# Structural Estimation of Marriage Models

Ma Xiang-yu

3/14/20

## Introduction

Suppose you meet two men. They're both of the same age cohort. One of the two married at the age of 21; the other at the age of 29. Which of them do you think had an easier time looking for partners? Who do you think would be choosier?

In this paper, I perform a replication (in spirit) of Linda Wong's 2003 paper, "The Structural Estimation of Marriage Models."[1] While doing so, I hope to provide some sort of answer to the question I pose above. Using a structural approach that embeds the numerical solution of a two-sided matching model within a maximum likelihood procedure, I estimate marriage models for two groups of men, (a) those who married before the age of 22, and (b) those who married after the age of 22. I find that those who married young experienced lower search friction during their pursuit of partners. While both groups face similar rates of marriage dissolution, those who married young experienced higher rates of partner arrival. As a result, they tend to be more selective about their choices of partner, and may actually demonstrate higher levels of positive assortative mating.

## Modeling Marriages

Following Wong 2003, this paper uses a structural approach to examine the question of who matches with whom. We use a two-sided matching model that allows for martial sorting in response to marriage market flexibility and agents' preferences. In this section, we begin with a discussion of who these agents are and how their attractiveness is measured. Next, we build a two-sided matching model by making assumptions about the utility structure of singlehood and marriage. Said assumptions about the utility assumptions allow us to construct an an optimal reservation policy that rational agents would follow. The numerical solution of this optimal reservation policy is then nested within a maximum likelihood procedure to estimate the structural parameters. The estimation procedure will be discussed in further detail in the section to come.

### The Agents

We begin by first considering who the agents are. In this paper, we assume that there are two groups of infinite-lived agents: men and women. At each point in time, agents are in either of two states, single or married. Only single agents search for marriage partners. Let the arrival rate of partners for single agents be distributed according to a Poisson distribution with parameter $\lambda$. Given an arrival of partner, an agent decide has to decide whether or not accept a "proposal" from said partner, and form a marriage-pair. Marriage-pairs also dissolve exogenously at rate $\delta$. Upon dissolution, agents flow back to the single pool. If no match is formed, the individual searches again.

[1] I say in spirit because I deviate from the analysis here and there – a full explication of where this paper departs from Wong's is noted in the Conclusion section of the paper.

**Measuring Agent Attractiveness**

Agents are ex ante heterogeneous with respect to their *attractiveness value*, $x_i$. In this paper, I'll assume there to be 5 categories of attractiveness for men and women. $x_i$ reflects the attractivensss value of a person of attractiveness-type $i$ .

Each attractiveness-type represents a discrete category of a marriage index, $z$, of an individual. It reflects an individual's "rank" in the marriage sorting market. In Wong's original construction, $z$ is generated by the individual's logarithm of wage $w$ and education $e$:

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

where $\alpha$ is a scaler parameter. $\alpha$ measures the sensitivity of spousal demand to changes in wages and was a parameter that could be estimated. However, due to a difficulty in processing the PSID education data, this paper proceeds with wage data alone. This effectively equates to a calibration of the $\alpha$ parameter to 1, such that education does not have an impact on attractiveness. $z$ is ordered and then discretized into 5 equal partitions. These are the 5 attractiveness-types. Within each attractive-type interval, $z$ is, bounded by $z_{Li} < z \leq z_{Hi}$, where $z_{Li}$ indicates the lowest $z$ that makes someone an attractive-type $i$ individual and $z_{Hi}$ indicates the highest $z$ that makes someone an attractivenesstype $i$ individual. The set of z within each $i$th interval is then mapped to an attractive value, $x_i$, according to the following rule,

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

Each $x_i$ value represents the attractiveness value of a attractiveness-type $i$ individual. $x_i$ is a piece-wise constant within the corresponding $i$th interval of $z_{(1)} \leq \cdots \leq z_{(N)}$. Thus, $x_i$, $x_j$ and corresponding empirical type distributions for men and women are generated from values of $w$.

