Understanding Mate Preferences from Two-Sided Matching Markets: Identification, Estimation and Policy Analysis

Yu-Wei Hsieh[*][†]
Department of Economics
New York University

First Draft Nov. 13th, 2011

[*]Author for correspondence. Tel: +1 917-288-0113; Address: NYU, 19 W. 4th Street, 6FL, New York, NY 10012; Email: yuwei.hsieh@nyu.edu

[†]I am indebted to Christphor Flinn and Bryan Graham for stimulating, encouraging and guiding me through this research project. I am also grateful to Baiyu Dong, Michael Jasson, Konrad Menzel, Dennis Nekipelov, Demian Pouzo, Jim Powell, Bernardo Silveira, Daniel Xu, Shengxing Zhang for several useful discussions. I would like to thank the seminar participants in Trans-Atlantic Doctoral Conference at London Business School, UC-Berkeley and NYU. Their suggestions lead to a better exposition for this paper. I am very thankful to Daniel Xu for letting me use his workstation. All the usual disclaimers apply.

Abstract

In this paper we estimate the utility functions over partner's characteristics from the aggregate matching patterns, using a structural two-sided matching model without transfer. A distinct feature of our approach is the ability to separately identify utility functions of men and women, which is absent in the matching models with transfer; e.g., Choo and Siow (2006). We argue that the deferred-acceptance algorithm can be represented as a special demand/supply system. It thus leads to a fast algorithm to compute the joint distribution of characteristics of married couples implied by the model without simulation. Based on it, we present a set of new nonparametric identification results and propose consistent estimators, which are free from the curse of dimensionality induced by large number of players. Testable implications of Men (Women) optimal stable matching are also investigated. Furthermore, we consider the inference problem without imposing equilibrium selection. We show that a set of moment inequalities can be derived from the no-blocking-pair condition without solving the game. Besides marriage market, our estimators can also apply to other two-sided matching markets in labor and IO.

We use the data of Current Population Survey to estimate the utility functions over spouse's education level. We find that men care more about the similarity in spouse's education than women do. Meanwhile, women's education level is becoming more attractive to men compared to 20 years ago. It is well known that marriage market is one of the key considerations of education choice, but such endogeneity problem in empirical two-sided matching models is usually assumed away. Therefore, we develop a structural pre-marital educational choice model that accounts for marriage market prospect. We use the estimated preference to compute the "return to schooling within marriage", and estimate the cost distributions of education via the structural model. By shifting the cost distribution of women via the lump-sum tuition subsidy, we find that the college attendance rate increases for both women and men. Moreover, the new policy creates more college-educated couples, while the correlation of the joint distribution of education of married couples remains unchanged.

Keywords: Two-Sided Matching; Identification; Discrete Choice Model; Marriage Market; Pre-Marital Investment; Assortative Matching; Social Interaction; Education Cost; Deferred-Acceptance Algorithm; Structural Estimation; Testable Implication.

JEL classification: J11; J12; C21; C25; C78; C63

1 Introduction

Many economic systems can be described as two-sided matching markets, with a preference list for each agent over potential partners. Following Roth and Sotomayor (1990), the term "two-sided" refers to the case that agents in such market can be classified into one of two disjoint sets. There are several such examples: men and women, schools and students, employers and employees, sport teams and athletes,...etc. Among other factors, preference heterogeneity and resource constraints play the most important role in determining the market outcome. Because of preference heterogeneity, there may exist conflicts of interest among agents (e.g., workers want to earn more but employers wan to pay less). Because of resource constraints, agents have to compete for limited resources (e.g., students want to go to the top schools but schools only have limited capacity). To satisfy both parties' needs subject to the resource constraints, game theorists have long been interested in designing a mechanism to match agents. The leading cases are the labor markets between hospitals and medical students (Roth, 1984).

On the other hand, economists and sociologists are also interested in explaining several aggregate matching patterns. For example, it is well-documented that people have a tendency to marry someone who shares similar characteristics, a phenomenon known as assortative matching[1] (Kalmijn, 1998). Statistically, assortative matching implies that the joint distribution of characteristics of married couples is positively correlated.[2] The economics of partnership can be dated back to the seminal work of Beck (1973): Supposes agent can only produce in pairs and if the characteristics of agents are complements in the matching production function, than the resulting matching will be assortative. Becker's marriage model thus provide a foundation linking the the underlying preferences of both parties to the observed outcomes: a foundation for structural estimation. However, not until Choo and Siow (2006) who propose the first empirical version of Becker (1973) did empirical researchers have tools to estimate the preferences from the observed matchings.

Structural estimation is important in two aspects. First, it helps us further understand the motivation for matching, and the role of resource constraints in matching. Second, the estimated preference can be further used to conduct counterfactual analysis of markets outcomes in different environments. It thus provides a valuable tool to assess the likely outcomes of a new policy and help the decision makers to design the new mechanism.

The difficulty of structural estimation comes from the fact that the market outcome is the result of mutual choices made by individuals in both sides of the market. As a result,

[1] In the sociology literature, such phenomenon is also known as intermarriage or homogamy.

[2] Anti-assortative matching implies negative correlation.

the observed matching does not necessarily reveal the underlying preferences, but more likely to be the outcome of compromise of the resource constraints and the conflicts of interest. Therefore, the market outcome cannot be analyzed within the traditional single agent discrete choice models. For example, if a woman marries an ugly man, standard single agent discrete choice models will conclude that her utility of marrying an ugly guy $u(\text{ugly}) + \epsilon(\text{ugly})$ is greater than the utility of marrying a handsome man $u(\text{handsom}) + \epsilon(\text{handsome})$, where $u(\cdot)$ is the deterministic components of utility function and $\epsilon(\cdot)$ is the random utility components. The likelihood can then be derived base on the above restriction. If there are many women who have ugly spouses, the MLE will conclude that on average women prefer ugly men: $u(\text{ugly}) > u(\text{handsom})$. However, such conclusion is not necessary true in the two-sided matching markets. First, it is possible that handsome men are in short supply, so statistically we will observe that many women who have ugly spouses. Such resource constraints are characterized by the marginal distributions of types, and should be taken into account in econometric analysis. Second, it is possible that handsome men do not like her, not because she does not prefer a handsome guy. Third, women have to compete with each other for the spouse they want, and it is not easy to have a handsome spouse. As a result, in econometric analysis one should consider the complicated social interactions among players on both sides of the market. However, the problem encountered here even goes beyond the scope of discrete choice models with social interactions (e.g., Brock and Durlauf, 2007; Ciliberto and Tamer, 2009). Not only the payoff of each agent depends on the other players' actions, but also the choice set is endogenous. For instance, a man can only choose his spouse from the set of women who wish to marry him. Therefore, a direct inference based on the observed choice is not possible as we do not observe the relevant choice set for each agent.

To account for the resource constraints and social interactions among players, I borrow the idea from the matching game without transfer (NTU matching game)[3] as the foundation for structural estimation. It possesses several advantages. From the modeling perspective, it overcomes the difficulty of modeling agents' strategies when the number of conceivable strategies is very large (Roth and Sotomayor, 1990). For example in the marriage market, a guy can choose, what to dress in a party or what education level to obtain, to attract his potential spouse. Modeling all of them is prohibitively difficult. NTU matching games offer a relative tractable way to model such circumstance than non-cooperative game framework. First the utility function over potential partners for each agent is specified. Second, the collection of utility functions of all agents is then

[3]In the matching game without transfer, agents are not allowed to exchange money and hence there does not exist a price mechanism.

mapped to the equilibrium outcome. Exploiting such mapping forms the foundation for structural estimation and is the main goal of this paper. Empirically, it offers the possibility to separately identify utility functions for both sides of the market; c.f., Choo and Siow (2006).

Recently, applying the NTU matching game in empirical studies has received considerable attention. These include household economics (Del Boca and Flinn, 2011), marriage markets (Dagsvik, 2000; Logan, Hoff and Newton, 2008; Hitsch, Hortacsu and Ariely, 2010a,b), teacher labor market (Boyd et al., 2006), and some applications in industry organization (Sorensen, 2007; Chen, 2008). Although several researchers already suggested some estimation procedures, little is known about the identifiability of the utility functions in NTU matching models. Identification analysis in NTU matching models is challenging since NTU matching models often predict multiple equilibria. Most papers sidestep the problem of multiplicity by choosing utility specifications that exclude preference heterogeneity (e.g., Sorensen, 2007), or based on some sort of equilibrium selections (e.g., Boyd et al., 2006). The exception is Menzel (2011) who explicitly accounts for multiplicity within the Bayesian framework of Logan, Hoff and Newton (2008). In terms of estimation, the existing simulation-based approaches (Boyd et al., 2006) are computationally cumbersome and inaccurate since the number of players in matching games are usually quite large.

In this paper, I approach the identification and estimation problems of NTU matching models from different perspectives. The utility specifications are chosen to be as flexible as possible to allow for multiple equilibria. Observed couples are treated as an equilibrium outcome. Utilizing the NTU matching theory, our structural model provides a link between individual preferences and the aggregate matching patterns. It is capable of answering how the individual preferences and the marginal distributions of types affect the matching patterns. Two important cases are considered: with and without equilibrium selection. In the first case, we assume the equilibrium outcome is Men or Women optimal stable assignment. We seek to back out the utility functions over partner's characteristics from the contingency table of marriage types (joint distribution of characteristics of married couples). The idea is closely related to Berry, Levinsohn and Pakes (1995), who estimate the consumers' demand function from the market share data. However, our method is distinct from BLP as their method is a single agent discrete choice model. Under the aggregate matching assumption, we argue that the deferred-acceptance algorithm can be represented as a special demand and supply system. It leads to a fast deferred-acceptance algorithm that takes the utility functions of men and women and the marginal distributions of types as inputs, and delivers the contingency table of marriage types implied by the Men-Optimal or Women-Optimal assignment. We then proceed the econometric analysis by matching the contingency table implied by the model with the one observed in the data. We find that the non-parametric utility functions can only be partially identified, even the Men or Women optimal equilibrium is assumed. We then discuss empirical strategies that lead to a smaller identified set. We also propose consistent estimators, which are free from the curse of dimensionality induced by large number of players, a notorious problem in the existing simulation-based approaches. Finally, we study the testable implications of the structural model assuming M-optimal assignment. In the second case, we assume the equilibrium outcome is stable without specifying the equilibrium selection. We derive a set of moment inequalities from the no-blocking-pair condition. Such approach avoids the need to solve the matching game and hence is computationally attractive. The proposed estimator is distinct from that of Echenique, Lee and Shum (2010), in the sense that we also utilize the frequency information in the contingency table. A lower bound for the probability of the observed matching is also derived. In the empirical implementation, the marriage market data extracted from the Current Population Survey are used to estimate the preferences over spouse's education level. We find that men care more about the similarity in spouse's education than women. Meanwhile, women's education level is becoming more attractive to men than 20 years ago. Besides marriage market, the proposed estimators can also apply to other two-sided matching markets in labor and IO.

This paper goes one step further than simply estimating the utility functions over spouse's characteristics. Recall that the resource constraints also has a significant impact on the market outcome. For example, Banks (2011) pointed out that declining marriage rate for African-Americans is mainly driven by the increasing college attendance rate of black women. Since the college attendance rate of black women is far more greater than black men, and black women prefer single rather than marrying someone incompatible, many black women end up with never get married. In light of this story, it is also important to understand how the marriage market will react to the changing of men's and women's characteristics. However, virtually all of the empirical work on matching assume that the marginal distributions of men's and women's characteristics are exogenously given prior to matching market entry. Moreover, the counterfactual analysis in the context of two-sided matching is also absent in the literature. In this paper we propose a structural pre-marital education choice model that can incorporates the marriage market consideration. We do not impose strong assumptions on the market outcome such as perfect assortative matching; e.g., Peters and Siow (2002). Instead, we use the estimated preference to compute the "return to schooling within marriage", and estimate the cost distributions of education via the structural model. We find that women's education cost drops substantially in the past four decades, while men's education cost remains relatively stable. A counterfactual analysis is then performed to evaluate the treatment effect of lump-sum tuition subsidy on the market outcome. We find that the tuition subsidy for women increases the college attendance rate for both women and man. The new policy creates more college-educated couples, while leaving the correlation of the joint distribution of education of married couples unchanged. The increased college attendance rate of women is intuitive. Regarding men, since it is more easy to marry high-educated women because of the increased supply, men are more willing to invest in their human capital. Otherwise men will become incompatible to high-educated women, which in turn creates disutility in marriage.

This paper is also related to the models with transfer (TU matching game) in which agents are allowed to exchange money. The econometric analysis of TU matching models is initiated by Choo and Siow (2006), and subsequently extended by Fox (2010), Galichon and Salanie (2010), Graham (2011), among others. TU matching games model the competition among players and the resource constraints by introducing the price mechanism. If handsome men are in short supply and women also prefer handsome men, the market price for marrying a handsome man will go up to clear the market. The novelty of TU matching models is that it capable of decomposing the complex problem into two single-agent discrete choice problems, subject to the market clearing conditions. Although the common market prices are usually unavailable, they can be treated as unobserved "fixed effect" and hence can be removed. Graham (2011) shows that many identifying strategies and estimators in panel data can be transported to the TU matching game. However, the limitation of Choo and Siow (2006)-type TU matching models is that it is unable to separately identify men's and women's utility functions.[4] Loosely speaking, only the sum of men's and women's utility functions is identifiable.[5] Moreover, the market prices within the model are not identified. These facts make the TU matching model unattractive in terms of policy analysis since some key features of structural parameters are not identified.

The remaining sections are organized as follow. In section 2, we briefly review some important concepts in the NTU matching game. In section 3, we introduce the notion of aggregate matching, and a modified deferred-acceptance algorithm for computing the contingency table of marriage types. We then proceed to discuss the identification and estimation problems. In section 4, we estimate the utility functions over spouse's characteristics using the CPS data. In section 5, a structural model of pre-martial education choice is proposed. We estimate the implied distributions of education cost and evaluate the effect of lump-sum tuition subsidy policy on the marriage market. In section 6, we con-

[4]Chiappori, Salanie and Weiss (2011) is an exception but they did not perform counterfactual analysis either.

[5]The sum of utility functions of both sides of the market is known as matching surplus or matching production function.

sider the inference problem without imposing equilibrium selection. Section 7 concludes. Technical proofs are collected in the Appendix.

2 Review of Two-Sided Matchings without Transfer

Using economic theory to relate the individual preference to the observed choice is essential to all microeconometrics analysis. For example in standard discrete choice models, the individual utility function is related to the observed choice by assuming utility maximization. We shall extend this modeling approach to the two-sided matching market. Specifically, we relate the collection of individual preferences to the observed aggregate matching patterns by assuming some solution concept.

We shall first briefly review some important solution concepts in the one-to-one NTU matching game, and then draw the connection between the model and the data. Readers who are familiar with matching game theory can skip this section. The Matching is one-to-one. Each man matches with one woman and vice versa. For simplicity, we assume there are equal amount of men and women. We will denote by $\mathcal{M} = \{m_1, m_2, \ldots, m_N\}$ and $\mathcal{W} = \{w^1, w^2, \ldots, w_N\}$ the set of men and women, respectively. Each player has a preference list Q of his/her marriage candidate, ranking from the most favorite to the least favorite. For example, man i's preference list can be written as

$$Q(m_i) : w^1, w^5, \ldots, w^j, \ldots.$$

The preference profile, which is defined as the collection of men's and women's preference lists $\mathcal{Q} = (Q(m_1), \ldots, Q(m_N); Q(w^1), \ldots, Q(w^N))$, is the main object of interest we want to learn from the data. The only available data, the matching matrix μ, is a matrix indicating who marry who. We use 1 to indicate the marriage relationship between two agents. For example, if m_1 and w^2, and m_2 and w^1 are couples then the matching matrix μ is given by

	w^1	w^2
m_1	0	1
m_2	1	0

In the one-to-one matching, the row sum and column sum are smaller than or equal to 1. Given a preference profile $\mathcal{Q}$, agents try their best to win the best possible spouse. Since there may exist some conflicts among agents, game theorists already made considerable effort to design some assignment mechanisms $\mathcal{T}$ to clear the market. If we can characterize some properties of the assignment mechanism $\mathcal{T}$, we might be able to back out the preference profile $\mathcal{Q}$ through the relationship linking the model and the data: $\mathcal{T}(\mathcal{Q}) = \mu$. Next we introduce the solution concept considering which matchings are more likely to occur, and which are not. The definitions are taken from Roth and Sotomayor (1990).

Definition 1. *(Individual Rationality) A matching μ is individually rational if each agent is acceptable to his or her ate. That is, a matching is individually rational if it is not blocked by any agent.*

Since no one is forced to get married, an agent can always be better off by remaining single if his/her mate is unacceptable at matching μ. Therefore, an individually irrational matching is unlikely to happen. To simplify our analysis, we will assume that for each player, every marriage candidate is acceptable. Since we will also assume equal amount of men and women, eventually everyone will get married. A stronger solution concept is the no-blocking-pair condition.

Definition 2. *(Blocking Pair) Suppose there is a man m and a women w who are not matched to one another at μ, but who prefer each other to their mates at μ. Namely, $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$. The man and woman (m, w) are said to block the matching μ.*

A matching that exists blocking pairs are unlikely to sustain in the long run because those blocking pairs can be better off by marrying each other.

Definition 3. *(Stable Matching) A matching μ is stable if it is not blocked by any individual or any pair of agents.*

In general there usually exist multiple stable matchings given $\mathcal{Q}$. Therefore, the econometric analysis is more difficult than that of TU matching models. There are several ways to find the set of stable matchings. Roth and Sotomayor (1990), and Burkard, Dell'Amico and Martello (2009) survey this literature. The most influential one is the deferred-acceptance algorithm proposed by Gale and Shapley (1962) that can find the M and W-optimal matchings. It has been applied in many market design problems; e.g., high school assignment (Abdulkadiroglu, Pathak and Roth, 2005) and medical residents assignment (Kamada and Kojima, 2010).

Definition 4. *(deferred-acceptance algorithm)*
1. In the first round, each man proposes to the first woman on his preference list of acceptable women.

2. Each woman tentatively accepts the proposal that gives her the highest utility, and reject all other proposals and any proposal made by an unacceptable man.

3. Any man whose was rejected in the previous round then proposes to his next best choice, and he cannot propose to the women who have rejected him in the future round.

4. Each woman rejects all but her most favorite proposal from the set consisting of the new proposals made in the current round and the proposal she kept in the previous round.

5. The algorithm stops until no man is rejected.

They use this algorithm to prove the existence of a stable matching, regardless of the number of players and the preference profiles

Theorem 1. *(Gale and Shapley) A stable matching exists for every one-to-one, two-sided matching game.*

Finally, we introduce the notion of men and women optimal stable matchings:

Definition 5. *A stable matching μ is M-optimal if every man likes it at least as well as any other stable matchings; that is, if for every other stable matching $\mu^{'}$, $\mu \geq_{m_i} \mu^{'} \forall i$. Analogously, a stable matching μ is W-optimal if every woman likes it at least as well as any other stable matching, that is, if for every other stable matching $\mu^{'}$, $\mu \geq_{w^j} \mu^{'} \forall j$.*

M and W-optimal stable matching have a special role in the matching theory. When the preferences are strict, the set of stable matchings is a distributive lattice and bounded by the M and W-optimal matchings. In empirical study, researchers are also interested in these two polar cases.

