spouse Selection: Testing Economic Theories of Marriage and Household Behavior." *Journal of Political Economy*, 92 (1984): 712-732.
- Bloch, Francis and Harl Ryder. "Two-sided Search, Matching, and Match-makers." Typescript, Brown University (1998).
- Bunzel, Helle, Bent Christensen, Peter Jensen, Nicholas Kiefer, Lars Korsholm, Lars Muus, George Neumann, Michael Rosholm. "Specification and Estimation of Equilibrium Search Models." Working paper 97-08, Center for Labour Market and Social Research (1997).
- Burdett, Kenneth and Melvyn Coles. "Long-Term Partnership Formation: Marriage and Employment." *The Economics Journal*, 109 (1999): F307-F334.
- Burdett, Kenneth and Melvyn Coles. "Marriage and Class." *Quarterly Journal of Economics*, 112 (1997): 141-168.
- Chamberlain, Gary. "Heterogeneity, Omitted Variable Bias and Duration Dependence." Harvard Institute of Economics Research, discussion paper #691 (1979).
- Collins, E. and J. McNamara. "The Job Search Problem as an Employer-Candidate Game." *Journal of Applied Probability*, 28 (1990): 815-827.
- Duncan, Greg and Daniel Hill. "Assessing Quality of Household Panel Data: the Case of the Panel Study of Income Dynamics." *Journal of Business and Economic Statistics*, 7 (1989): 441-452.
- Gale, David and Lloyd Shapley. "College Admission and the Stability of Marriage." *American Mathematical Monthly*, 69 (1962): 9-15.
- Keller, Matthew, Del Thiessen, and Robert Young. "Mate Assortment in Dating and Married Couples." *Person Individual Difference*, 21 (1996): 217-221.
- Kiefer, Nicholas and George Neumann. "Wage Dispersion with Homogeneity: The Empirical Equilibrium Search Model." in H. Bunzel, P. Jensen, and N.C. Westergaard-Nielsen, eds., *Panel Data and Labour Market Dynamics*. New York: North-Holland (1993): 57-74.
- Montgomery, Mark and Donna Sulak. "Female First Marriage in East and Southeast Asia: A Kiefer-Neumann Model." *Journal of Development Economics*, 30 (1989): 225-240.
- Mortensen, Dale and George Neumann. "Estimating Structural Models of Unemployment and Job Duration." *Proceedings of the Third International Symposium in Economic Theory and Econometrics*. Cambridge University Press (1988): 335-355.

- Oi, Walter. "Heterogeneous Firms and the Organisation of Production." *Economic Inquiry*, 21 (1983): 247-270.
- Ondrich, Jan. "The Initial Conditions Problem in Work History Data." *Review of Economic Statistics*, 67 (1985): 411-421.
- Pencavel, John. "Assortative Mating by Schooling and the Work Behavior of Wives and Husbands." *American Economic Review*, 88 (1998): 326-29.
- Rosen, Sherwin. "The Economics of Superstars." *American Economic Review*, 71 (1981): 845-858.
- Roth, Alvin and Marilda Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press (1990).
- Smith, Lones. "The Marriage Model with Search Frictions." Typescript, MIT (1997).
- Spuhler, J.N.. "Assortative mating with respect to physical characteristics." *Social-Biology*, 29 (1982): 53-66.
- Suen, Wing and Hon-Kwong Lui. "A Direct Test of the Efficient Marriage Market Hypothesis." *Economic Inquiry*, 37 (1999): 29-46
- Vandenberg, Steven. "Assortative mating, or who marries whom?" *Behavior-Genetics*, 2 (1972): 127-157.
- Wong, Y. Linda. "A Search Interpretation on Stratification in the Marriage Market." Typescript, University of Iowa (1998).

Appendix.

Derivation of the reservation type

From (1), we have

$$\beta V(x_i) = x_i + \lambda \sum_{j \in A_i} [W(x_i, x_j) - V(x_i)] \gamma_j. \quad (\text{A1})$$

Substitute (2) into A1 to obtain

$$\beta V(x_i) = x_i + \lambda \sum_{j \in A_i} \left[\frac{x_i x_j}{2(\beta + \delta)} - \frac{\beta V(x_i)}{(\beta + \delta)} \right] \gamma_j. \quad (\text{A2})$$

The optimal policy gives $\frac{x_i R}{2(\beta + \delta)} + \frac{\delta V(x_i)}{(\beta + \delta)} = V(x_i)$, which can be simplified to obtain

$$\frac{x_i R_i}{2} = \beta V(x_i). \quad (\text{A3})$$

Substitute A3 into A2,

$$\frac{x_i R_i}{2} = x_i + \lambda \sum_{j \in A_i} \left[\frac{x_i x_j}{2(\beta + \delta)} - \frac{x_i R}{2(\beta + \delta)} \right] \gamma_j. \quad (\text{A4})$$

A4 is equivalent to (3): $R_i = 2 + \frac{\lambda}{\beta + \delta} \sum_{j \in A_i} (x_j - R_i) \gamma_j$.

Spousal Correlations, IPUMS 1960-1980

	wage	education	age
1960	0.3338	0.6160	0.7174
1970	0.3054	0.6063	0.7516
1980	0.3090	0.6277	0.6950

Source: The samples are drawn from the 1960, 1970 and 1980 decennial Census files. I create matched family data for non-institutionalized employed newlyweds, who have married once with current age no more than two years from the first age at marriage, and who have no children. Age is limited to those between 15 and 64, with 15 being the official marriageable age defined by the Census.

Some cares need to be taken in order to obtain meaningful wage correlation between spouses. To reduce measurement errors for wages, wage is trimmed using top-codes provided by the Census and the minimum hourly wage documented by the Bureau of Census. Those with wages that do not fall within admissible ranges are treated as missing. Wages used are weekly wages, and thus the minimum hourly wage is multiplied by 35 hours to get the lower wage bound.