## How Agents Match:

Next, we proceed with a model of how agents choose whom to marry. We assume that a man of attractive type $i$ elects to marry (or not marry) with the objective of maximizing his expected discounted value in his future utility stream. While an agent is single, the instantaneous utility is the attractive value of the agent's attractiveness type. When an agent is married, instantaneous utility is assumed to be an equal split of match production $x_i x_j$, where $x_i$ represents the type value of agent $i$ and $x_j$ the type value of another agent $j$. Upon a meeting, the types of the two potential partners are revealed to each other, and each agent decides whether or not to proceed with marriage. $\gamma^i$ denotes the proportion of types among single men who would propose to a woman of match type $j$ if they were to meet, that is, $\gamma^i = \sum_{k=1}^{i} \gamma_k = \Pr(X < x_{i+1}|x_j)$, where $\gamma_i$ denotes the probability mass function of attractive type $i$.

If agents discount future income at rate $\beta$, then the value of being single is

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

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 comprises 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,

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

An agent of type $i$ considers a marriage proposal *acceptable* if the potential partner is above the value of the reservation type $R_i$. Let $r$ represent the ranking of R, so $R_i = x_{r_i}$. Let $M_i$ be the value of the maximum attainable type for agents of type $i$ and $m$ to 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 of any random match 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.

The reservation type for men is the solution to the following equation:

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

This obtains a unique solution since as $R_i$ increases, the right-hand side term $\sum_{j \in A_i}(x_j - R_i)\gamma_j$ decreases. The reservation for women is the following:

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

These two sets of joint matching strategies for men and women are Nash equilibrium solutions to the noncooperative stationary game of matching. Positive assortative matching is the predicted equilibrium outcome. In this model, an increase in reservation type can be explained by either a rise in the arrival rate of partners ($\lambda$), a fall in the separation rate ($\delta$), or a combination of both. The ratio $\frac{\lambda}{\beta+\delta}$ can be thought of as a measure of the inverse of search friction. As $\lambda$ approaches infinitely, partners arrive instantaneously. Friction disappears, and each agent matches with his or her most desirable type. Strict positive assortative matching results. Conversely when $\lambda$ approaches zero, chances to meet partners become vanishingly rare and agents are at their least selective.

### Two Groups of Agents

In this paper, we compare two sub-populations of agents: those who married young, and those got married older. *Young* as this paper defines it refers to those who experienced their first marriage at the age of 21 and below. This roughly corresponds to the age where an adult might first complete their BA degree. The *married older* group refers simply to those who experienced their first marriage above 21. The guiding hypothesis here is that different processes might be guiding the marriage-matches for each group. Here, we have two competing hunches.

The first hunch is that the first group, those who married young, experience *lower* search friction. The young are typically more likely to be in social contexts where chance romantic meetings happen – the arrival rates are higher. The second hunch is that the those who married young are *less picky* then their older counterparts. They have a decreased proclivity for positive assortative mating. Chalk this up to the romantic idealism of youth, which leads to heterogeneity in tastes; or even more simply to a decrease capacity to evaluate accurately the attractiveness of a partner. These two hunches conflict with one another, however. Lower search friction should lead to greater positive assortative matching. A less strict matching criteria should decrease search friction. We shall use the structural model to see which hunch the data supports.

## Data

This paper draws on longitudinal survey data from the Panel Study of Income Dynamics (PSID, 1968-1993). Specifically, we use data on the age at first marriage, the couple's wages at first marriage, and the duration of marriage. The

Table 1: Sample Summary Statistics

| Variable | Total | Married While Young | Married While Older |
|---|---|---|---|
| No. of Persons | 1400.0 | 567.0 | 833.0 |
| No. of Men | 647.0 | 200.0 | 447.0 |
| No. of Women | 753.0 | 367.0 | 386.0 |
| No. of Married | 931.0 | 345.0 | 586.0 |
| Mean Age Married | 24.0 | 19.5 | 27.0 |
| Mean Wage at Marriage | 3.7 | 2.6 | 4.4 |
| Mean single-time | 9.0 | 4.5 | 12.0 |
| Mean married-time | 19.1 | 18.1 | 19.8 |

data come from the PSID's family and individual samples. The family files do not contain the marriage history of the respondents and the individual files do not contain the detailed sociodemographic attributes required. We use the individual file to create an eligible sample population. A person is included within this eligible sample population if they are (a) either a household head, or a partner of the household head, (b) were married between 1968 and 1993. The total eligible sample comprises 1400 individuals, of whom 647 are male and 753 are female. Further sample statistics can be found below in Table 1.

# Estimation Strategy

As mentioned before, the estimation procedure here involves an analytical solution embedded within a maximum likelihood procedure. Given estimates of $$, we solve for the optimal matching sets for men and women. Using the solutions to the optimal matching sets, we calculate the likelihood of an attractive-type $i$ agent marrying a type $j$ agent. We choose the parameter values that maximizes the sum of the log likelihoods for all men in our sample population.