We motive the identification issue by a simple example. Consider the matching market with two men and two women (2-by-2 matching market). In such a market, there are 2^4 possible preference profile $\mathcal{Q}$, with two potential outcomes: $\{(m_1, w_1), (m_2, w_2)\}$ and $\{(m_1, w_2), (m_2, w_1)\}$. Now let's consider some preference profiles and the corresponding stable matchings.

- **Conflicting Preference profile**

$$\begin{aligned} Q(m_1): &\quad w^1, w^2 \\ Q(m_2): &\quad w^2, w^1 \\ Q(w^1): &\quad m_2, m_1 \\ Q(w^2): &\quad m_1, m_2 \end{aligned}$$

Under this preference profile, $\{(m_1, w_1), (m_2, w_2)\}$ is the M-optimal stable matching while $\{(m_1, w_2), (m_2, w_1)\}$ is the W-optimal stable matching. I call this system as conflicting preference profile because man (woman)i's most preferred woman (man) does not like him (her).

- **Monotone Preference Profile**

$$\begin{aligned} Q(m_1) &: \quad w^1, w^2 \\ Q(m_2) &: \quad w^1, w^2 \\ Q(w^1) &: \quad m_1, m_2 \\ Q(w^2) &: \quad m_1, m_2 \end{aligned}$$

Under this preference profile, $\{(m_1, w_1), (m_2, w_2)\}$ is the only one stable matching. I call this system as monotone preference profile because every man has the same preference over women, and so does every woman. Sorensen (2007) and Chen (2008) belong to this class.

- **Mutually-Most-Preferred System**

$$\begin{aligned} Q(m_1) &: \quad w^1, w^2 \\ Q(m_2) &: \quad w^2, w^1 \\ Q(w^1) &: \quad m_1, m_2 \\ Q(w^2) &: \quad m_2, m_1 \end{aligned}$$

Under this preference profile, $\{(m_1, w_1), (m_2, w_2)\}$ is the only one stable matching. I call this system as mutually-most-preferred profile because man i's most preferred woman also likes him most. There is no conflict between players.

Clearly, it is easy to see that many different preference profiles all lead to the same matching μ. Equivalently, it is difficult to tell which preference profile is true by observing the matching outcome. The problem comes from the fact that there are only two potential outcomes, while there are 16 preference profiles we want to separately identify. Namely, the underlying model is too complex relative to the available information. The problem is even worse when the number of players increases ($N!^{2N}$ preference profiles versus $N!$ outcomes). Similar problems also encountered in econometric analysis of Nash entry games; e.g., Bajari, Han, Hong and Ridder (2011). The presence of multiple equilibria further complicates the analysis. To tackle this problem, we propose a model called aggregate matchings in the following section. Each agent has a random utility over potential spouses, and can be described by a finite dimensional parameters. The number of players is assumed to be large, so that some distributional features of μ can be estimated. Instead of estimating preference profiles through $\mathcal{T}(\mathcal{Q}) = \mu$, we proceed by matching the distributional features of $\mathcal{Q}$ and μ.

3 Aggregate Matchings and the Deferred-Acceptance Algorithm

We consider the case when there are a large number of players and their characteristics only take discrete values. This data scenario is usually encountered in marriage markets; e.g., Choo and Siow (2006), and Logan, Hoff and Newton (2008). It deserves a different treatment in two aspects: First, although theoretically the simulation-based estimators in the existing literature can be applied to the matching market with many players, practically the computational problem render it difficult to implement. As the number of players increases, the preference profiles we need to investigate increase exponentially, and for each preference profile we have to apply some algorithm to find all stable matchings. Second, the sets of players in different marriage markets are different. In contrast to the Nash entry game literature (Berry and Tamer, 2006) in which we observed the same set of players across different markets, usually the outcome space in two-sided matching models cannot be defined based on players' identities. There are few exceptions, however, for example the panel study of matched employer-employee data.

To tackle the first problem, we assume agents as having preferences over observable characteristics, instead of over individuals.[6] Specifically, men's preferences are only driven by women's observed types and vice versa. Namely, if woman j and k are of the same type, then man i is indifferent between j and k. Regarding the second problem, we only consider the joint distribution of marriage types derived from the matching matrix μ in each market, rather than the matching μ itself. Namely, players' identities do not matter. Only the joint distribution of marriage types matters in our analysis. Following Echenique, Lee and Shum (2010), we will refer this data scenario and modeling approach as aggregate matchings. It essentially consists of two ingredients: one assumption of dimension reduction on preferences and one on the data. We shall proceed by first describing the data available to the researchers, and then describing the underlying structure of the model. We then impose an equilibrium concept that links the underlying structure to the data, and discuss its identifiability.

[6]To avoid the curse of dimensionality, it is a common practice in the empirical IO literature to model consumers' preferences over product characteristics, rather than the individual products.

3.1 Aggregate Matchings

First we define the data available to the researchers.

Definition 6. *(Aggregate Matchings: Data)*
An M-by-M aggregate matching $\mathbf{C}$ *is the following M-by-M contingency table of marriage types in terms of frequency:*

type of men/women	*1*	...	*M*	*row sum*
1	C_{11}	...	C_{1M}	f_1^m
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
M	C_{M1}	...	C_{MM}	f_M^m
column sum	f_1^w	...	f_M^w	

where C_{ij} *is the* (i, j) *entry of the matrix* $\mathbf{C}$*, and it represents the proportion of type* (i, j) *matches. The sum of the* j*-th column is the proportion of the* j*-type women, denoted by* f_j^w*. Analogously, The sum of the* i*-th row is the proportion of the* i*-type men, denoted by* f_i^m*. We will refer to* $\boldsymbol{f^m} = (f_1^m, \ldots, f_M^m)$ *as the marginal distribution of men's characteristics, and* $\boldsymbol{f^w} = (f_1^w, \ldots, f_M^w)$ *as the marginal distribution of women's characteristics. To rule out the degenerate case, we further assume each element of* $\boldsymbol{f^m}$ *and* $\boldsymbol{f^w}$ *is non-zero.*

To simplify the analysis, we only consider the all-marriage case and hence the definition 6 does not include the case when agents can remain single. Also notice that the marginal distributions $\boldsymbol{f^m}$ and $\boldsymbol{f^w}$ characterize the resource constraints. Next we specify the individual preference:

Assumption 1. *(Aggregate Matchings: Preference)*
We assume that women can be categorized into M types, $Z^j \in \{1, \ldots, M\}$ $\forall j$ *and men can be categorized into M types,* $X_i \in \{1, \ldots, M\}$ $\forall i$*, where* X_i *and* Z^j *are man i and women j's characteristic.*[7] *Their preferences only depend on potential partners' observed type. Mathematically, the utility of man i matches with woman j is specified as:*

$$U_i(j) = \delta_m(X_i, Z^j) + \epsilon_i(Z^j),$$

Analogously, the utility of woman j matches with man i is specified as:

$$V^j(i) = \delta_w(X_i, Z^j) + \eta^j(X_i).$$

where δ_m *and* δ_w *are the deterministic components of the utility functions, and* $\epsilon_i(Z^j)$ *and* $\eta^j(X_i)$ *are the random components.*

[7] In general men's type space are allowed to be different from women's.

Under this utility specification, different agents who belong to the same category are allowed to have different preference over potential partners due to the utility shocks. However, each agent equally values his/her potential partners who belong to the same category. This setting is also a crucial identifying assumption in TU matching games; see Galichon and Salanie (2010), Graham (2011), and Chiappori, Salanie and Weiss (2011) for a discussion. The assumption 1 essentially reduce the number of possible preference profiles to a great extent. Each man only draws M utility shocks instead of N shocks, and each woman only draws M utility shocks instead of N shocks. Since we do not specify the utility of remaining single, we implicitly assume that agents are mutually acceptable to each other. We further assume the unobservable preference shocks $\boldsymbol{\epsilon}_i = (\epsilon_i(1), \ldots, \epsilon_i(M))$ and $\boldsymbol{\eta}^j = (\eta^j(1), \ldots, \eta^j(M))$ satisfy

Assumption 2. *$\boldsymbol{\epsilon}_i$ ($\boldsymbol{\eta}^j$) is a M-dimensional random vector, with each component $\epsilon_i(j); j = 1, \ldots, M$ ($\eta^j(i); i = 1, \ldots, M$) being i.i.d. drawn from F_ϵ (F_η). F_ϵ and F_η are assumed to be absolutely continuous w.r.t the Lebesgue measure, with unbounded supports on $\mathbb{R}$. Moreover, we assume the unobservable preference shocks between any two agents are independent; i.e., $\boldsymbol{\epsilon}_i \perp \boldsymbol{\epsilon}_j$, $\boldsymbol{\eta}^i \perp \boldsymbol{\eta}^j$, and $\boldsymbol{\epsilon}_i \perp \boldsymbol{\eta}^j$for all $i \neq j$.*

Given the utility specifications for (δ_m, δ_w) and the assumptions on (F_ϵ, F_η), we can compute the probability for a particular preference profile conditional on the agent's type:

Definition 7. *(Conditional Probability of Preference Profile)*
Let $(l_1, l_2, \ldots, l_M)$ be a permutation of $1, 2, \ldots, M$. The conditional choice probability for an k-type man who's preference list is $(l_1, l_2, \ldots, l_M)$ is denoted by $P^m_{l_1 l_2 \ldots l_M|k} = Pr(\delta_m(X_i, Z^j = l_1) + e_i(l_1) > \delta_m(X_i, Z^j = l_2) + e_i(l_2) > \cdots > \delta_m(X_i, Z^j = l_M) + e_i(l_M)|X_i = k)$. Similarly we define $P^w_{l_1 l_2 \ldots l_M|k}$ as a k-type woman's conditional choice probability for the preference list $(l_1, l_2, \ldots, l_M)$. The set of all possible conditional probabilities of preference profiles for men (women) will be denoted by $\boldsymbol{P}^m$ ($\boldsymbol{P}^w$).

As will become clear later, the conditional probabilities of preference profiles characterize the proportion of agents who adopt a particular proposing/accepting strategy in the deferred-acceptance algorithm. Finally, we define the structure for the aggregate matching model:

Definition 8. *(Structure)*
We will denote by S the collection of $(\mathbf{X}, \mathbf{Z}, \delta_m, \delta_w, F_\epsilon, F_\eta)$, where $\mathbf{X} = (X_1, \ldots, X_N)$ and $\mathbf{Z} = (Z_1, \ldots, Z^N)$. X_i and Z^j are i.i.d. distributed according $\boldsymbol{f^m}$ and $\boldsymbol{f^w}$ respectively. (δ_m, δ_w) are assumed to satisfy assumption 1 and (F_ϵ, F_η) are assumed to satisfy assumption 2. Notice that $(\mathbf{X}, \mathbf{Z})$ are observable while the random utility shocks are the unobserv-

ables, and hence are the stochastic components of the matching markets. A structure s is an element of S.

We need a solution concept that links the structure s to the observed outcomes.[8] A natural candidate is stable matching because if the observed matching is not stable, it is not sustainable in the long run. A stable matching in aggregate matchings can be viewed as a mapping $\mathcal{T}$ from s to the contingency table of marriage types $\mathbf{C}(s;\mathcal{T})$. Notice that although different stable assignments may match different agents, it is possible that the derived contingency tables of marriage types $\mathbf{C}(s;\mathcal{T})$ are the same. In this section we will focus on the case of M-optimal and W-optimal assignment. Econometrics analysis without assuming a specific equilibrium selection will be discussed later. We will simply write $\mathbf{C}(s)$ if there is no ambiguity about the underlying assignment mechanism.

An estimation method tailored for M or W-optimal assignment[9] is important in three aspects. First, the deferred-acceptance algorithm and its ramifications has been widely applied in many market design problems. The leading cases are the labor markets of medical students (Roth, 1984), and the high school assignment problems (Abdulkadiroglu, Pathak and Roth, 2005). Thus understanding the inference problem assuming M-optimal equilibrium is interesting on its own right. Second, even though the actual mechanism is unknown, assuming M (W) optimal assignment may still provide some useful results as the set of stable matchings is bounded by M and W optimal equilibrium. It is as if we can analyze two extreme cases (e.g., Boyd et al., 2006; Del Boca and Flinn, 2011). Third, using the M or W-optimal equilibria as the identifying assumption has an advantage over other assumptions. It models the problem of endogenous choice set and agents' strategies in a way that is closely related to the classical discrete choice models. Some insights from discrete choice models can then be directly applied to the two-sided matching models.

3.2 Simulation-Based Approaches

It is well-known that the existing simulation-based estimators (Boyd et al., 2006) are constrained by the number of players, because the contingency table $\mathbf{C}$ is obtained indirectly: First, at the k-th simulation we simulate a preference list for each player according to $s = (\mathbf{X}, \mathbf{Z}, \delta_m, \delta_w, F_\epsilon, F_\eta)$. Second, the matching matrix μ_k is obtained by the deferred-acceptance algorithm. Third, the corresponding contingency table $\mathbf{C}_k$ is derived from μ_k

[8]For example, utility maximization is the solution concept linking the utility functions to the observed choices.

[9]It is important to know that assuming M-optimal is different from assuming the underlying mechanism is implemented by the deferred-acceptance algorithm. It is merely an algorithm to find the M-optimal matching.

by aggregation. The above procedure is then repeated K times and the simulated contingency table of marriage types implied by s is given by $\tilde{\mathbf{C}}(s) = \frac{1}{K}\sum_{k=1}^{K} \mathbf{C}_k$. Finally, the true structure $s_o \in \mathcal{S}$ can be solved by matching the features of $\tilde{\mathbf{C}}(s)$ and the features of the observed contingency table of marriage types $\hat{\mathbf{C}}$. Ideally we would like to compute the exact $\mathbf{C}(s)$ implied by the model, instead of the simulated version $\tilde{\mathbf{C}}(s)$. However, in a market with 10 couples, there are $(10!)^{20} \doteq 1.56 \times 10^{131}$ possible preference profiles need to be evaluated in order to obtain $\mathbf{C}(s)$. Even the state of the art algorithm is quite efficient, executing it 10^{131} times is prohibitively difficult. Consequently, in practice one can only choose K to be a relatively small number. However, the potential pitfall of such practice is that $\tilde{\mathbf{C}}(s)$ might substantially differ from $\mathbf{C}(s)$ since only a tiny portion of the preference profiles is evaluated.

3.3 A Modified Deferred-Acceptance Algorithm

To facilitate any inference procedure, it is necessary to have a fast algorithm to compute the contingency table of marriage types $\mathbf{C}$(s). Under the aggregate matching settings, we propose a modified deferred-acceptance algorithm to compute $\mathbf{C}(s)$ directly without first computing the matching matrix μ. The inputs of our algorithm consist of two part: the marginal distributions of types $(\boldsymbol{f^m}, \boldsymbol{f^w})$ and the parameter vector β characterizing the preferences. Specifically, our algorithm only requires the distributional information and hence is free from the curse of dimensionality caused by the large number of players. We exploit the fact that the deferred-acceptance algorithm is essentially a mechanism to sequentially equate the demand-supply system without the price; c.f., Choo and Siow (2006). The distributions of demand and supply are created according to the marginal distributions of types $(\boldsymbol{f^m}, \boldsymbol{f^w})$ and the set of conditional probabilities of preference profiles $(\boldsymbol{P}^m, \boldsymbol{P}^w)$. A striking result is that the structural parameters may not be point identified, even the equilibrium selection is specified.

Under the assumption of aggregate matchings, the preference only depend on the observed covariate. As a result, the preference is not strict. The deferred-acceptance algorithm continue to produce a stable matching, so long as the algorithm includes a rule to break ties. As discussed in Roth and Sotomayor (1990), the tie-breaking procedure can be viewed as any way of converting the original preference profile $\mathcal{Q}$ into a strict preference profile $\mathcal{Q}'$ that differ from $\mathcal{Q}$ only in those comparisons in which indifference was expressed. We impose two tie-breaking rules to simplify the implementation of the original deferred-acceptance algorithm. These two rules are deducted from the property of assumption 1.

Rule 1: *We assume that each man draws a random strict preference list according to* *some distribution over the set of women who belong to the same category, and so does each woman.*

Under rule 1, we can preserve the original preference ordering in aggregate matching game but break ties among the set of observationally equivalent agents. Next we show that agents' strategies in the deferred-acceptance algorithm can be simplified due to assumption 1:

Rule 2: *A woman will not reject the offer she got in the previous round, unless another man who is strictly preferable proposes to her. Men only propose to those women who are not engaged.*

The idea of rule 2 is the following. Suppose man m prefers H-type women. Since there are many H-type women available and he only cares about women's type, compete with other men for the same H-type woman will not increase his utility. Therefore, he should propose to any available H-type woman. On the other hand, suppose an H-type woman $w^{'}$ who is engaged with another H-type man $m^{'}$ in the previous round. Since rejecting $m^{'}$ and accept m neither increases the utility for $w^{'}$ nor changes the contingency table, we assumes that w' will reject m if m propose to her. Knowing this fact in advance, m will not propose to any H-type woman who is engaged. He shall propose to a H-type woman who is still available, or propose to a woman of different type according to his preference list.

Under Rule 1 and 2 the deferred-acceptance algorithm can be interpreted as a special demand/supply system. All we need to do is to keep track of the number of matches created in each round, and the residual demand and supply. We use a simple 2-by-2 case to illustrate the idea. A pseudo code of the modified deferred-acceptance algorithm for the general M-by-M case is provided in the appendix. Consider the following configuration for the utility functions and the marginal distributions of types. We assume $(\delta_m, \delta_w) = \mathbf{0}$, which implies that agents randomly determine their preference list. Consequently, the conditional choice probabilities for preference profiles are given by:

$$P^m_{HL|H} = P^m_{HL|L} = P^w_{HL|H} = P^w_{HL|L} = 0.5$$

These four probabilities suffice to characterize the set of all conditional probabilities of preference profiles because of the adding-up constraints; e.g., $P^m_{LH|H} = 1 - P^m_{HL|H}$. Next we set the marginal distributions to be $\boldsymbol{f}^m = \boldsymbol{f}^w = (0.7, 0.3)$.

type of men/women	H	L	row sum
H	C_{HH}	C_{HL}	0.7
L	C_{LH}	C_{LL}	0.3
column sum	0.7	0.3	

We shall demonstrate how the algorithm generates the contingency table: $\text{vec}(\mathbf{C}) = (C_{HH}, C_{LH}, C_{HL}, C_{LL})'$.

In the first round, each man proposes to his most preferred woman. As a result, there are $P^m_{HL|H} \cdot f^m_H = 0.35$ unit measure of H-type men who propose to H-type women, and $P^m_{LH|H} \cdot f^m_H = 0.35$ unit measure of H-type men who propose to L-type women. This is the distribution of supply of the H-type men. Analogously, we can compute the distribution of supply of the L-type men. Next we consider the demand system generated from women's distribution and preference. Clearly, there are $P^w_{HL|H} \cdot f^w_H = 0.35$ unit measure of H-type women who prefer a H-type man and there are $P^w_{LH|H} \cdot f^w_H = 0.35$ unit measure of H-type women who prefer a L-type man. The distribution of the demand of the L-type women can be computed following the same procedure. The demand-supply system in the first round can then be represented by the following bimatrix of the demand and supply, with the supply in the first argument and the demand in the second argument

1st round: Bimatrix of the demand and supply

type of men/women	H	L
H	(0.35,0.35)	(0.35,0.15)
L	(0.15,0.35)	(0.15,0.15)

Clearly there are 4 sub-markets but only part of them have equal number of demand and supply. The modified deferred-acceptance algorithm provides a way to clear such system. Consider the (H, H)-type matchings, there are 0.35 unit measure of H-type men who propose to the type H women, and there are 0.35 unit measure of H-type women who will potentially accept their proposal. Since demand equals to supply, the (H, H)-type sub-market clears in the first round. Similar argument applies to the (L, L)-type sub-market. However, there exist excess supply of H-type men in the (H, L) sub-market and excess demand of L-type men in the (L, H)-type sub-market. Living in a world without price, only $\min(0.35, 0.15) = 0.15$ unit measure of (H, L)-type matchings will be created in the first round. $0.35 - \min(0.35, 0.15) = 0.2$ unit measure of H-type men will be rejected since each woman can only accept one proposal. The same argument applies to the (L, H)-type sub-market. To sum up, the modified deferred-acceptance algorithm produce two outputs: one is the number of matchings created, and one is a bimatrix describing the residual demand and supply.

1st round: Number of matchings created $\mathbf{C_1}$

	H	L
H	0.35	0.15
L	0.15	0.15

1st round: Bimatrix of the residual demand and supply

	H	L
H	(0,0)	(0.2,0)
L	(0,0.2)	(0,0)

Since there is no L-type women available, under rule 2 those who have been rejected by the L-type women shall propose to their second best choice, the type H women. One the other hand, since there is no L-type men available, the remaining H-type women shall compete for the proposals submitted by H-type men. Consequently, the bimatrix of the demand and supply is

2nd round: Bimatrix of the demand and supply

	H	L
H	(0.2,0.2)	(0,0)
L	(0,0)	(0,0)

Following the same argument, the modified deferred-acceptance algorithm produces the number of matchings created and the bimatrix of the residual demand and supply

2nd round: Number of matchings created $\mathbf{C_2}$

	H	L
H	0.2	0
L	0	0

2nd round: Bimatrix of the residual demand and supply

	H	L
H	(0,0)	(0,0)
L	(0,0)	(0,0)

The algorithm stops in the second round since there is no residual demand and supply. The desired contingency table is obtained by summing over the matrix of the number of matchings created in each round: $\mathbf{C}(s) = \mathbf{C_1} + \mathbf{C_2}$:

type of men/women	H	L	row sum
H	0.55	0.15	0.7
L	0.15	0.15	0.3
column sum	0.7	0.3	

Analogous to the original deferred-acceptance, the modified deferred-acceptance algorithm also converges in finite steps:

Theorem 2. *The modified deferred-acceptance algorithm converges in finite steps.*

Proof: see appendix.

At first glance one might think the contingency table of marriage types should be the one generated by the independent copula if agents' preferences have nothing to do with their types:

type of men/women	H	L	row sum
H	0.49	0.21	0.7
L	0.21	0.09	0.3
column sum	0.7	0.3	

However, it is interesting to note that in $\mathbf{C}(s)$ there exist positive assortative matching between men's and women's characteristics, even though their preference lists are completely random. The lesson from this example is that the relationship between assortative matching and complementarity are weak in the NTU matching games. This phenomena was also discussed in Becker and Murphy (2000). Graham (2011) further extends their theory to an important identification result in the TU matching games. There is a clear relationship between assortative matching and complementarity in the TU matching games. He shows that the sign of the local complementarity ϕ of the matching production function is determined by $sgn(\phi) = sgn(C_{HH} - f_H^m f_H^w)$ (simply compare the contingency table with the one generated by independent copula). In this case, $sgn(\phi) = (0.55 - 0.49) > 0$, implying individual types are complement. But in fact it is zero by construction. Therefore, estimators based on the TU matching assumptions may yield misleading results in some cases. On the other hand, estimators based on the NTU matching assumptions are robust to the potential pitfall emphasized in Becker and Murphy (2000).

3.4 Identification and Testable Implications

In this section we show the nonparametric identification results for the deterministic components of utility functions (δ_m, δ_w), through the relationship linking the model and the data: $\mathbf{C}(\hat{\boldsymbol{f}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) = \hat{\mathbf{C}}$. The marginal distributions $(\hat{\boldsymbol{f}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}^{\boldsymbol{w}})$ will be treated as known in the identification analysis because they are observable to the econometricians. The deterministic components of utility functions (δ_m, δ_w) will be treated as nonparametric functions and are the parameters of interest. Equivalently, we can instead treat the conditional probabilities of preference profiles $(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})$ as the parameters of interest. Given distributional specifications for the utility shocks (F_η, F_ϵ) satisfying assumptions 2, (δ_m, δ_w) can be obtained by inverting $(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})$. This is a useful technique in the dynamic discrete choice models; see Hotz and Miller (1993). Studying identification is difficult as the mapping from models' primitives to the data is only implicitly defined by the modified deferred acceptance algorithm. We provide some useful characterizations for the mapping $\mathbf{C}(\boldsymbol{f}^{\boldsymbol{m}}, \boldsymbol{f}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})$.

Theorem 3. *The contingency table* $\mathbf{C}(\boldsymbol{f}^{\boldsymbol{m}}, \boldsymbol{f}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})$ *is continuous in* $(\boldsymbol{f}^{\boldsymbol{m}}, \boldsymbol{f}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})$

Proof: see appendix.

The intuition why $\mathbf{C}$ is continuous is straightforward. If we only shift the marginal distributions and the conditional probabilities of preference profiles a little bit, the demand/supply system and agents' strategies (the distributions of men's proposals and women's demand orders) will not change too much. Our next theorem is concerned with the identifiability and testable implication of the proposed structural NTU matching model:

Theorem 4. *Given a 2-by-2 contingency table* $\hat{\mathbf{C}}$ *with marginal distributions* $(\hat{\boldsymbol{f}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}^{\boldsymbol{w}})$*, it is always possible to find* $(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})$ *to rationalized it. Namely, there always exist* $(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})$ *such that* $\mathbf{C}(\hat{\boldsymbol{f}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) = \hat{\mathbf{C}}$

Proof: see appendix.

Theorem 4 states that it is always possible to rationalize a 2-by-2[10] contingency table as the outcome of M- or W-optimal assignment. It guarantees that the identified set

$$\Theta \equiv \{(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) | \mathbf{C}(\hat{\boldsymbol{f}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) = \hat{\mathbf{C}}\}$$

[10] We conjecture that the proof can be generalized to the contingency table with arbitrary dimension. It is left as a future research topic.

will be non-empty. This result has an important implication in terms of hypothesis test and model selection. Since the structural estimation critically hinges on the modeling choices, there is a need to test such choices when possible. Theorem 4 implies that the structural NTU matching model we propose does not place a refutable restriction on the observed aggregate matching patterns. Consequently, it is impossible to test whether the true assignment mechanism is M-optimal or not:

Corollary 1. *In the 2-by-2 aggregate matchings it is impossible to test the following hypothesis under assumption 1 and 2 if (δ_m, δ_w) are nonparametric*

$$\begin{aligned} &\mathrm{H}_0 : \mathcal{T} \quad \textit{is } M\textit{-optimal stable assignment} \\ &\mathrm{H}_1 : \mathcal{T} \quad \textit{is not } M\textit{-optimal stable assignment} \end{aligned}$$

In particular, it is impossible to test M-optimal stable assignment against W-optimal stable assignment (or vise versa) under assumption 1 and 2 if (δ_m, δ_w) are nonparametric

$$\begin{aligned} &\mathrm{H}_0 : \mathcal{T} \quad \textit{is } M\textit{-optimal stable assignment} \\ &\mathrm{H}_1 : \mathcal{T} \quad \textit{is } W\textit{-optimal stable assignment} \end{aligned}$$

Similar results are obtained in the TU matching models. In Choo and Siow (2006), the matching production function can be written as a function (with a close form) of the contingency table. As a result, it is also impossible to statistically choose between the TU and NTU matching models. To generate falsifiability of the model, multiple observations of aggregate matchings $\hat{\mathbf{C}}_t$, $t = 1, 2, ...T$ are needed. In Choo and Siow (2006), observing multiple aggregate matchings immediately generates overidentifying restrictions because their model is exactly identified, and hence the TU matching model can be tested. See also Chiappori, Salanie and Weiss (2011). On the hand, Echenique (2008) shows that the NTU matching model is testable if multiple matchings μ_t are observed. However, he adopts a quite different framework compared to this paper. In particular, he assumes the same set of players are involved across different markets and there is no utility shock. As a result, the observed difference between each matching μ_t is solely driven by the equilibrium selection. In my structural NTU matching model, the model is testable if and only if there exist a sequence of $\hat{\mathbf{C}}_t$, $t = 1, 2, ..., T$ such that

$$\Theta \equiv \textstyle\bigcap_{t=1}^{T} \Theta_t = \emptyset, \;\text{ where }\; \Theta_t = \{(\boldsymbol{P^m}, \boldsymbol{P^w}) | \mathbf{C}(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P^m}, \boldsymbol{P^w}) = \hat{\mathbf{C}}_t\}.$$

It turns out that the structural NTU matching model assuming M-optimal assignment is also testable:

Theorem 5. *There exist a sequence of* $\hat{\mathbf{C}}_t$*,* $t = 1, 2, ..., T$ *such that* $\Theta = \emptyset$*.*

Proof: Choose $\hat{\mathbf{C}}_1 \neq \hat{\mathbf{C}}_2$ but $(\hat{\boldsymbol{f}}_{\mathbf{1}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\mathbf{1}}^{\boldsymbol{w}})$=$(\hat{\boldsymbol{f}}_{\mathbf{2}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\mathbf{2}}^{\boldsymbol{w}})$. Since the mapping of deferred-acceptance algorithm $\mathbf{C}(\cdot)$ is not set-valued, there does not exist a $(\boldsymbol{P^m}, \boldsymbol{P^w})$ to rationalize $(\hat{\mathbf{C}}_1, \hat{\mathbf{C}}_2)$. *Q.E.D.*

Notice that we prove theorem 5 by constructing a sequence of $\hat{\mathbf{C}}_t$ which can only be rationalized by a model with multiple equilibria. However, the set of $\hat{\mathbf{C}}_t$ that can reject our model is larger than this simple case. If the correlation structure changes dramatically in $\{\hat{\mathbf{C}}_t\}_{t=1}^T$, then the model will be rejected. For example, if men's type and women's type are comonotonic in $\hat{\mathbf{C}}_1$ while they are countercomonotonic in $\hat{\mathbf{C}}_2$, then there does not exist $(\boldsymbol{P^m}, \boldsymbol{P^w})$ to rationalize $(\hat{\mathbf{C}}_1, \hat{\mathbf{C}}_2)$.

Next we discuss the identification problem of aggregate matchings. We say $(\boldsymbol{P^m}, \boldsymbol{P^w})$ is point identified if there exists an unique $(\boldsymbol{P^m}, \boldsymbol{P^w})$ that uniquely solves $\mathbf{C}(\hat{\boldsymbol{f}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}^{\boldsymbol{w}}, \boldsymbol{P^m}, \boldsymbol{P^w}) = \hat{\mathbf{C}}$. If there exists multiple solutions, then (δ_M, δ_W) is partially identified. In general, the aggregate matching models are not nonparametrically point identified because the number of unknowns is greater than the number of available restrictions. For a given M-by-M contingency table, there are only $(M-1)^2$ non-redundant restrictions. However there are $(M!-1)^{2M}$ conditional probabilities of preference profiles to be identified, a number which is far greater than the number of restrictions. We conduct several case studies to illustrate the properties of identified set.

Case 1: 2-by-2 Aggregate Matchings

Recall that in the 2-by-2 case, the contingency table can be represented by any cell of the table once we conditional on the marginal distributions. Without loss of generality, we use the probability of observing a (H, H)-type matching to represent the contingency table. The proportion of (H, H)-type matchings indexes the degree of assortative matching. The high the number, the stronger the assortative matching. It also provides information for identifying $(\boldsymbol{P^m}, \boldsymbol{P^w})$. For example, if $P^m_{HL|H}$ and $P^w_{HL|H}$ are high while $P^m_{HL|L}$ and $P^w_{HL|L}$ are low, the probabilities in the main diagonal will be high.[11] Thus the complementarity in agents' types are related to the assortative matchings in their types, though such relationship is weaker than that of TU matching models. The reason being that there are four unknown conditional probabilities of preference profiles $(P^m_{HL|H}, P^m_{HL|L}, P^w_{HL|H}, P^w_{HL|L})$, and there is only one free parameter in the contingency table. As a result, those conditional choice probabilities are not point identified in general. In order to visualize the identified set, we restrict the dimension of the parameter

[11]See also the proof of theorem 4 for more details.

space and then plot the mapping from the $(\boldsymbol{P^m}, \boldsymbol{P^w})$ to the contingency table. Suppose $\delta_m = \delta_w = \delta$, and $\delta(H,H) = \delta(L,L)$ and $\delta(H,L) = \delta(L,H)$. These assumptions imply that $P^m_{HL|H} = P^w_{HL|H} \equiv x$ and $P^m_{HL|L} = P^w_{HL|L} \equiv y$. Even though the heterogeneity is restricted, such setup can still answer some interesting economic questions such as assortative matching; also see Graham, Imbens and Ridder (2010). For example, if x is high and y is low, then it implies that agents prefer assortative matching. If x is low and y is high, then it implies that agents prefer anti-assortative matching. We consider two configurations of marginal distributions: $(f^m_H, f^w_H) = (0.3, 0.7)$ and $(0.7, 0.7)$.[12] Figure 1 and 3 depict the mapping from $(\boldsymbol{P^m}, \boldsymbol{P^w})$ to the contingency table $z = \mathbf{C}(f^m_H, f^w_H, x, y)$, given different (f^m_H, f^w_H). Figure 2 and 4 are the corresponding contour plots. Each contour line characterizes the identified set of (x, y), given the probability of observing a (H,H)-type matching is z. The contour line possess a very specific shape. In fact, it is similar to the contour plot of Leontief production function. This is because the contingency table is obtained by sequentially taking the minimum of demand and supply. The size of the identified set varies in different cases. In Figure 2, if the probability of observing a (H,H)-type matching is 0.28, then the identified set is quite large. On the other hand, if the probability of observing a (H,H)-type matching is 0.1, then the identified set shrinks substantially. Another interesting feature is that the identified set is asymmetric. In Figure 2, if the probability of observing a (H,H)-type matching is 0.12, then then identified set essentially contains two components: $(x = [0, 0.4], y = 0.6)$ and $(x = 0.4, y = [0.6, 1])$. Even though one parameter is partially identified, another parameter is point identified.

Case 2: 2-by-1 Aggregate Matchings

Next we consider a special case in which the data provides no information for the conditional probabilities of preference profiles. If there is only one type for one-side of the market, then the contingency always equal to the marginal distribution of the other side of the market. The ideal is simple: if handsome men do not exist, then we will observed every women marry an ugly men, regardless of their preference. Mathematically, the number of restriction is zero in the 2-by-1 case.

Case 3: The Role of Variations in Marginal Distributions

So far our analysis assumes that there is only one contingency table available. The identified set can be further reduced if multiple distinct contingency tables $\hat{\mathbf{C}}_t$, $t = 1, 2, ...T$ (multiple independent matching markets) are available because $\Theta_t \subseteq \bigcap_{t=1}^{T} \Theta_t$. The idea

[12] We have tried several different configurations and they all yield qualitatively similar results. They are available upon request.

of variations in marginal distributions may help to shrink the identified set is the following. Consider two preference profiles: women prefer ugly men or women prefer handsome men. Men's preference over women's appearance is assumed to be random. If there is no handsome man, then according to case 2 these two preference profile are not distinguishable. However, if we increase the supply of handsome men, the proportion of women who marry handsome men under the second profile will be greater than the first one because the market demand for handsome men will increase substantially compared to the first profile. Variations in marginal distributions are essentially instrumental variables. Fixing the marginal distribution of women's characteristics, women's preference can be traced out by shifting the marginal distribution of men's characteristics. Fixing the marginal distribution of men's characteristics, men's preference can be traced out by shifting the marginal distribution of women's characteristics. As the modified deferred-acceptance algorithm can be interpreted as a demand/supply system, it is not surprising that the classical identification argument of simultaneous equation models can also be applied to the context of two-sided matchings.

We further conduct an experiment to demonstrate how large the variations of marginal distributions should be in order to distinguish two similar configurations of $(\boldsymbol{P^m}, \boldsymbol{P^w})$: $\theta \equiv (P^m_{HL|H}, P^m_{HL|L}, P^w_{HL|H}, P^w_{HL|L}) = (0.8, 0.8, 0.7, 0.6)$ and $\theta' = (0.7, 0.8, 0.7, 0.6)$. In figure 5 we depict the probability of observing a (H, H)-type matching as a function of marginal distributions (f^m_H, f^w_H), conditional on $(\boldsymbol{P^m}, \boldsymbol{P^w}) = \theta$ or θ'. In figure 6 we depict $\mathbf{C}(f^m_H, f^w_H, \theta) - \mathbf{C}(f^m_H, f^w_H, \theta')$. It is easy to see that θ and θ' are observationally distinguishable because there exist some configurations of (f^m_H, f^w_H) that lead to different contingency tables. However, there also exists a large area in which θ and θ' yield exactly the same contingency table. Therefore, in order to separately identified two sets of conditional probabilities of preference profiles that are close to each other, one may need sufficiently large variation in covariate distributions.

To sum up, we recommend some methods to reduce the identified set for practitioners. First, increase the number of types, which implies richer restrictions from the data. Since large type space also implies large number of unknown conditional probabilities of preference profiles, we can reduce its dimensionality by choosing some parsimonious specifications for the utility functions. Second, collect the data from independent matching markets, which also implies more restrictions from the data. We conclude this section by noting that assumption 1 and 2 are introduced for convenience only. Other utility specification and error structure are possible. For example, we can instead assume the utility shocks are match-specific, which is quite common in the bargaining literature: $U_i(j) = \delta_m(X_i, Z^j) + \lambda \cdot \epsilon_{ij}(X_i, Z^j)$ and $V^j(i) = \delta_w(X_i, Z^j) + (1 - \lambda) \cdot \epsilon_{ij}(X_i, Z^j)$, where

$\lambda \in [0,1]$. Our algorithm still works as long as the conditional probabilities of preference profiles can be computed. Different utility specifications merely imposes different structure on $(\boldsymbol{P}^m, \boldsymbol{P}^w)$, and hence results in different shape of the identification region.