## Determining Matching Sets

Equilibrium matching sets are endogenously determined in the model using an analytical procedure. All parameters $(\lambda, \delta, \alpha)$ are assumed to be the same between both men and women. Given a vector of parameter values, a matching algorithm is proposed to compute the acceptable set of partners for each type of individual for the matching model. The acceptance set of partners can be identified by solving the equilibrium acceptance set in descending of types, starting from the higher type from each gender.

There are three steps to the procedure. To adumbrate:

1. We start with the highest type of men ($i$) and women ($j$). For these individuals, we set the maximum attainable type to $J = 5$.
2. We solve for $R$ for men and women using the equation above. $\{R_{i=5}, M_{i=5}\} \times \{R_{j=5}, M_{j=5}\}$ defines the first acceptance area.
3. We proceed with the remaining $i$th types. For any $i$th individual, where $i < 5$, two cases can occur. First, let's consider the case where $M_i$ is identifiable. This occurs when $\{i|R_j \leq x_i\}$ isn't empty for some $j > j'$, i.e. $i$ is accepted by some \$ j >j'\$. Less formally, this means that there exists some individuals of type $j$ of the opposite gender, whose reservation type is smaller than the match type of individual $i$, who *would* accept a marriage proposal of an individual of type $i$. In such cases, we set $M_i = \max_{j>j'}\{j|R_j \leq x_i\}$ and solve for the reservation type for the men's side, $R_i$. Then, we repeat for the women's side. Alternatively, $M_i$ may be unidentifiable. This occurs when $\{j|R_j \leq x_i\}$ is empty for a $j < j'$. In such cases, we reverse the role of $i$ and $j$ and solve for $\{R_j, M_j\}$, aiming at determining the acceptance sets of additional $j = j-1, j-2\cdots$, until the first woman to accept the $i$th man, that is, $R_j \leq x_i$. 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 $x_j$ and repeat step 3.

## Determining Likelihood

With solutions to the optimal matching sets in hand, we proceed to calculate the likelihood of our observing the sample. But what likelihood more precisely? We're calculating the likelihood of a person's period of singlehood, and of his marriage to chosen partner. Since all couples are single at the beget (this is by construction/selection), we are able to obtain information on the duration of singlehood. Let $T_{0b}$ be the time single prior to the first interview in the PSID, and $T_{0f}$ be the time spent single post interview in the PSID, the total duration of single is $T_0 = T_{0b} + T_{0f}$.

Both $T_{0b}$ and$T_{0f}$ are assumed to be independently and identically distributed with an exponential distribution with parameter $\lambda(\gamma^m - \gamma^r)$. Conditional on being type $i$, the individual contribution of singlehood duration until and including the time of exit into marriage is:

$$L_{0i} = \lambda(\gamma^m - \gamma^r) \exp[-\lambda(\gamma^m - \gamma^r)(T_{0b} + T_{0f})],$$

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

Marriage is independent of singlehood. The event immediately following type $i$'s singlehood duration is the realization of with whom to match. This event is given by the density of accepted types, $f(x_j|x_i) = f(x_j|j \in A_i)$. Let $N_j$ be the number of type $j$ agents, and $\sum_{j \in A_i} N_j$ be the number of potential partner acceptable to a type $i$ man. The acceptance criterion of a type $i$ man is endogenously determined by the matching algorithm described above. Given a type $i$ man, the probability that a 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,

$$f(x_j|x_i) = \frac{N_j}{\sum_{k \in A_i} N_k}.$$

Marriage duration $T_1$ has an exponential distribution with parameter $\delta$. The total type $i$ individual likelihood contribution to events between entering marriage and possible separation is equal to the following:

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

where $T_1 > 0 $. The total type $i$ individual likelihood contribution who is single for duration $T_0$ and who tend proceeds to get married for a duration of $T_1$ is equal the product of the two likelihoods above:

$$L_{ij} = L_{0i} L_{1ij}.$$

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

$$L = \prod_{i=1}^{n} L_{ij}$$

Using optimizers from the SciPy library, we choose values of $\lambda$ and $\delta$ that minimize the negative natural log of $L$.