3.5 Two-Step Estimators

Our identification argument naturally leads to a two-step estimator based on the relationship between the contingency table implied by the model and the observed contingency table: $\mathbf{C}(\hat{\boldsymbol{f}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) = \hat{\mathbf{C}}$. Regarding the sampling process, we assume that the researchers observe several independent aggregate matching markets

Assumption 3. *We have i.i.d. samples of T independent matching markets. In market t, we independently sample N_t married couples with characteristics (X_k^t, Z_k^t), where the superscript t represents the t-th market and the the subscript k represents the k-th couple. X and Z are men's and women's types.*

In the first step, we compute sample moments that are related to the contingency table of marriage types. By assumption 3, it immediately follows that the population contingency table $\mathbf{C}^t$ can be consistently estimated by the sample analog $\hat{\mathbf{C}}^t$, with the (i,j)-th element defined by

$$\hat{C}_{ij}^t = \frac{1}{N_t} \textstyle\sum_{k=1}^{N_t} \mathbf{I}_{[X_k^t = i, Z_k^t = j]} \tag{1}$$

Analogously, the population marginal distributions $(\boldsymbol{f_t^m}, \boldsymbol{f_t^w})$ can be consistently estimated by the sample analog $(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}})$, with the i-th element defined by

$$\begin{aligned} \hat{f}_{i,t}^m &= \frac{1}{N_t} \textstyle\sum_{k=1}^{N_t} \mathbf{I}_{[X_k^t = i]} \\ \hat{f}_{i,t}^w &= \frac{1}{N_t} \textstyle\sum_{k=1}^{N_t} \mathbf{I}_{[Z_k^t = i]} \end{aligned} \tag{2}$$

In the second step, the set estimator of conditional probabilities of preference profiles $\hat{\Theta}$ can be obtained by solving the following minimization problem

$$\begin{aligned} &\hat{\Theta} = \mathrm{argmin}_{(\boldsymbol{P^m}, \boldsymbol{P^w})} \hat{Q}_n(\boldsymbol{P^m}, \boldsymbol{P^w}), \\ &\hat{Q}_n(\boldsymbol{P^m}, \boldsymbol{P^w}) = \textstyle\sum_{t=1}^{T} ||\mathbf{C}(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P^m}, \boldsymbol{P^w}) - \hat{\mathbf{C}}^t||^2, \end{aligned} \tag{3}$$

where $||\cdot||$ is the Euclidean matrix norm defined on the non-redundant elements

$$||\mathbf{C}|| = \left(\textstyle\sum_{i=1}^{M-1} \sum_{j=1}^{M-1} |C_{ij}|^2 \right)^{1/2}.$$

If suitable parametric assumptions on (δ_m, δ_w) are imposed to achieved point identification, the preference parameters $\hat{\beta}$ can be obtained by solving

$$\begin{aligned} &\hat{\beta} = \text{argmin}_{\beta} \hat{Q}_n(\beta), \\ &\hat{Q}_n(\beta) = \textstyle\sum_{t=1}^{T} ||\mathbf{C}(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}(\beta), \boldsymbol{P}^{\boldsymbol{w}}(\beta)) - \hat{\mathbf{C}}^t||^2. \end{aligned} \tag{4}$$

The total sample size is $n = \sum_{t=1}^{T} N_t$. We study the fixed T, large n asymptotic properties. We also assume that $N_t/n \to r_t$, where $0 < r_t < 1$ for all t. The key property to guarantee the consistency of $\hat{\Theta}$ and $\hat{\beta}$ is the uniform convergence of the criterion function $(\mathbf{C}(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) - \hat{\mathbf{C}}^t)$; see Newey and McFadden (1994), and Chernozhukov, Hong and Tamer (2007).

Lemma 1. *Under assumption 3,* $\sup_{(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})} ||(\mathbf{C}(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) - \hat{\mathbf{C}}^t) - (\mathbf{C}(\boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{m}}, \boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) - \mathbf{C}^t)|| \to 0$ *in probability.*

Proof: see appendix.

We show that $\hat{\beta}$ is a consistent estimator:

Theorem 6. *(Consistency of $\hat{\beta}$)*
Suppose the utility function δ_m and δ_w are parametrize as continuous functions of $\beta \in \Theta$, where Θ is a compact subset of R^{k}*, and the distributions of random utility shocks satisfy the assumption 2. Under assumption 3, $\hat{\beta}$ is a consistent estimator for*
$\beta_0 = \text{argmin}_{\beta} Q_0(\beta) \equiv \sum_{t=1}^{T} ||\mathbf{C}(\boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{m}}, \boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}(\beta), \boldsymbol{P}^{\boldsymbol{w}}(\beta)) - \mathbf{C}^t||^2$.

Proof: see appendix.

To show the consistency of the set estimator $\hat{\Theta}$, we need the following lemma:

Lemma 2. *(Convergence Rate of* $\mathbf{C}(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) - \hat{\mathbf{C}}^t$*)*
Suppose $(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})$ *belongs to the population identified set Θ_I, then by assumption 3 we have* $\sup_{\Theta_I}(\mathbf{C}(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) - \hat{\mathbf{C}}^t) = O_p(n^{-1/2})$

Proof: see appendix.

Theorem 7. *(Consistency of $\hat{\Theta}$)*
Under assumption 3, $\hat{\Theta}$ is a consistent estimator for
$\Theta_I = \text{argmin}_{(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})} Q(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) \equiv \sum_{t=1}^{T} ||\mathbf{C}(\boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{m}}, \boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) - \mathbf{C}^t||^2$.

Proof: see appendix.

Deriving the limiting distribution for $\hat{\beta}$ is challenging as it is difficult to obtain the Bahadur representation; $\mathbf{C}(\cdot)$ is only implicitly defined by some algorithm. However, by lemma 2 we know that $\sup_{(\boldsymbol{P^m},\boldsymbol{P^w})\in\Theta_I} n\hat{Q}_n(\boldsymbol{P^m},\boldsymbol{P^w}) = O_p(1)$. Under additional mild regularity conditions, we can obtain the critical value by subsampling the empirical process of $n\hat{Q}_n(\boldsymbol{P^m},\boldsymbol{P^w})$. We refer the implementation details to Chernozhukov, Hong and Tamer (2007), and Romano and Shaikh (2010).

4 Estimation of Preference over Partner's Characteristics

4.1 Data

Defining the marriage market is of first order importance to the empirical study. However, it is difficult to have a fully satisfactory solution since the actual marriage market faced by each agent is unobserved to the econometricians. In the empirical implementation, we use the data of married couples extracted from IPUMS-CPS. One advantage of our estimator is that it only requires the knowledge of the joint distribution of marriage types. Therefore, even the actual marriage market cannot be observed exactly, we believe that the CPS data provides a good approximation to the distribution of agents' characteristics in the real world marriage market.

The data is constructed in the following way. First, we sample married couples in 1980, 1985, 1990, 1995, 2000, 2005, and 2010. Because of the sampling scheme the CPS adopted, duplicated observations will be included if we sample too frequently. Therefore, only March survey data are used. In order to have a thicker cells in the contingency table, we further aggregate the data of 1980&1985, 1990&1995 and 2000&2005. Therefore, what we sample is the stock of marriage at a given point in time.[13] Second, in each year we define the marriage markets according to the age category and region of residence. To be included in the final sample, their age should be older than or equal to 20, and younger than or equal to 60. We define 4 age categories based on husbands' age: 20-30, 31-40, 41-50, 51-60. We drop the observation whenever wife's age is more than 9 years younger or older than her husband. Regarding the region of residence, 51 states in US are used to define the location of marriage markets. In the CPS data, all married couples have the same code for the region of residence. Using these two criterions, we define 204 independent marriage markets for each year. Although somewhat arbitrary, we believe that age and

[13] An alternative sampling method is to sample the flow of marriage in a given period of time. For example, Echenique, Lee and Shum (2010) sample the new marriage data from the US Bureau of Vital Statistics. See also the discussion of Kalmijn (1998) for comparing these two sampling methods.

location are useful measures to define different marriage markets for the following two reasons. First, people tend to marry someone who belong to the same cohort. In real world we do observed people marry someone with great difference if age. However, since these samples only constitute a small portion, they are ignored in our analysis. Second, it is difficult to interact with someone living in different state on a day-to-day basis due to the geographical constraint.

We study preference over spouse's education level. Three education level are define: less than high school (type L), high school and partial college (type M), and 4-year college completion (type H).[14] It is well documented that the education level of men and women are increasing in the past four decades. The college attendance rate of women even exceeds that of men in recent years. The marginal distribution of education level for men and women are depicted in Figure 7. Clearly the change of marginal distributions of types will have some impacts on the marriage market, because the fundamental structure of demand and supply changes. For example, if college-educated men prefer college-educated women and the proportion of male college graduates increases, then the market demand for female college graduates will increase accordingly. Different marginal distributions also imply different resource constraints. Below we summarize the joint distributions of married couples' education level over time:

Contingency table of marriage types in 1980&85

men/ women	L	M	H	row sum
L	0.468	0.077	0.005	0.550
M	0.142	0.177	0.018	0.337
H	0.018	0.063	0.032	0.113
column sum	0.628	0.317	0.055	1
corr coef=0.55				
Kendall's $\tau = 0.52$				
Odds Ratio	10.34	4.13	14.84	

Contingency table of marriage types in 1990&95

[14] We use the variable *educ* in the CPS data. An agent will be classified as less than high school if $2 \leq educ \leq 72$, high school and partial college if $73 \leq educ \leq 110$, and 4-year college completion if $111 \leq educ \leq 125$.

men/ women	L	M	H	row sum
L	0.237	0.076	0.005	0.318
M	0.088	0.340	0.047	0.475
H	0.009	0.092	0.106	0.207
column sum	0.334	0.508	0.158	1
corr coef=0.63				
Kendall's $\tau = 0.602$				
Odds Ratio	17.65	5.35	14.96	

Contingency table of marriage types in 2000&2005

men/ women	L	M	H	row sum
L	0.058	0.046	0.004	0.108
M	0.034	0.440	0.094	0.568
H	0.003	0.107	0.213	0.323
column sum	0.095	0.593	0.311	1
corr coef=0.57				
Kendall's $\tau = 0.55$				
Odds Ratio	26.8	6.27	11.44	

Contingency table of marriage types in 2010&2010

men/ women	L	M	H	row sum
L	0.051	0.040	0.005	0.096
M	0.029	0.411	0.116	0.556
H	0.002	0.096	0.252	0.350
column sum	0.082	0.547	0.373	1
corr coef=0.62				
Kendall's $\tau = 0.56$				
Odds Ratio	31.92	6.42	11.25	

It is interesting to note that the probabilities in the main diagonal are far more greater than the probabilities in the off-diagonal cells: People tend to marry with someone who belong to the same education category. Moreover, the contingency tables change over the study period, partly because the marginal distributions change substantially. However, the degree of assortative matching, measured by the correlation coefficient and Kendall's τ, is quite stable over time. Another interesting fact is that the proportion of type (M,L) marriage drops substantially. In 1980s, it constitutes 14% of marriage but in 2000s it only constitutes 3%. On the other hand, the proportion of type (L,M) marriage are relatively stable over time. In 1980s, 7.6% are of type (L,M) marriage while in 2000s it drops to 4%.

An alternative measure of the degree of assortative matching is the odds ratio (Kalmijn, 1998). It is defined as the odds that an A-type man marries an A-type woman (rather than a non-A-type woman), divided by the odds that a non-A-type man marries an A-type woman; i.e., $Odds_A = (C_{AA}/C_{AA^c})/(C_{A^cA}/C_{A^cA^c})$ where C_{AA} is the proportion of (A, A) type marriage. Odds ratios greater than one indicate people tend to marry someone within their group. The larger the odds ratio, the stronger such tendency is. We summarize the odds ratios in the last row of the table. For example, in 2010 the odds ratios for L-type, M-type, H-type agents are 31.92, 6.42, 11.25 respectively. The odds ratios reveal some interesting trends that is not captured by the correlation coefficients. We observed that the odds ratios for L-type agents increase dramatically over time. On the other hand, the odds ratios for H-type agents decrease slightly. The odds ratios of M-type agents are the smallest among the three education categories, and it is also the most stable one over time. They are more likely to marry someone who belongs to different education category. These facts suggest that the change of marginal distributions may not be the only driving force for the change of the marriage market outcomes. It is possible that agents' preferences also change over time.

4.2 Utility Specification

The utility of man i marry woman j is specified as

$$U_i(j) = \beta_1 Z^j + \beta_2 |X_i - Z^j| + \epsilon_i(Z^j),$$

and the utility of woman j marry man i is given by

$$V^j(i) = \beta_3 X_i + \beta_4 |X_i - Z^j| + \eta^j(X_i),$$

where X_i and Z^j are respectively man i's and woman j's education level. This specification is motivated by Kalmijn (1998) who gives a throughout discussion on the role of education in marriage markets. Similar specifications are also adopted in previous studies; e.g., Logan, Hoff and Newton (2008); Echenique, Lee and Shum (2010). First, education tends to link with better labor market outcome. Therefore, the expected sign of β_1 (β_3) is positive as having an educated spouse one typically enjoys higher household income. We call the first term $\beta_1 Z^j$ ($\beta_3 X_i$) as the "main effect" of education. Second, differences in education level usually lead to differences in knowledge, taste and lifestyle. Since marriage involves many joint activities, having similar taste and lifestyle can enlarge the opportunity of joint activities. Having a similar knowledge also simplifies the joint household decision problem, and provides a common basis for mutual understanding. Therefore, the expected sign of β_2 (β_4) is negative. We interpret the second term $\beta_2|X_i - Z^j|$ ($\beta_4|X_i - Z^j|$) as the mismatch effect. Agents' preference profile will depends on the relative magnitude of the main and the mismatch effect. If the main effect is large enough, then everyone would prefer to marry someone who has a college degree. If the mismatch effect is non-negligible, then it is not optimal for an lesser-educated agent to marry a highly-educated spouse as the mismatch effect would dominate the main effect. The utility shocks $\epsilon_i(Z^j)$ and $\eta^j(X_i)$ are assumed to follow type I extreme value distribution. The IIA property of type I extreme value leads to a close form expression for the conditional probability of preference profiles; see McFadden (1984). Such assumption is especially useful when the type space is large since the large dimensional numerical integration can be avoid.

4.3 Estimation Results

We estimate the proposed utility function using the 2-step estimator suggested in equation 4. For each year, we have 204 (51 regions of residence and 4 age categories) restrictions on the contingency tables for the parameters of interest. Since each contingency table contains 4 non-redundant restriction, we end up with 816 moment restriction for 4 utility parameters. The results are summarized in Table 1 (M-optimal) and Table 2 (W-optimal). All coefficients are significant at 5% significance level. Regarding the main effects, we found substantial preference heterogeneity between men and women. In the case of M-optimal, besides 1990&95, the main effect for men is larger than that of women. In the case of W-optimal, the differences of main effects between men and women are smaller than the case of M-optimal. In the case of M-optimal, besides 1980&85, we found that men's main effects are increasing over the past few decades. A possible explanation is that the gender difference in wage is decreasing, and hence women's human capital is becoming more attractive to men. We also find similar pattern in the case of W-optimal.

Regarding the mismatch effect, the mismatch effect for men is generally larger than that of women. The exception is the case of M-optimal in 1980&85. It is also interesting to note that the estimated men's mismatch effects are quite stable over time. To sum up, we find that men care more about similarity in education level when choosing marital partners. On the other hand, women care less about spouses' education level. However this result should be interpreted with some caution. As the parameters in discrete choice models are only identified up to scale normalization, an alternative explanation is that the variance of random utility shocks are larger for women.[15] Both explanations imply that

[15]Similar results that indicate the variance of utility shocks for women is larger can be found in Chiappori,

women's partner choice behavior is more difficult to predict by education.

5 Pre-Martial Education Choice and Counterfactual Analysis

Virtually all of the empirical work on matching assumes that observed individual heterogeneity is exogenously given prior to matching market entry. The pioneer work by Flinn (2011), and Chiappori, Iyigun and Weiss (2009) point out the potential pitfall of such an assumption. Take education as an example, going to college not only improves the future income stream, which can improves the attractiveness in the marriage market, but also increases the opportunity to marry high-educated spouse. Consequently, a policy analysis will be misleading if we fail to take into account agents' optimization behaviors. Although it is quite common to treat education as endogenous in labor economics literature, such problem is usually ignore in the context of empirical matching work. As pointed out in Chiappori, Salanie and Weiss (2011), it is quite difficult to build an empirical education choice model that incorporates the marriage market consideration. The reason being that there is lack of a good measurement of return to schooling within marriage. All we observed is who marry who. In contrast to the case of labor market, the return can be quantified by the wage distribution.

We propose an estimable education choice model that also incorporates the marriage market consideration. Based on the estimator we proposed, the expected utility of marriage can be computed from the contingency table of marriage types. It serves the role of return to schooling within marriage. Our model is distinct from other pre-marital investment games in the sense that we do not impose strong assumptions on the marriage outcome. Instead, we use the observed marriage outcome to back out the structure parameters governing the education choice and preferences over education level.

5.1 Model

I propose a simple two-period game to model agents' education choices and the aggregate marriage market outcome. At the first period, agents decide whether to enter college or not. If the agent goes to college, he/she will incur an education cost c_i, which is his/her own private information. Such cost may reflect the opportunity cost of entering into labor market late and preference heterogeneity. Since we will not explicitly model the labor market consideration, the education cost c_i here is the cost net of the labor market return. The education choice made in the first stage will then change agents' type in the second

Salanie and Weiss (2011) in the context of TU matching game.

stage, leading to a different preference over potential mates. Moreover, it will also changes agents' relative attractiveness in the marriage market. Agents need to forecast all these effects to determine the benefit of going to college. At the second time period, each agent draws a utility shocks according to assumption 1 and form his/her preference profiles over potential mates. The matchmaker then use the deferred-acceptance algorithm to match men and women. We further assume agents' information structure. Agents know the parameter values in the utility functions, the distributions of education cost in the first stage, and the distribution of preference shocks in the second stage, for both sides of the market.

Given the assignment rule of the second stage game, man i's problem at the first stage is given by

$$\begin{aligned} D_i^m &= 1 \text{ if he choose to go to college} \\ &= \mathbf{I}\{EU_m(1) - c_i > EU_m(0)\}, \end{aligned}$$

where $\mathcal{I}$ is indicator function. $EU_m(1)$ is the expected utility of a male college graduate in the second stage, and $EU_m(0)$ is the expected utility of a male non-college graduate. Let β be the vector of utility parameters, which is known to every player. Let P_M^e and P_W^e be the expected college attendance rate for men and women, respectively. These are the belief system of the game. To compute the expected utility, agents first use the deferred-acceptance algorithm to compute the implied aggregate matching at the second stage using (P_M^e, P_W^e, β). $EU_m(\cdot)$ can then be computed from the corresponding conditional distribution of the contingency table and the utility function. Therefore, $EU_m(\cdot)$ is a function of (P_M^e, P_W^e, β); $EU_m(\cdot) = EU_m(\cdot|P_M^e, P_W^e; \beta)$. Analogously, woman j's problem can be expressed as

$$\begin{aligned} D_j^w &= 1 \text{ if she choose to go to college} \\ &= \mathbf{I}\{EU_w(1) - c_i > EU_w(0)\}, \end{aligned}$$

where $EU_w(\cdot) = EU_w(\cdot|P_M^e, P_W^e; \beta)$. Clearly, the realized college attendance rate for men and women (P_M^s, P_W^s) are given by

$$\begin{aligned} P_M^s &= Pr(D^m = 1) = Pr(\Delta EU_m > c_m) = F_{c_m}(\Delta EU_m(P_M^e, P_W^e; \beta)) \text{ and} \\ P_W^s &= Pr(D^w = 1) = Pr(\Delta EU_w > c_w) = F_{c_w}(\Delta EU_w(P_M^e, P_W^e; \beta)), \end{aligned}$$

where $\Delta EU_j \equiv [EU_j(1) - EU_j(0)], j = \{m, w\}$, and (F_{c_m}, F_{c_w}) are the distribution functions of men's and women's education cost respectively. Finally, an equilibrium solution concept is needed to complete the specification of the model. We impose a rational expectation-type condition: $P_M^e = P_M^r = P_M$ and $P_W^e = P_W^r = P_W$.

Definition 9. *Given* $(F_{c_m}, F_{w_m}, \beta)$*, a pair of* (P_M, P_W) *constitute an equilibrium if it solves the following fixed point equations.*

$$\begin{aligned} P_M &= F_{c_m}(\Delta EU_m(P_M, P_W; \beta)) \\ P_W &= F_{c_w}(\Delta EU_w(P_M, P_W; \beta)). \end{aligned} \tag{5}$$

The equilibrium condition requires that each man (woman) has a correct expectation of how many men and women will go to college. It can also be interpreted as a variant of discrete choice social interaction models (e.g., Brock and Durlauf, 2007) in which agent's choice depends on the average choice made by other players. Similar assumptions of rational expectation have been made in the literature of pre-marital investment game to facilitate the computation of the expected utility of marriage game in the second stage; e.g., Peters and Siow (2002); Peters (2006). However, this paper differs from their approaches in an important aspect. To secure the existence of an equilibrium, a strong assumption on marriage outcome is usually assumed. Both papers assume perfect assortative matching in the second stage game. However, in real world we never observe such marriage outcome, and the model is immediately rejected by the data. By contrast, we only make weaker identification assumptions in the marriage market. The expected utility can be directly estimated from the contingency table of marriage types. Furthermore, we will show that the cost distribution can also be back out from the data. Consequently, we are able to perform a counterfactual analysis to investigate the treatment effect of shifting the marginal distribution of types, or preference over potential mates. To my best knowledge, such exercise has not yet been perform in the context of two-sided matching games and pre-marital investment games.

5.2 Estimation of the Distribution of Education Cost

Utilizing the fixed point equation (5), we show the distribution of education cost can be estimated from the observed contingency table of marriage types. We assume F_{c_m} is a one-parameter continuous distribution, with density $f_{c_m} > 0$ on $\mathcal{R}$. In particular, we will assume F_{c_m} is Gumbel distribution with an unknown location parameter θ_{c_m} and the scale parameter is normalized to one.[16] Gumbel distribution has a special function form of quantile function that enables us to solve θ_{c_m} analytically:

$$\theta_{c_m} = \Delta EU_m(P_M, P_W; \beta) + \log[-\log(P_M)] \tag{6}$$

[16]Given the nature of the discrete choice model, it is a convention to normalized the scale parameter to one. For example, the variance is not identified in Probit regression models and is assumed to be one.

Estimation of θ_{c_m} is straightforward. First, the proportion of college graduates P_M is estimated from the contingency table of marriage types. Second, $\Delta EU_m(P_M, P_W; \beta)$ can be easily computed from the estimated utility function and the contingency table.[17]

We compute the contingency table for each year using all observations, without using the age category and region of residence to define sub-markets. The estimated average education cost are depicted in Figure 8 (M-optimal) and Figure 9 (W-optimal).[18] Under the assumption of W-optimal, we find that women's education cost drops substantially since 1980, while men's education cost is quite stable over time. At 2010, women's average education cost is lower than men's. We also observe qualitatively similar pattern in the case of M-optimal. Since education cost c_i is the only source of heterogeneity in our model, such pattern is needed to rationalize the increasing college attendance rate of women.

There is a huge literature that try to explain the increased education level of women in the past few decades. First, there are more colleges established in recent years, leading to more education opportunity for women. Second, the return of schooling for women increases since the gender gap in wage drops. However, most papers explain the increase education level from the standpoint of labor market. In this paper we offer a quite different point of view, from the perspective of marriage market. If women's education cost drops exogenously, the college attendance rate for women will increase. This is the direct effect of cost reduction. It has another indirect effect on men: Given there are more female college graduate available in the market, men's college attendance rate should increase accordingly for the following two reasons. First, there is now a greater opportunity to marry an educated women. Second, a man will enjoy the main effect from an educated wife and suffer from no mismatch effect if he also go to college. As a result, men's college attendance rate also increases, even men's education cost distribution remains unchanged. Notice that P_M will not over-response in our model. If men's college attendance rate increase too much, the excess supply of men would implied higher probability to marry an uneducated women, resulting in lower expected gain of college attendance.

[17]The utility function we estimated before is based on a 3-by-3 contingency table, while the education choice model we proposed here is a binary choice model. To deal with the inconsistency of the model and the data, we assume that when agent does not go to college, he/she will be randomly assign to either type 1 ("less than high school") or type 2 ("high school and partial college") according to some distribution. Such distribution can be directly calculated by $(\frac{Pr(\text{type 1})}{Pr(\text{type 1})+Pr(\text{type 2})}, \frac{Pr(\text{type 2})}{Pr(\text{type 1})+Pr(\text{type 2})})$ from the contingency table of marriage types. $EU_j(0)$ is then computed by the sum of expected utility of being type 1 and type 2, weighted by the corresponding probability mass.

[18]The average costs are obtained by adding the Euler constant to the estimated location parameters $(\hat{\theta}_{c_m}, \hat{\theta}_{c_w})$.

5.3 Counterfactual Analysis

In this section we investigate the effect of changing marginal distribution of types on the marriage market, by exogenously shift the cost distribution.[19] Consider the case of 1990&95, the contingency table of marriage types and the estimated location parameters under the assumption of W-optimal equilibrium are summarized below:

Contingency table of marriage types in 1990&95

men/ women	L	M	H	row sum
L	0.237	0.076	0.005	0.318
M	0.088	0.340	0.047	0.475
H	0.009	0.092	0.106	0.207
column sum	0.334	0.508	0.158	1

$\hat{\theta}_{c_m} = 0.3$
$\hat{\theta}_{c_w} = 1$
corr coef=0.63
Kendall's $\tau = 0.602$

If we reduce θ_{c_w} to 0.8 while holding θ_{c_m} fixed, the implied contingency table becomes

[19] One can also analyze the effect of changing preference on the marriage market, holding the marginal distributions of types fixed.

Imputed contingency table of marriage types in 1990&95				
men/ women	L	M	H	row sum
L	0.201	0.067	0.010	0.278
M	0.067	0.301	0.047	0.415
H	0.019	0.068	0.22	0.307
column sum	0.287	0.437	0.276	1
$\hat{\theta}_{c_m} = 0.3$				
$\hat{\theta}_{c_w} = 0.8$				
corr coef=0.68				
Kendall's $\tau = 0.642$				

Even a 20% drop in θ_{c_w} will raise the proportion of (H,H) type marriage from 10% to 20%. Let's consider another scenario: $(\theta_{c_m}, \theta_{c_w}) = (0.3, 0.3)$. The contingency table becomes:

Imputed contingency table of marriage types in 1990&95				
men/ women	L	M	H	row sum
L	0.165	0.050	0.033	0.248
M	0.055	0.234	0.081	0.370
H	0.018	0.078	0.286	0.382
column sum	0.238	0.362	0.4	1
$\hat{\theta}_{c_m} = 0.3$				
$\hat{\theta}_{c_w} = 0.3$				
corr coef=0.62				
Kendall's $\tau = 0.578$				

A 70% drop in θ_{c_w} results in almost 300% increase in the (H,H) type marriage. Women's college attendance rate is even greater than men's. It is also interesting to note that the degree of assortative matching, measured by correlation coefficient and Kendall's τ, remains unchanged. To sum up, the tuition subsidy creates more college-educated couples, while the correlation of the joint distribution of education of married couples remains unchanged. Namely, the location shift of education cost leads to location shift of the contingency table, rather than shift its copula.

We believe our model provides a valuable tool for policy analysis in real world. For example, a decision maker can use it to evaluate the effect of a new tuition subsidy policy. In the theoretical ground, the proposed model allows us to endogenize the marginal distribution of types, which is usually assumed exogenously given in the empirical matching literature. We conclude this section by discussing the limitation of our education choice model. Because the complexity of the ΔEU_j function which depends on the deferred-acceptance algorithm, in this study we are unable to provide a sufficient condition to guarantee the existence of an equilibrium. But in our empirical exercise, it seems that the equilibrium does exist if the postulated location parameters do not differ from the estimated parameters too much.

6 Extension: Moment Inequality Approach

In this section we discuss an alternative approach without assuming certain assignment mechanism. Since the number of equilibria may be large when there are many players, finding all equilibria is computationally cumbersome. Instead of solving the game, we exploit the no-blocking-pair (NBP) condition to derive a simple estimable restrictions. We extend the result of Echenique, Lee and Shum (2010) in two aspects. First, we utilize the frequency information of contingency table of observed matchings to derive the moment inequalities. Second, a lower bound for the probability of the observed matchings is also provided. We first use a 2-by-2 aggregate matching to illustrate the idea. It can then be easily extended to general cases.

Since the NBP condition is defined on the individual level, we should slightly modify the definition of the aggregate matching market:

Definition 10. *(Aggregate Matchings: Data)*
An 2-by-2 aggregate matching $\mathbf{D}$ *is the following 2-by-2 contingency table of marriage types in terms of count:*

type of men/women	*1*	*2*	*row sum*
1	N_{11}	N_{12}	N_1^m
2	N_{21}	N_{22}	N_2^m
column sum	N_1^w	N_2^w	

where N_{ij} *is the* (i,j) *entry of the matrix* $\mathbf{D}$*, and it represents the number of type* (i,j) *matches. The sum of the* j*-th column is the number of the* j*-type women, denoted by* N_j^w*. Analogously, The sum of the* i*-th row is the number of the* i*-type men, denoted by* N_i^m*. We will refer to* $\boldsymbol{N^m} = (N_1^m, N_2^m)$ *as the realized marginal distribution of men's types, and* $\boldsymbol{N^w} = (N_1^w, N_2^w)$ *as the realized marginal distribution of women's types.*

Besides the definition of available data is modified, other definitions and assumptions remains the same as in definition 8 for the stochastic aggregate matching market. In this section, the identification analysis shall proceed as a study of the relationship between the structural parameters $(\boldsymbol{P^m}, \boldsymbol{P^w})$ and the distribution of the contingency table $\mathbf{D}$. The following lemma can be used to form the lower bound of $Pr(\mathbf{D}|\boldsymbol{N^m}, \boldsymbol{N^w})$.

Lemma 3. *The observed matching* $\mathbf{D}$ *can be rationalized by the mutually-most-preferred preference profile. Namely, within each cell of the contingency table, each man and each woman prefer his/her status quo matching: There are* N_{11} *men who prefer type 1 to type 2 man;* N_{11} *women who prefer type 1 to type 2 men,* N_{12} *men who prefer type 2 to type 1 women;* N_{12} *women who prefer type 1 to type 2 men,* N_{21} *men prefer type 1 to type 2 women;* N_{21} *women who prefer type 2 to type 1 men, and* N_{22} *men who prefer type 2 to type 1 women;* N_{22} *women who prefer type 2 to type 1 men.*

Under this preference profile, the observed matching is both M and W optimal stable matching. Everybody gets his/her most preferred partners. Since the mutually-most-preferred preference profile leads to $\mathbf{D}$, it immediately follows that

$$Pr(\text{mutually-most-preferred}|\boldsymbol{N^m}, \boldsymbol{N^w}) \leq Pr(\mathbf{D}|\boldsymbol{N^m}, \boldsymbol{N^w})$$

To control for the relative scarcity, we shall conditional on the realized marginal distribution of agents' types. The i.i.d. sampling procedure in assumption 3 identifies $Pr(\mathbf{D}|\boldsymbol{N^m}, \boldsymbol{N^w})$. By assumption 2, $Pr(\text{mutually-most-preferred}|\boldsymbol{N^m}, \boldsymbol{N^w})$ can be factored into

$$\begin{aligned}&\Big(\frac{N_1^m!}{N_{11}!N_{12}!}(P^m_{12|1})^{N_{11}}(1-P^m_{12|1})^{N_{12}}\Big)\Big(\frac{N_2^m!}{N_{21}!N_{22}!}(P^m_{12|2})^{N_{21}}(1-P^m_{12|2})^{N_{22}}\Big)\times\\&\Big(\frac{N_1^w!}{N_{11}!N_{21}!}(P^w_{12|1})^{N_{11}}(1-P^w_{12|1})^{N_{21}}\Big)\Big(\frac{N_2^w!}{N_{12}!N_{22}!}(P^w_{12|2})^{N_{12}}(1-P^w_{12|2})^{N_{22}}\Big)\end{aligned} \tag{LB}$$

Next we show the upper bound for $Pr(\mathbf{D}|\boldsymbol{N^m}, \boldsymbol{N^w})$ can be derived from the no-blocking-pair (NBP) condition. It immediately follows that

$$Pr(\mathbf{D}|\boldsymbol{N^m}, \boldsymbol{N^w}) \leq Pr(NBP|\boldsymbol{N^m}, \boldsymbol{N^w}).$$

The event on the LHS implies the event on the RHS by stability. However, the event on the RHS does not necessarily imply to the event on the LHS due to multiple equilibria. Given the preference profile satisfying NBP condition, it may lead to a different contingency table implied by the other stable aggregate matching. Since we are agnostic about the equilibrium selection, the inequality cannot be replaced by the equality. Next we show how to compute $Pr(NBP|\boldsymbol{N^m}, \boldsymbol{N^w})$. We have to rule out blocking pairs of any kind.

This can be done by comparing any couple in $(1,1)$-type sub-market versus any couple in $(2,2)$-type sub-market, and comparing any couple in $(1,2)$-type sub-market versus any couple in $(2,1)$-type sub-market. Notice that there is no need to compare couples between type $(1,1)$ and $(1,2)$. A switch of husbands by two women will leave the contingency table unchanged (Graham, 2011). Therefore all we need is comparing two matches (i,j) and $(i^{'},j^{'})$ such that $X_i \neq X_{i'}$ and $Z^j \neq Z^{j^{'}}$. It follows that the NBP condition in the 2-by-2 case can be expressed as the following event:

$$
\begin{aligned}
&NBP = (\text{No type 1 man in the } (1,1)\text{-type sub-market prefers type 2 women} \cup \\
&\text{No type 2 woman in the } (2,2)\text{-type sub-market prefers type 1 men}) \\
&\cap(\text{No type 1 woman in the } (1,1)\text{-type sub-market prefers type 2 men} \cup \\
&\text{No type 2 man in the } (2,2)\text{-type sub-market prefers type 1 women}) \\
&\cap(\text{No type 1 man in the } (1,2)\text{-type sub-market prefers type 1 women} \cup \\
&\text{No type 1 woman in the } (2,1)\text{-type sub-market prefers type 1 men}) \\
&\cap(\text{No type 2 woman in the } (1,2)\text{-type sub-market prefers type 2 men} \cup \\
&\text{No type 2 man in the } (2,1)\text{-type sub-market prefers type 2 women}) \\
&\equiv A_1 \cap A_2 \cap A_3 \cap A_4
\end{aligned} \tag{5}
$$

In general the probability of NBP is difficult to compute, but the independence assumption of utility shocks admits a simple closed-form expression:

Theorem 8. *Under assumption 2,* $Pr(NBP|\boldsymbol{N^m}, \boldsymbol{N^w}) = \prod_{i=1}^{4} Pr(A_i|\boldsymbol{N^m}, \boldsymbol{N^w})$ *in the 2-by-2 case and is equal to*

$$
\begin{aligned}
&\Big((P^m_{12|1})^{N_{11}} + (1-P^w_{12|2})^{N_{22}} - (P^m_{12|1})^{N_{11}}(1-P^w_{12|2})^{N_{22}}\Big) \\
&\times\Big((P^w_{12|1})^{N_{11}} + (1-P^m_{12|2})^{N_{22}} - (P^w_{12|1})^{N_{11}}(1-P^m_{12|2})^{N_{22}}\Big) \\
&\times\Big((1-P^m_{12|1})^{N_{12}} + (1-P^w_{12|1})^{N_{21}} - (1-P^m_{12|1})^{N_{12}}(1-P^w_{12|1})^{N_{21}}\Big) \\
&\times\Big((P^w_{12|2})^{N_{12}} + (P^m_{12|2})^{N_{21}} - (P^w_{12|2})^{N_{12}}(P^m_{12|2})^{N_{21}}\Big)
\end{aligned} \tag{UB}
$$

Proof: see appendix.

To sum up, our identification arguments suggest the following theorem

Theorem 9. *The no-blocking-pair condition implies the following moment inequalities for the aggregate matching markets:*

$$LB \leq Pr(\mu|\boldsymbol{N^m}, \boldsymbol{N^w}) \leq UB.$$

Notice that the bound in theorem 9 may not be sharp. Besides the mutually-most-preferred profile, there may exist other preference profiles that also lead to $\mathbf{D}$. However, to find all such profiles one has to find all equilibria for all possible preference profiles, which is almost impossible in large markets. Therefore, even though the bound we provide only identify the outer region of the sharp identified set, it is feasible to implement in practice. The proposed upper bound and lower bound can be easily calculated, and are free from the curse of dimensionality induced by large number of players. Moreover, we improve on the bound derived by Echenique, Lee and Shum (2010) in two aspects. First, we exploit the frequency information of contingency table of observed matchings to derive the moment inequalities. Second, a lower bound for the probability of the observed matchings is also provided.