# Results

## Comparing Parameter Estimates

Table 2: Model Results

| | | Point Estimate | S.E. | Lower Bound (95%) | Upper Bound (95%) |
|---|---|---|---|---|---|
| Men Who Married While Young | $\lambda$ | 0.759 | 0.106 | 0.551 | 0.966 |
| | $\delta$ | 0.053 | 0.011 | 0.031 | 0.075 |
| Men Who Married While Older | $\lambda$ | 0.254 | 0.051 | 0.154 | 0.353 |
| | $\delta$ | 0.050 | 0.009 | 0.032 | 0.068 |

We begin by comparing the parameter estimates, which are shown in Table 2. We can see that the $\lambda$ estimates for the *married young* group are higher than those for the older group (0.76 > 0.25). The $\delta$ estimates on the other hand are more or less identical (approx. 0.05 for both groups). But let's put a more substantive interpretation onto these results. The approximately identical sets of $\delta$ values tell us that the marriage dissolution rates for the two subpopulations are indistinguishable from one another. The differences in $\lambda$, meanwhile, tell us the prospective partners arrive at a significantly higher rate for those in the married-young group. This corroborates our first hunch from earlier, that younger folks experience lower search friction in their pursuit of the one.

## Comparing Optimal Match Sets

Table 3: Optimal Match Sets for Men

| | Attractiveness-type | Max | Reserve |
|---|---|---|---|
| Married While Young | 5 | 5.0 | 5.0 |
| | 4 | 4.0 | 3.0 |
| | 3 | 4.0 | 3.0 |
| | 2 | 2.0 | 2.0 |
| | 1 | 2.0 | 2.0 |
| Married While Older | 5 | 5.0 | 4.0 |
| | 4 | 5.0 | 4.0 |
| | 3 | 3.0 | 2.0 |
| | 2 | 3.0 | 2.0 |
| | 1 | 1.0 | 1.0 |

Table 4: Optimal Match Sets for Women

| | Attractiveness-type | Max | Reserve |
|---|---|---|---|
| Married While Young | 5 | 5.0 | 5.0 |
| | 4 | 4.0 | 3.0 |
| | 3 | 4.0 | 3.0 |
| | 2 | 2.0 | 1.0 |
| | 1 | 1.0 | 1.0 |
| Married While Older | 5 | 5.0 | 4.0 |
| | 4 | 5.0 | 4.0 |
| | 3 | 3.0 | 2.0 |
| | 2 | 3.0 | 2.0 |
| | 1 | 1.0 | 1.0 |

Next, we move on to a comparison of optimal match sets. Table 3 shows the optimal match sets for men; Table 4 shows the same for women. We can observe the greater positive assortative matching in the . Of course, this is necessarily true given the optimal reservation-match policy we're assuming: higher $\lambda$ values, ceteris paribus, leads to greater assortative mating. In the cases of both men and women, we can see that the those of the higher attractiveness-type, type 5, are only willing to match with each other. In the case of men, we can see that men of the lowest attractiveness type, type 1, aren't even willing to "settle" for their women counterpart, the type 1 women.

## Conclusion

There isn't too much to say here. Using a structural approach that embeds the numerical solution of a two-sided matching model within a maximum likelihood procedure, I estimate marriage models for two groups of men, (a) those who married before the age of 22, and (b) those who married after the age of 22. I use the model estimates to adjudicate between two rivalrous set of hunches. I find that those who married young experienced lower search friction during their pursuit of partners. While both groups face similar rates of marriage dissolution, those who married young experienced higher rates of partner arrival. As a result, they tend to be more selective about their choices of partner, and may actually demonstrate higher levels of positive assortative mating.

I would like to spend a moment here to ponder the limitations of this particular replication. This began with an intention to replicate Wong's study in earnest, but the task quickly turned more formidable than I'd initially expected. PSID data require(s/d) far more wrangling than initially expected. Beyond the occlusion of certain important sociodemographic attributes (namely race and education), I also omitted some data-imputation procedures that Wong carried out, and omitted treatment of data censoring. To correct for over-abundance of zeros in wage-hours among women, Wong does a Heckman two-step procedure. Not being entirely familiar with this, I did not do so. These departures definitely weakened and impoverished the replication – for one, it required an artificial calibration of the alpha parameter that scales the relationship between wage and education. Second, I did not perform the correction of "classification errors" that Wong performed. All of this means that the results here can not be expected to be commensurate with Wong's.

## References

Wong, Linda Y. 2003. "Structural Estimation of Marriage Models." *Journal of Labor Economics* 21(3):699–727.