6.1 Estimation

Following Manski and Tamer (2002), and Ciliberto and Tamer (2009), the moment inequality in theorem 8 can be estimated by the modified minimum distance estimator:

$$\begin{aligned} Q_T(\theta) = \ & \frac{1}{T}\textstyle\sum_{t=1}^{T} \Big\{ \Big\| [P_T(\mathbf{D}_t|\boldsymbol{N^m}, \boldsymbol{N^w}) - LB(\boldsymbol{N^m}, \boldsymbol{N^w}; \theta)]_- \Big\| + \\ & \Big\| [P_T(\mathbf{D}_t|\boldsymbol{N^m}, \boldsymbol{N^w}) - UB(\boldsymbol{N^m}, \boldsymbol{N^w}; \theta)]_+ \Big\| \Big\}. \end{aligned}$$

where $(a)_- = (a\mathbf{I}[a \leq 0])$ and $(a)_+ = (a\mathbf{I}[a \geq 0])$. $\|\cdot\|$ is the Euclidian norm. $P_T(\mathbf{D}_t|\boldsymbol{N^m}, \boldsymbol{N^w})$ is the estimated probability of observed matching, conditional on covariates. The solution of $Q_T(\theta)$ is set-valued in general. Several inferential methods for set-valued estimators have been proposed recently; e.g., Chernozhukov, Hong and Tamer (2007). There are two major differences between the estimator of aggregate matching game and that of entry game studied in Ciliberto and Tamer (2009). Fist, UB and LB here can be computed analytically, while in Ciliberto and Tamer (2009) it is obtained by simulation and the game is solved for each draw. Second, they use a nonparametric estimator for the choice probability $P_T(\mathbf{D}_t|\boldsymbol{N^m}, \boldsymbol{N^w})$. However, nonparametric estimation is difficult to implement in aggregate matching games because of the sparse observations. In contrast to the entry game, the set of players can be different in different markets (consider the marriage couples in NY and CA), and so does the number of players. If the number of players is different in each market, then there is only one observation to estimate $P_T(\mathbf{D}_t|\boldsymbol{N^m}, \boldsymbol{N^w})$. To circumvent this problem, we propose a likelihood-based approach to estimate the conditional choice probabilities.

We exploit the fact that the contingency table follows a multinomial distribution. If we can estimate parameters characterizing such distribution, the conditional choice probabilities can then be obtained. Suppose the probability of observing a (i, j)-type matching is given by C_{ij}, then the probability of observing $\mathbf{D}$ with N couples is given by

$$Pr(\mathbf{D}) = \frac{N!}{N_{11}!N_{12}!N_{21}!N_{22}!} C_{11}^{N_{11}} C_{12}^{N_{12}} C_{21}^{N_{21}} C_{22}^{N_{22}}.$$

The likelihood of observing $\{\mathbf{D}_t\}_{t=1}^T$ is

$$\prod_{t=1}^T \left[\frac{N^t!}{N_{11}^t!N_{12}^t!N_{21}^t!N_{22}^t!} C_{11}^{N_{11}^t} C_{12}^{N_{12}^t} C_{21}^{N_{21}^t} C_{22}^{N_{22}^t} \right],$$

and C_{ij} can be consistently estimated by MLE. Next we demonstrate how to calculate the conditional choice probability $Pr(\mathbf{D}|\boldsymbol{N^m}, \boldsymbol{N^w})$. Conditional on the realized marginal distribution $(\boldsymbol{N^m}, \boldsymbol{N^w})$, the set of all possible stable matching is given by

$$\begin{aligned}
&\Omega_\nu(\boldsymbol{N^m}, \boldsymbol{N^w}) \\
&= \{N_{11} \text{ is an integer}|N_{11} \in [0 \cdot \mathbf{I}_{[N_1^w \le N_2^m]} + (N_1^w - N_2^m) \cdot \mathbf{I}_{[N_1^w > N_2^m]}, \min(N_1^m, N_1^w)]\}.
\end{aligned}$$

The second equality provides a simple characterization of $\Omega_\nu(\boldsymbol{N^m}, \boldsymbol{N^w})$. It says that the set of all possible stable matching can be indexed by N_{11}, and N_{11} must belong to some interval. This is because in the 2-by-2 contingency table, knowing the number in one of the four cells suffices to characterize the whole table, as the numbers in the rest three cells are determined by the adding-up constraints. It follows that

$$Pr(\mathbf{D}|\boldsymbol{N^m}, \boldsymbol{N^w}) = \frac{Pr(\mathbf{D})}{\sum_{\boldsymbol{\nu} \in \Omega_\nu(\boldsymbol{N^m}, \boldsymbol{N^w})} Pr(\boldsymbol{\nu})}.$$

7 Conclusion

We contribute to the empirical two-sided matching literature in two aspects. First, we present some new identification results with and without equilibrium selection, and propose consistent estimators which are free from the curse of dimensionality induced by the large number of players. Second, the effects of resource constraints on the market outcome are evaluated via a structural pre-marital education choice model. To my best knowledge, the counterfactual analysis in two-sided matching game is absent from the existing literature.

Several future extensions are possible. First, in this paper we do not consider the case when agents can remain single. Introducing such outside option may help achieve tighter identification region. Other properties of stable matchings can be further exploited. For example, the set of players who remain single is the same for all stable matchings. Therefore, we can ignore the problem of multiplicity when computing the marriage rate implied by the model. Second, our approach does not allow for continuous covariates. When this is the case, the researchers have to discretize the continuous variables first. Such procedure is nevertheless arbitrary and may create the problem of misclassification, and hence it is important to integrate the misclassification problem into the analysis.

8 Appendix

8.1 Pseudo Code for the Modified Deferred-Acceptance Algorithm

The modified deferred-acceptance algorithm require two inputs: the conditional probabilities of preference profiles $(\boldsymbol{P}^m, \boldsymbol{P}^w)$ characterizing the preference profile, and the marginal distributions of types $(\boldsymbol{f}^m, \boldsymbol{f}^w)$. Suppose men and women can be categorized into N types. As a result, conditional on the type agents can be further categorize into $N!$ types according to their preference list. Such preference list will determine their proposing/accepting strategy throughout the algorithm. The preference list will be indexed by l, with $l[k]$ being the k-th element of the preference list l (the k-th most favorite type). The size of the i-type men who have preference list l is given by $\mathrm{size}^{\mathrm{m}}_{\mathrm{l|i}} = f^m_i \times P^m_{l|i}$. We shall create an array $\mathrm{S}_{\mathrm{info}}$ to store the size of men with different type and preference list. Similarly we can define $\mathrm{size}^{\mathrm{w}}_{\mathrm{l|j}} = f^w_j \times P^w_{l|j}$ and $\mathrm{D}_{\mathrm{info}}$. We illustrate the algorithm assuming M-optimal. Finally we define a couple of notations. We refer to M_{DS} as the bimatrix of market demand and supply, $\mathrm{M}_{\mathrm{RDS}}$ as the bimatrix of residual demand and supply, $\mathbf{C^i}$ as the number of matches created in round i, $\mathrm{ES}_{\mathrm{ij}}$ as the excess supply in the (i, j)-type sub-market, and $\mathrm{ED}_{\mathrm{ij}}$ as the excess demand in the (i, j)-type sub-market.

step 1. Create the Market Supply

For the i-type men with preference list l, they first check whether the $l[1]$-type women are available or not from $\mathrm{D}_{\mathrm{info}}$. If yes, they propose to the $l[1]$-type women. The market supply in the $(i, l[1])$-type sub-market (the left argument of the $(i, l[1])$ cell in M_{DS})is then increased by $\mathrm{size}^{\mathrm{m}}_{\mathrm{l|i}}$. If not, they proceed to check the availability of the $l[2]$-type women.

This searching procedure continues until they find someone to propose to. Apply the above procedure to the other men who have different type and preference list.

step 2. Create the Market Demand

For the j-type women with preference list l, they first investigate the market supply in $\mathrm{M_{DS}}$. If the left argument of the $(l[1], j)$ cell in $\mathrm{M_{DS}}$ is nonzero, they submit the market demand orders to that sub-market. Namely, the right argument of the $(l[1], j)$ cell is increased by $\mathrm{size^{w}_{l|j}}$. If the left argument of the $(l[1], j)$ cell is zero, they proceed to check the left argument of the $(l[2], j)$ cell. This searching procedure continues until they find someone to submit the demand orders. Apply the above procedure to the other women who have different types and preference lists.

step 3. Match the Demand and Supply

The number of matches created in the 1-st round $\mathbf{C}^1$ is obtained by taking the minimum of each cell in $\mathrm{M_{DS}}$. Namely, the (i, j) cell in $\mathbf{C}^1$ equals to the minimum of the (i, j) cell in $\mathrm{M_{DS}}$:

$$C^1_{ij} = \min(\mathrm{M_{DS}}(i, j))$$

step 4. Calculate the Residual Supply and Update men's Distribution of Types and Preference Lists

We refer to $\Omega^{\mathrm{m}}_{\mathrm{j|i}}$ as the set of preference lists of the i-type men who propose to the j-type women in this round. If the left argument of $\mathrm{M_{DS}}(i, j)$ is smaller than or equal to C^1_{ij}, then there exists no excess supply. In $\mathrm{S_{info}}$, set the size of the i-type men who submit their proposals to the (i, j) sub-market to zero. Namely, set $\mathrm{size^{m}_{l|i}} = 0, \forall l \in \Omega^{\mathrm{m}}_{\mathrm{j|i}}$. If the left argument of $\mathrm{M_{DS}}(i, j)$ is greater than C^1_{ij}, then there exists excess supply of the i-type men. The excess supply $\mathrm{ES_{ij}}$ equals to the left argument of $\mathrm{M_{DS}}(i, j)$ minus C^1_{ij}. The corresponding elements of $\mathrm{S_{info}}$ are updated by

$$\mathrm{size^{m}_{l|i}} = \mathrm{ES_{ij}} \times \frac{\mathrm{size^{m}_{l|i}}}{\sum_{l \in \Omega^{\mathrm{m}}_{\mathrm{j|i}}} \mathrm{size^{m}_{l|i}}}.$$

Apply the above procedure to all cells in $\mathrm{M_{DS}}$ and $\mathbf{C^1}$.

step 5. Calculate the Residual Demand and Update women's Distribution of Types and Preference Lists

We refer to $\Omega^{\mathrm{w}}_{\mathrm{i|j}}$ as the set of preference lists of the j-type women who submit their demand orders to the i-type men in this round. If the right argument of $\mathrm{M_{DS}}(i, j)$ is smaller than or equal to C^1_{ij}, then there exists no excess demand. In $\mathrm{D_{info}}$, set the size of j-type women who submit their demand orders to the (i, j) sub-market to zero. Namely, set $\mathrm{size}^{\mathrm{w}}_{\mathrm{l|j}} = 0, \forall l \in \Omega^{\mathrm{w}}_{\mathrm{i|j}}$. If the right argument of $\mathrm{M_{DS}}(i, j)$ is greater than C^1_{ij}, then there exists excess demand. The excess demand $\mathrm{ED_{ij}}$ equals to the right argument of $\mathrm{M_{DS}}(i, j)$ minus C^1_{ij}. The corresponding elements of $\mathrm{D_{info}}$ are updated by

$$\mathrm{size}^{\mathrm{w}}_{\mathrm{l|j}} = \mathrm{ED_{ij}} \times \frac{\mathrm{size}^{\mathrm{w}}_{\mathrm{l|j}}}{\sum_{l \in \Omega^{\mathrm{w}}_{\mathrm{i|j}}} \mathrm{size}^{\mathrm{w}}_{\mathrm{l|j}}}.$$

Apply the above procedure to all cells in $\mathrm{M_{DS}}$ and $\mathbf{C^1}$.

step 6. Stopping Rule

Set $\mathrm{M_{DS}}$ and $\mathrm{M_{RDS}}$ to null matrix. Return to step 1 and iterate until there is no excess supply and demand. The desired contingency table is given by $\mathbf{C}(\boldsymbol{f^m}, \boldsymbol{f^w}, \boldsymbol{P^m}, \boldsymbol{P^w}) = \mathbf{C^1} + \mathbf{C^2} + \cdots + \mathbf{C^K}$, where K is the number of iterations.

8.2 Proofs

Proof of Theorem 2

First, at each round, each type of men will propose to some types of women. Consequently, there will be non-zero supply in some sub-markets. Since women only submit their demand orders to the sub-markets with non-zero supply, there exists at least one sub-market with non-zero demand and supply. Hence, at each round a positive number of matches will be created. Second, suppose at round 1 there are m_1 unit measure of matches have been created. Clearly, the size of residual men will equal to the size of residual women, which is $1 - m_1$. Therefore, at each round the total size of men equals to the total size of woman, implying that everyone can potentially get married in this round. In the end of the algorithm it is impossible to remain single due to shortage of men or women. Third, to keep the size equal in each round, it must be that in some sub-markets there exist residual men and in some sub-markets there exists residual women. Consider the (i, j)-type sub-market in which there exists residual j-type women. The algorithm will set the size to zero for those i-type men who propose to the i-type women in this round. Therefore, some types of men with some preference lists will be eliminated in each round, and some types of women with some preference lists will be eliminated in each round. Since the size of residual men and women is a strictly decreasing sequence and there are only finitely many types in which the size will be reduced at each iteration, at some point there will be only one type of men and one type of women with equal size. They will get married in this round, and the algorithm thus converges. *Q.E.D.*

Proof of Theorem 3

First, notice that the $\text{size}^{\text{m}}_{\text{l}|\text{i}}$ and $\text{size}^{\text{w}}_{\text{l}|\text{j}}$ is a continuous function of $(\boldsymbol{f^m}, \boldsymbol{f^w}, \boldsymbol{P^m}, \boldsymbol{P^w})$, which implies that the distribution of proposals made by men and the distribution of demand orders made by women are continuous in $(\boldsymbol{f^m}, \boldsymbol{f^w}, \boldsymbol{P^m}, \boldsymbol{P^w})$ in the first round. Second, consider two different configurations of marginal distributions $(\boldsymbol{f^m_1}, \boldsymbol{f^w_1})$ and $(\boldsymbol{f^m_2}, \boldsymbol{f^w_2})$, with $||(\boldsymbol{f^m_1}, \boldsymbol{f^w_1}) - (\boldsymbol{f^m_2}, \boldsymbol{f^w_2})|| < \epsilon$. Recall that men's proposing strategies in the first round only depends on $\boldsymbol{f^m}$. As long as the difference between $\boldsymbol{f^w_1}$ and $\boldsymbol{f^w_2}$ is small, men's proposing strategies will be the same. Same argument also applies to women's strategies of submitting the demand order. These two facts imply that the bimatrix of market demand and supply M_{DS} is continuous. Since the number of matches created $\mathbf{C^1}$ is obtained by taking the minimum of M_{DS} element-wisely and the function $\min(\cdot, \cdot)$ is a continuous operator, it follows that $\mathbf{C^1}$ is continuous in $(\boldsymbol{f^m}, \boldsymbol{f^w}, \boldsymbol{P^m}, \boldsymbol{P^w})$. Moreover, recall that the residual demand/supply is obtained by $\max(\text{M}_{\text{DS}} - \mathbf{C^1})$. A difference of two continuous function is still continuous, and hence the updated sequences of $\text{size}^{\text{m}}_{\text{l}|\text{i}}$ and $\text{size}^{\text{w}}_{\text{l}|\text{j}}$ for the second round. Repeat the above argument to each round. It follows that $\mathbf{C}^i$ is continuous for all i. Therefore, the sum of $\mathbf{C}^i$ is also continuous. *Q.E.D.*

Proof of Theorem 4

We proceed by first showing that the set of all possible contingency table given $(\boldsymbol{\hat{f}^m}, \boldsymbol{\hat{f}^w})$ can be characterized by the Frechet-Hoeffding bound. The upper bound can be reached by choosing $(P^m_{12|1}, P^m_{12|2}, P^w_{12|1}, P^w_{12|2}) = (1, 0, 1, 0)$. On the other hand, the lower bound can be reached by choosing $(P^m_{12|1}, P^m_{12|2}, P^w_{12|1}, P^w_{12|2}) = (0, 1, 0, 1)$. Second, the existence of $(\boldsymbol{P^m}, \boldsymbol{P^w})$ such that $\mathbf{C}(\boldsymbol{\hat{f}^m}, \boldsymbol{\hat{f}^w}, \boldsymbol{P^m}, \boldsymbol{P^w}) = \hat{\mathbf{C}}$ is followed by the intermediate value theorem.

claim 1.

Notice that in the 2-by-2 case, the contingency table can be fully characterized by a single number because of the adding-up constraints. Without loss of generality, it suffices to characterize the possible range of C_{11}, the proportion of (1, 1)-type couples. According to the Frechet-Hoeffding bound, we have $\max(f^m_1 + f^w_1 - 1, 0) \leq C_{11} \leq \min(f^m_1, f^w_1)$. This bound is sharp; see Nelson (2010).

claim 2.

The upper bound $\min(f^m_1, f^w_1)$ can be rationalized by choosing $(P^m_{12|1}, P^m_{12|2}, P^w_{12|1}, P^w_{12|2}) = (1, 0, 1, 0)$. The ideal is that the upper bound represents the case when men's and women's types are comonotonic. If men's preferences perfectly coincide with that of women, their types will be comonotonic. Under this setup, the bimatrix of demand and supply $\mathrm{M_{DS}}$ in the first round is given by

type of men/women	1	2	men's marginal
1	(f_1^m, f_1^w)	$(0,0)$	f_1^m
2	$(0,0)$	$(1-f_1^m, 1-f_1^w)$	$1-f_1^m$
women's margianl	f_1^w	$1-f_1^w$	

Clearly the number of $(1,1)$-type matches created in this round is $C_{11}^1 = \min(f_1^m, f_1^w)$. Moreover, in the subsequent round no more $(1,1)$-type will be created since there is no more 1-type men or 1-type women available. We conclude that $C_{11} = \min(f_1^m, f_1^w)$.

claim 3.

The upper bound $\min(f_1^m, f_1^w)$ can be rationalized by choosing $(P_{12|1}^m, P_{12|2}^m, P_{12|1}^w, P_{12|2}^w) = (0,1,0,1)$. The ideal is that the lower bound represents the case when men's and women's types are countermonotonic. If men's preferences perfectly mismatch with that of women, their types will be countermonotonic. Under this setup, the bimatrix of demand and supply $\mathrm{M_{DS}}$ in the first round is given by

type of men/women	1	2	men's marginal
1	$(0,0)$	$(f_1^m, 1-f_1^w)$	f_1^m
2	$(1-f_1^m, f_1^w)$	$(0,0)$	$1-f_1^m$
women's margianl	f_1^w	$1-f_1^w$	

Three cases are possible in the first round. If $1 - f_1^m = f_1^w$, then four sub-markets clear in the first round: $C_{11} = 0$. If $1 - f_1^m < f_1^w$, then $C_{21}^1 = 1 - f_1^m$. No more $(2,1)$-type couples will be created in the subsequent rounds because 2-type men no longer available. Since $\mathbf{C}$ must satisfies the adding-up constraints, it immediately follows that $C_{11} = f_1^w - (1 - f_1^m) > 0$. If $1 - f_1^m > f_1^w$ then $C_{21}^1 = f_1^w$. No more $(2,1)$-type couples will be created in the subsequent rounds because 1-type women no longer available. Therefore, it immediately follows that $C_{11} = 0$. We conclude that $C_{11} = \max(f_1^m + f_1^w - 1, 0)$.

claim 4.

Notice that $(P_{12|1}^m, P_{12|2}^m, P_{12|1}^w, P_{12|2}^w) \in [0,1]^4$ is a connected set, and by theorem 3 $\mathbf{C}$ is continuous. By claim 1,2,3 and the intermediate value theorem, it follows that there exist a sequence of $(P_{12|1}^m, P_{12|2}^m, P_{12|1}^w, P_{12|2}^w)$ such that $\mathbf{C}(\hat{\boldsymbol{f}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})$ equals to a given $\hat{\mathbf{C}}$ with marginal distributions $(\hat{\boldsymbol{f}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}^{\boldsymbol{w}})$ *Q.E.D.*

Proof of Lemma 1

By triangular inequality, the object of interest is bounded by

$$||\mathbf{C}(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) - \mathbf{C}(\boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{m}}, \boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})|| + ||\mathbf{C}^t - \hat{\mathbf{C}}^t||.$$

Since $E|\mathbf{I}_{[X_k^t=i, Z_k^t=j]}| < \infty$, and by the i.i.d. sampling assumption, $||\hat{\mathbf{C}}^t - \mathbf{C}^t|| = o_p(1)$ by the law of large numbers. Notice that the convergence does not depend on the unknown parameters $(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})$, hence the convergence is also uniform on $(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) \in \Theta$. Analogously, $(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}) \to (\boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{m}}, \boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{w}})$ almost surely, uniformly on Θ. Therefore, as N_t is large enough, $||(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) - (\boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{m}}, \boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})|| < \epsilon$ almost surely. By theorem 3, $\mathbf{C}$ is continuous in $(\boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{m}}, \boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})$. Since all arguments of $\mathbf{C}$ are probabilities, its domain is compact and the result of theorem 3 can be strengthened to be uniform continuous. It follows that $\sup_{(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})} ||\mathbf{C}(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) - \mathbf{C}(\boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{m}}, \boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})|| = o_p(1)$. *Q.E.D.*

Proof of Theorem 6

We prove the consistency by checking the regularity conditions in Newey and McFadden (1994). First, $Q_0(\beta)$ is uniquely minimized at β_0 and Θ is compact is trivially satisfied. Second, under the maintained assumptions $(\boldsymbol{P}^{\boldsymbol{m}}(\beta), \boldsymbol{p}^{\boldsymbol{w}}(\beta))$ are continuous in β. By theorem 3, it immediately follows that $Q_0(\beta)$ is continuous in β. The uniform convergence of $\hat{Q}_n(\beta)$ to $Q_0(\beta)$ is a direct result of lemma 1. *Q.E.D.*

Proof of Lemma 2

Since $(\boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) \in \Theta_I$,

$$\mathbf{C}(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) - \hat{\mathbf{C}}^t$$

$$= (\mathbf{C}(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}}) - \mathbf{C}(\boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{m}}, \boldsymbol{f}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})) + (\hat{\mathbf{C}}^t - \mathbf{C}^t)$$

Clearly, $(\hat{\mathbf{C}}^t - \mathbf{C}^t) = O_p(n^{-1/2})$ by the central limit theorem. The proof for the other is more involved. We will show the matrix representing the number of matches created in round k under $(\hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}_{\boldsymbol{t}}^{\boldsymbol{w}}, \boldsymbol{P}^{\boldsymbol{m}}, \boldsymbol{P}^{\boldsymbol{w}})$, denoted by $\hat{\mathbf{C}}^k$, converges to $\mathbf{C}^k$ at rate $\sqrt{n}$. Consider the (i,j)-type matchings created in the first round. $\hat{C}_{ij}^1$ is given by

$$\min\left(\hat{f}_i^m(\textstyle\sum_{l\in\Omega_{j|i}^{m,1}} P_{l|i}^m), \hat{f}_j^w(\sum_{l\in\Omega_{i|j}^{w,1}} P_{l|j}^w)\right)$$

Since $(\hat{f}_i^m, \hat{f}_j^w)$ converges to (f_i^m, f_j^w) at rate $\sqrt{n}$ by the central limit theorem, it follows that $\hat{C}_{ij}^1 - C_{ij}^1 = O_p(n^{-1/2})$. At the second round, the marginal distributions and the conditional probabilities of preference profiles are updated according to the residual demand

and supply. We use the notation $\hat{f}_i^{m,2}$ to represent the total size of i-type men in the second round. Similarly, we use the notation $\hat{P}_{l|i}^{m,2}$ to represent the conditional probability of preference profiles of a i-type man whose preference list is l at the second round. The number of (i,j)-type matchings created at the second round under $(\hat{\boldsymbol{f}}^{\boldsymbol{m}}, \hat{\boldsymbol{f}}^{\boldsymbol{w}})$, $\hat{C}_{ij}^2$, is given by

$$\min\left(\hat{f}_i^{m,2}(\textstyle\sum_{l\in\hat{\Omega}_{j|i}^{m,2}} \hat{P}_{l|i}^{m,2}), \hat{f}_i^{w,2}(\sum_{l\in\hat{\Omega}_{j|i}^{w,2}} \hat{P}_{l|i}^{w,2})\right),$$

while the population version C_{ij}^2 is given by

$$\min\left(f_i^{m,2}(\textstyle\sum_{l\in\Omega_{j|i}^{m,2}} P_{l|i}^{m,2}), f_i^{w,2}(\sum_{l\in\Omega_{j|i}^{w,2}} P_{l|i}^{w,2})\right).$$

The difference between the first components of $\hat{C}_{ij}^2$ and C_{ij}^2 can be written as

$$\left(\hat{f}_i^{m,2}(\textstyle\sum_{l\in\hat{\Omega}_{j|i}^{m,2}} \hat{P}_{l|i}^{m,2}) - \hat{f}_i^{m,2}(\sum_{l\in\Omega_{j|i}^{m,2}} P_{l|i}^{m,2})\right) + \left(\hat{f}_i^{m,2}(\sum_{l\in\Omega_{j|i}^{m,2}} P_{l|i}^{m,2}) - f_i^{m,2}(\sum_{l\in\Omega_{j|i}^{m,2}} P_{l|i}^{m,2})\right).$$

The next task is to evaluate the convergence rate of $\hat{f}_i^{m,2} - f_i^{m,2}$ and $\hat{P}_{l|i}^{m,2} - P_{l|i}^{m,2}$. $\hat{f}_i^{m,2}$ can be further decomposed into the sum of excess supply created in the first round

$$\textstyle\sum_{j=1}^M \hat{\text{ES}}_{ij}^1 = \sum_{j=1}^M \left(\hat{f}_i^m(\sum_{l\in\Omega_{j|i}^{m,1}} P_{l|i}^m) - \min(\hat{f}_i^m(\sum_{l\in\Omega_{j|i}^{m,1}} P_{l|i}^m), \hat{f}_j^w(\sum_{l\in\Omega_{i|j}^{w,1}} P_{l|j}^w))\right),$$

and $f_j^{m,2}$ can be expressed as similar form

$$\textstyle\sum_{j=1}^M \text{ES}_{ij}^1 = \sum_{j=1}^M \left(f_i^m(\sum_{l\in\Omega_{j|i}^{m,1}} P_{l|i}^m) - \min(f_i^m(\sum_{l\in\Omega_{j|i}^{m,1}} P_{l|i}^m), f_j^w(\sum_{l\in\Omega_{i|j}^{w,1}} P_{l|j}^w))\right).$$

Following the same argument as in the first round, we conclude that $\hat{f}_i^{m,2} - f_j^{m,2} = O_p(n^{-1/2})$. Finally we show that $\hat{P}_{l|i}^{m,2}$ can be decomposed into

$$\hat{P}_{l|i}^{m,2} = \frac{\hat{\text{size}}_{l|i}^{m,2}}{\sum_l \hat{\text{size}}_{l|i}^{m,2}},$$

$$\hat{\text{size}}_{l|i}^{m,2} = \hat{\text{ES}}_{ij}^1 \frac{\hat{\text{size}}_{l|i}^{m,1}}{\sum_{l\in\Omega_{j|i}^{m,1}} \hat{\text{size}}_{l|i}^{m,1}} = \hat{\text{ES}}_{ij}^1 \frac{P_{l|i}^{m,1}}{\sum_{l\in\Omega_{j|i}^{m,1}} P_{l|i}^{m,1}} \quad \text{for some } j$$

Analogously, $P_{l|i}^{m,2}$ can be written as

$$\text{ES}_{ij}^1 \frac{P_{l|i}^{m,1}}{\sum_{l\in\Omega_{j|i}^{m,1}} P_{l|i}^{m,1}},$$

and we just show that $\hat{\text{ES}}_{ij}^1 - \text{ES}_{ij}^1 = O_p(n^{-1/2})$. We conclude that C_{ij}^2 converges to C_{ij}^2 at rate $\sqrt{n}$. The same argument for all other rounds and other cells. Notice that our argument does not depend on the value of $(\boldsymbol{P^m}, \boldsymbol{P^w})$, and hence the convergence is uniform. *Q.E.D.*

Proof of Theorem 7

We proceed by checking the regularity condition C1 in Chernozhukov, Hong and Tamer (2007). Condition C1-a and c are trivially satisfied. Condition C1-b is implied by theorem 3. Condition C1-d is a direct consequence of lemma 1 and C1-e follows from lemma 2. $Q.E.D.$

Proof of Theorem 8

Before proceeding, some notations are introduced here to interpret the no-blocking-pair condition in terms of econometric terminology. We show how to compute $Pr(A_1|\boldsymbol{N^m}, \boldsymbol{N^w})$. Other conditional probabilities can be computed in the same manner. Take any matched pair $(i, j) \in (1, 1)$ sub-market and $(i^{'}, j^{'}) \in (2, 2)$ sub-market. Define the deviation indicator for man $i \in (1, 1)$.

$$d_i(j^{'} > j) = \mathbf{I}[\delta_m(1, 2) + \epsilon_i(2) > \delta_m(1, 1) + \epsilon_i(1)].$$

If $d_i(j^{'} > j) = 1$ then man i has an incentive to deviate from the status quo matching. The deviation indicator equals to 1 with probability $P^m_{21|1}$. Similarly the deviation indicator for woman $j^{'} \in (2, 2)$ sub-market is defined as

$$d_{j^{'}}(i^{>}i^{'}) = \mathbf{I}[\delta_w(1, 2) + \eta^{j^{'}}(1) > \delta_w(2, 2) + \eta^{j^{'}}(2)].$$

To form a blocking pair, both man i and woman $j^{'}$ have to agree on matching together. Define the blocking-pair indicator $Y_{ij^{'}} = d_i(j^{'} > j) \cdot d_{j^{'}}(i^{>}i^{'})$. Then $(i, j^{'})$ is a blocking pair if and only if $Y_{ij^{'}} = 1$. Notice that the blocking-pair indicator takes the form of double hurdle model (e.g. Jones, 1989).

Consider all matched pairs $\{(i_1, j_1), \ldots, (i_{N_{11}}, j_{N_{11}})\} \in (1, 1)$ sub-market and $\{(i^{'}_1, j^{'}_1), \ldots, (i^{'}_{N_{22}}, j^{'}_{N_{22}})\} \in (2, 2)$ sub-market. The event A_1 implies that $Y_{i_p j^{'}_k} = 0 \forall p \forall k$. Next we show how to compute the probability of event A_1 implied by the double hurdle structure. It can be decomposed into

$$\begin{aligned} &Pr(Y_{i_p j^{'}_k} = 0 \ \ \forall i_p \in (1, 1), \forall j^{'}_k \in (2, 2)) \\ &= Pr(\cdot|d_{i_1}(\cdot) = 1)Pr(d_{i_1}(\cdot) = 1) + Pr(\cdot|d_{i_1}(\cdot) = 0)Pr(d_{i_1}(\cdot) = 0), \end{aligned} \tag{A.1}$$

where $d_{i_1}(\cdot) \equiv d_{i_1}(j^{'}_1 > j_1)$.[20] If $d_{i_1}(\cdot) = 1$, it must be that $d_{j^{'}_k}(\cdot) = 0$ for all $j^{'}_k \in (2, 2)$ to guarantee $Y_{i_1 j^{'}_k} = 0$. It immediately follows that

$$Pr(\cdot|d_{i_1}(\cdot) = 1) = \textstyle\prod_{j^{'}_k \in (2,2)} Pr(d_{j^{'}_k}(\cdot) = 0). \tag{A.2}$$

[20] Recall that we assume agents' preferences only depend on observables. Therefore $d_{i_p}(j^{'}_x > j_p) = d_{i_p}(j^{'}_y > j_p)$ as long as $j^{'}_x$ and $j^{'}_y$ belong to the same type.

Now consider the second component of (A.1). $Pr(\cdot|d_{i_1}(\cdot)=0)$ can be further decompose into

$$\begin{aligned} Pr(\cdot|d_{i_1}(\cdot)=0) &= Pr(\cdot|d_{i_1}(\cdot)=0, d_{i_2}(\cdot)=1)Pr(d_{i_2}(\cdot)=1) \\ &+ Pr(\cdot|d_{i_1}(\cdot)=0, d_{i_2}(\cdot)=0)Pr(d_{i_2}(\cdot)=0). \end{aligned}$$

Again, If $d_{i_2}(\cdot)=1$, it must be that $d_{j_k'}(\cdot)=0$ for all $j_k' \in (2,2)$ in order to guarantee $Y_{i_2 j_k'}=0$. It immediately follows that

$$Pr(\cdot|d_{i_1}(\cdot)=0, d_{i_2}(\cdot)=1) = \prod_{j_k' \in (2,2)} Pr(d_{j_k'}(\cdot)=0). \tag{A.3}$$

Combining (A.1)-(A.3), we obtain

$$\begin{aligned} &Pr(Y_{i_p j_k'}=0 \;\; \forall i_p \in (1,1), \forall j_k' \in (2,2)) \\ &= Pr(d_{i_1}(\cdot)=1) \prod_{j_k' \in (2,2)} Pr(d_{j_k'}(\cdot)=0) \\ &+ Pr(d_{i_1}(\cdot)=0)Pr(d_{i_2}(\cdot)=1) \prod_{j_k' \in (2,2)} Pr(d_{j_k'}(\cdot)=0) \\ &+ Pr(d_{i_1}(\cdot)=0)Pr(d_{i_2}(\cdot)=0)Pr(\cdot|d_{i_1}(\cdot)=0, d_{i_2}(\cdot)=0). \end{aligned} \tag{A.4}$$

We can keep expanding $Pr(\cdot|d_{i_1}(\cdot)=0, d_{i_2}(\cdot)=0)$ by conditioning on d_{i_3} and then d_{i_4} utile $d_{i_{N_{11}}}$. Finally we can write (A.4) as

$$\begin{aligned} &Pr(Y_{i_p j_k'}=0 \;\; \forall i_p \in (1,1), \forall j_k' \in (2,2)) \\ &= \prod_{j_k' \in (2,2)} Pr(d_{j_k'}(\cdot)=0) \Big(Pr(d_{i_1}(\cdot)=1) + Pr(d_{i_1}(\cdot)=0)Pr(d_{i_2}(\cdot)=1) + \dots \Big) \\ &+ P(\cdot|d_{i_1}=0, \dots, d_{i_{N_{11}}}=0) \prod_{i_p \in (1,1)} Pr(d_{i_p}=0). \end{aligned} \tag{A.5}$$

When no man wants to deviate from the status quo matching ($d_{i_1}=0, \dots, d_{i_{N_{11}}}=0$), then $Y_{i_p j_k'}=0$ are trivially satisfied. Hence the term $P(\cdot|d_{i_1}=0, \dots, d_{i_{N_{11}}}=0)$ equals to 1. It follows that

$$\begin{aligned} &Pr(Y_{i_p j_k'}=0 \;\; \forall i_p \in (1,1), \forall j_k' \in (2,2)) \\ &= (P^m_{12|1})^{N_{11}} + (1-P^w_{12|2})^{N_{22}} - (P^m_{12|1})^{N_{11}}(1-P^w_{12|2})^{N_{22}} \end{aligned}$$

Q.E.D.

References

Abdulkadiroglu, A., P. A. Pathak and A. E. Roth (2005). The New York City High School Match, *AEA Papers and Proceedings*, **95**, 364–367.

Bajari, P., J. Hahn, H. Hong, and G. Ridder (2011). A Note on Semiparametric Estimation of Finite Mixtures of Discrete Choice Models with Application to Game Theoretic Models, *International Economic Review*, **52**, 807–824.

Banks, R. R. (2011). *Is Marriage for White People?: How the African American Marriage Decline Affects Everyone*, Dutton Adult.

Becker, G. S. (1973). A theory of marriage: part I, *Journal of Political Economy*, **81**, 813–846.

Becker, G. S. and K. M. Murphy (2000). *Social Economics: Market Behavior in a Social Environment*, Cambridge, MA: Harvard University Press.

Berry, S., J. Levinsohn and A. Pakes (1995). Automobile Prices in Market Equilibrium, *Econometrica*, **63**, 841–890.

Berry, S. and E. Tamer (2006). Identification in Models of Oligopoly Entry.

Boyd, D., H. Lankford, S. Loeb and J. Wyckoff (2006). Analyzing the Determinants of the Matching of Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets, NBER working paper.

Brock, W. A. and S. N. Durlauf (2007). Identification of Binary Choice Models with Social Interactions. *Journal of Econometrics*, **140**, 52–75.

Burkard, R., M. Dell'Amico and S. Martello (2009). *Assignment Problems*, Society for Industrial and Applied Mathematics, Philadelphia, PA.

Chen, J. (2008). Two-Sided Matching and Spread Determinants in the Loan Market, *Working Paper*

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and Inference on Identified Parameter Sets, *Econometrica*, **75**, 1243–1284.

Chiappori, P.-A., M. Iyigun and Y. Weiss (209). Investment in Schooling and the Marriage Market, *American Economic Review*, **99**, 1689–1713.

Chiappori, P.-A., B. Salanie and Y. Weiss (2011). Partner Choice and the Marital College Premium, *working paper*

Choo, E. and A. Siow (2006). Who Marries Whom and Why, *Journal of Political Economy*, **114**, 175–201.

Ciliberto, F. and E. Tamer (2009). Market Structure and Multiple Equilibria in Airline Markets. *Econometrica*, **77**, 1791–1828.

Dagsvik, J. H. (2000). Aggregation in Matching Markets, *International Economic Review*, **41**, 27–56.

Del Boca, D. and C. J. Flinn (2011). Household Time Allocation and Modes of Behavior: A Theory of Sorts. *Working Paper.*

Echenique, F. (2008). What Matchings can be Stable ? The Testable Implications of Matching Theory, *Mathematics of Operations Research*, **33**, 757–768.

Echenique, F., S. Lee and M. Shum (2010). Aggregate Matchings. *Working Paper.*

Flinn, C. J. (2011). Labor Market Search and Schooling Investment. *Working Paper.*

Fox, J. T. (2010). Identification in Matching Games, *Quantitative Economics*, **1**, 203–254.

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

Galichon, A. and B. Salanié (2010). Matching with Trade-offs: Revealed Preference over Competing Characteristics. *Working Paper.*

Graham, B. S., G. Imbens and G. Ridders (2010). Complementarity and Aggregate Implications of Assortative Matching: a Nonparametric Analysis, *Working Paper.*

Graham, B. S. (2011). Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers, *Handbook of Social Economics*, **1B**, 965–1052, Amsterdam, North-Holland.

Hitsch, G. J., A. Hortacsu and D. Ariely (2010a). What Makes You Click? Mate Preferences and Matching Outcomes in Online Dating, *Quantitative Marketing and Economics*, forthcoming.

Hitsch, G. J., A. Hortacsu and D. Ariely (2010b). Matching and Sorting in Online Dating, *American Economic Review*, forthcoming.

Hotz, V. J. and R. A. Miller (1993). Conditional Choice Probabilities and Estimation of Dynamic Models, *Annual Review of Sociology*, **60**, 497–529.

Jones, A. M. (1989). A Double-Hurdle Model of Cigarette Consumption, *Journal of Applied Econometrics*, **4**, 23–39.

Kalmijn, M. (1998). Intermarriage and Homogamy: Causes, Patterns, Trends, *Annual Review of Sociology*, **24**, 395–421.

Kamada Y. and F. Kojima (2010). Improving Efficiency in Matching Markets with Regional Caps: The Case of the JAPAN Residency Matching Program, *Working Paper.*

Logan, J. A., P. D. Hoff, and M. A. Newton (2008). Two-Sided Estimation of Mate Preferences for Similarities in Age, Education, and Religion, *Journal of the American Statistical Association*, **103**, 559–569.

McFadden, D. A. (1984) Econometric Analysis of Qualitative Response Models. *Handbook of Econometrics*, **Vol. 2**,chapter 24, 1395–1457.

Menzel, K. (2011). MCMC Methods for Incomplete Structural Models of Social Interactions, *Working Paper.*

Nelson, R. B. (2010). *An introduction to Copulas*, New York, NY: Springer.

Peters, M. and A. Siow (2002). Competing Premarital Investments, *Journal of Political Economy*, **110**, 592–608.

Peters, M. (2006). Pre-Marital Investment Game, *Working Paper.*

Roman, J. P. and A. M. Shaikh (2010). Inference for the Identified Set in Partially Identified Econometric Models, *Econometrica*, **78**, 169–211.

Roth, A. E. (1984). The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory, *Journal of Political Economy*, **92**, 991–1016.

Roth, A. E. and M. A. O. Sotomayor (1990). *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Econometric Society Monographs No. 18, Cambridge University Press.

Sorensen, M. (2007). How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital, *Journal of Finance*, **LXII**, 2725–2762.

Table 1: Estimation of Utility over Spouse's Education Level and Difference, Assuming M-Optimal Equilibrium

	Men		Women	
Time	β_1	β_2	β_3	β_4
2010	0.905	-1.810	0.551	-0.062
	[.38,.93]	[-1.81,-1.43]	[0.0,.4]	[-.55,-.25]
2000&05	0.485	-1.725	0.179	-0.401
	[.38,.63]	[-1.78,-1.66]	[.14,.35]	[-.4,-.28]
1990&95	0.188	-1.422	0.294	-0.638
	[.15,.31]	[-1.46,-1.40]	[.26,.44]	[-.78,-.51]
1980&85	1.240	-0.474	0.189	-0.653
	[.07,1.25]	[-.63,-.26]	[.16,.42]	[-.75,-.49]

The 95% subsampling confidence intervals are reported in brackets below the corresponding point estimates. The utility of man i marry woman j is given by $U_i(j) = \beta_1 Z^j + \beta_2 |X_i - Z^j| + \epsilon_i(Z^j)$, and the utility of woman j marry man i is given by $V^j(i) = \beta_3 X_i + \beta_4 |X_i - Z^j| + \eta^j(X_i)$, where X_i and Z^j are respectively man's and woman's education category

Table 2: Estimation of Utility over Spouse's Education Level and Difference, Assuming W-Optimal Equilibrium

	Men		Women	
Time	β_1	β_2	β_3	β_4
2010	0.590	-1.622	0.631	-0.278
	[.47,.85]	[-1.81,-1.52]	[.5,.74]	[-.41,-.19]
2000&05	0.600	-1.632	0.461	-0.101
	[.56,.73]	[-1.68,-1.59]	[.43,.57]	[-.15,-.01]
1990&95	0.225	-1.325	0.469	-0.803
	[.08,.22]	[-1.43,-1.28]	[.24,.75]	[-1.04,-.39]
1980&85	0.342	-0.652	0.524	-0.090
	[0.05,1.55]	[-1.45,0.49]	[0.07,2.4]	[-4.39,0.17]

The 95% subsampling confidence intervals are reported in brackets below the corresponding point estimates. The utility of man i marry woman j is given by $U_i(j) = \beta_1 Z^j + \beta_2 |X_i - Z^j| + \epsilon_i(Z^j)$, and the utility of woman j marry man i is given by $V^j(i) = \beta_3 X_i + \beta_4 |X_i - Z^j| + \eta^j(X_i)$, where X_i and Z^j are respectively man's and woman's education category.

Figure 1: The 3D plot of the probability of observing a (H, H)-type matching as the function of $(\boldsymbol{P^m}, \boldsymbol{P^w})$. The marginal distributions of types are $(f_H^m, f_H^w) = (0.3, 0.7)$. The x-axis represents $P_{HL|H}^m = P_{HL|H}^w$, the y-axis represents $P_{HL|L}^m = P_{HL|L}^w$, and the z-axis represents the probability of observing a (H, H)-type matching.

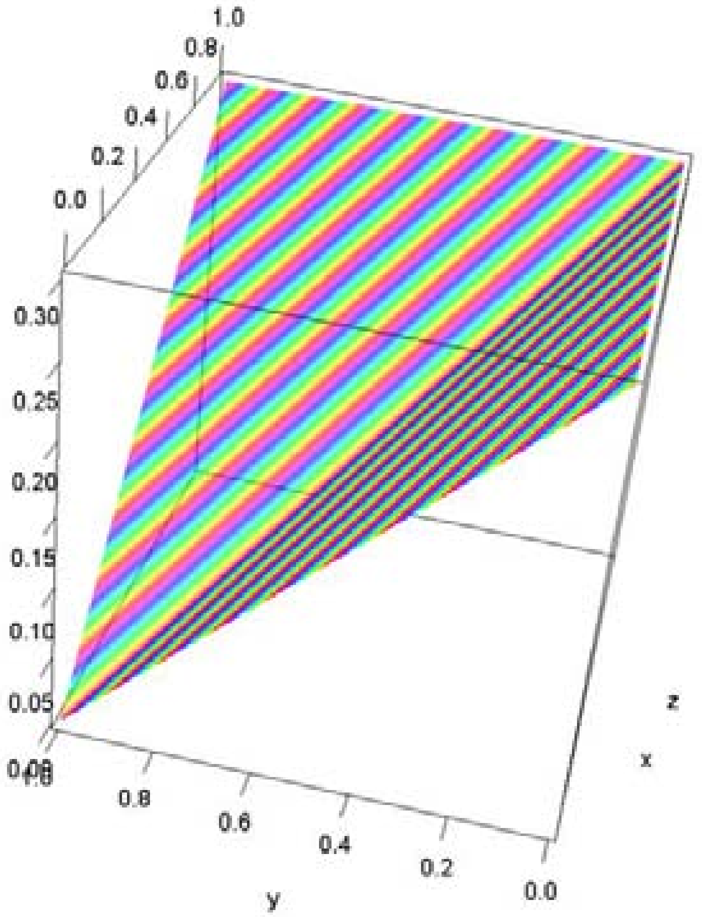


Figure 2: The contour plot of the probability of observing a (H, H)-type matching as the function of $(\boldsymbol{P^m}, \boldsymbol{P^w})$. The marginal distributions of types are $(f_H^m, f_H^w) = (0.3, 0.7)$. The x-axis represents $P_{HL|H}^m = P_{HL|H}^w$, the y-axis represents $P_{HL|L}^m = P_{HL|L}^w$, and the contour line represents the probability of observing a (H, H)-type matching.

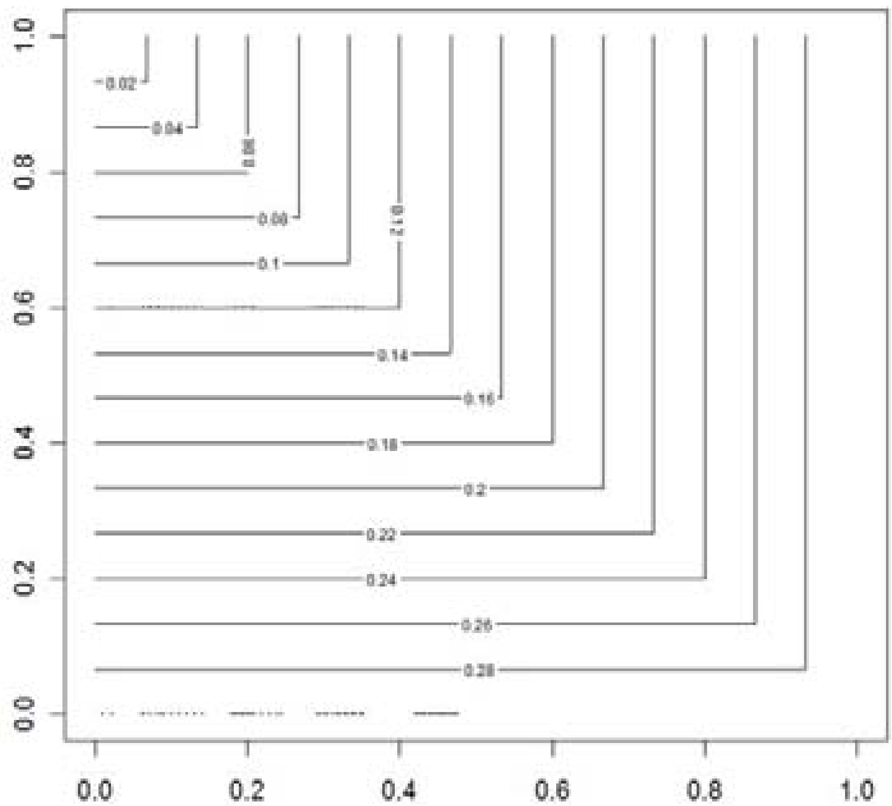


Figure 3: The 3D plot of the probability of observing a (H, H)-type matching as the function of $(\boldsymbol{P^m}, \boldsymbol{P^w})$. The marginal distributions of types are $(f_H^m, f_H^w) = (0.7, 0.7)$. The x-axis represents $P_{HL|H}^m = P_{HL|H}^w$, the y-axis represents $P_{HL|L}^m = P_{HL|L}^w$, and the z-axis represents the probability of observing a (H, H)-type matching.

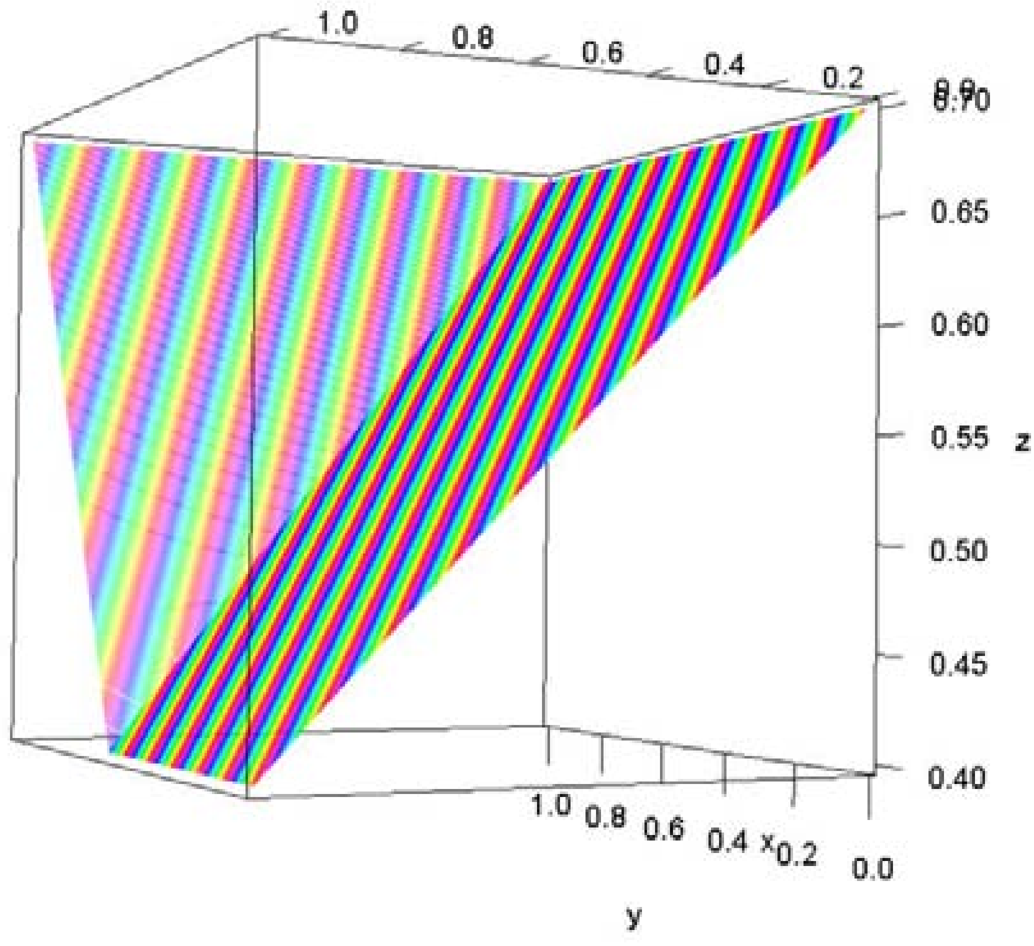


Figure 4: The contour plot of the probability of observing a (H, H)-type matching as the function of $(\boldsymbol{P^m}, \boldsymbol{P^w})$. The marginal distributions of types are $(f_H^m, f_H^w) = (0.7, 0.7)$. The x-axis represents $P^m_{HL|H} = P^w_{HL|H}$, the y-axis represents $P^m_{HL|L} = P^w_{HL|L}$, and the contour line represents the probability of observing a (H, H)-type matching.

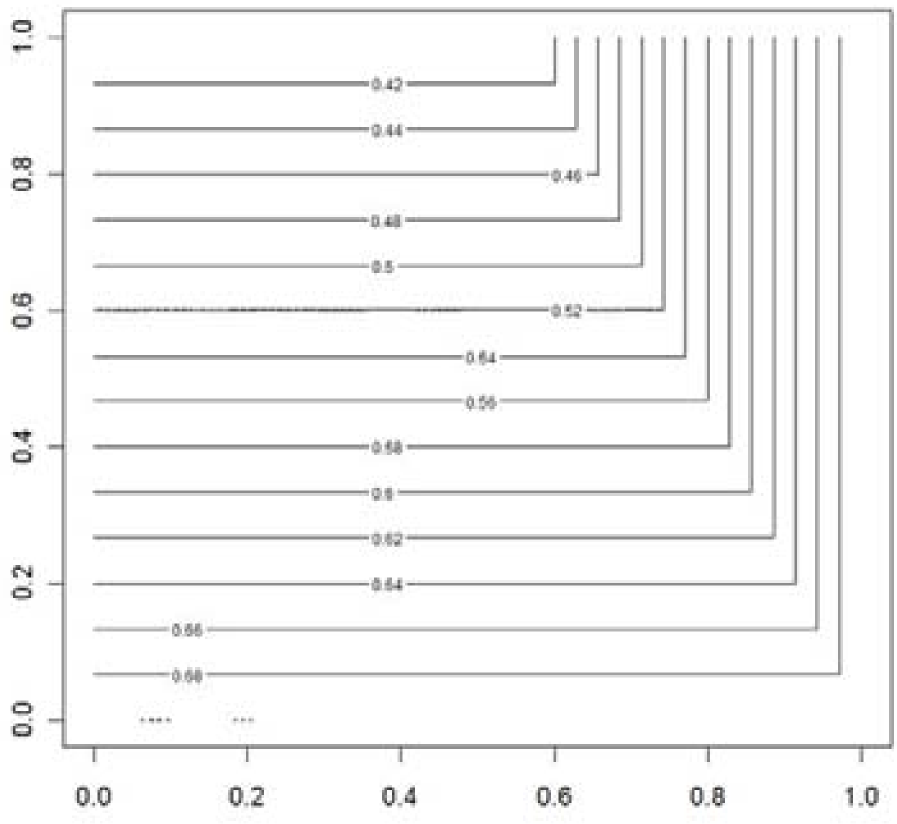
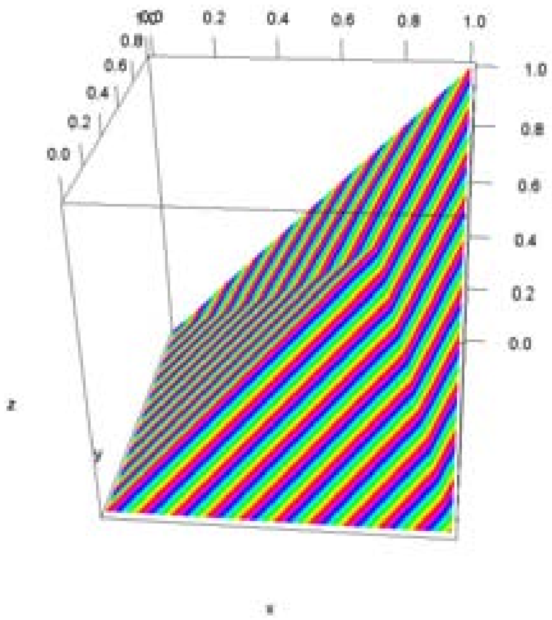


Figure 5: The 3D plot of the probability of observing a (H, H)-type matching as the function of the marginal distributions. The x-axis represents f_H^m, the y-axis represents f_H^w , and the z-axis represents the probability of observing a type (H, H) matching. The graph on the top panel corresponds to $\theta = (0.8, 0.8, 0.7, 0.6)$ and the graph on the bottom panel corresponds to $\theta^{'} = (0.7, 0.8, 0.7, 0.6)$.



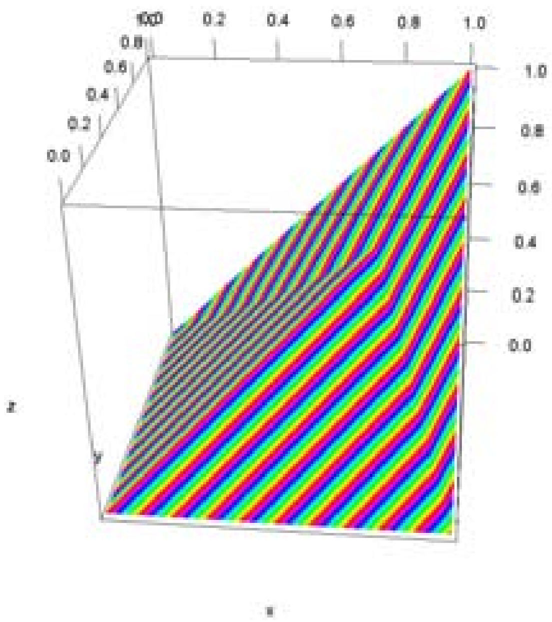


Figure 6: The 3D plot of $\mathbf{C}(f_H^m, f_H^w, \theta) - \mathcal{T}(f_H^m, f_H^w, \theta')$. The x-axis represents f_H^m and the y-axis represents f_H^w. $\theta = (0.8, 0.8, 0.7, 0.6)$ and $\theta' = (0.7, 0.8, 0.7, 0.6)$.

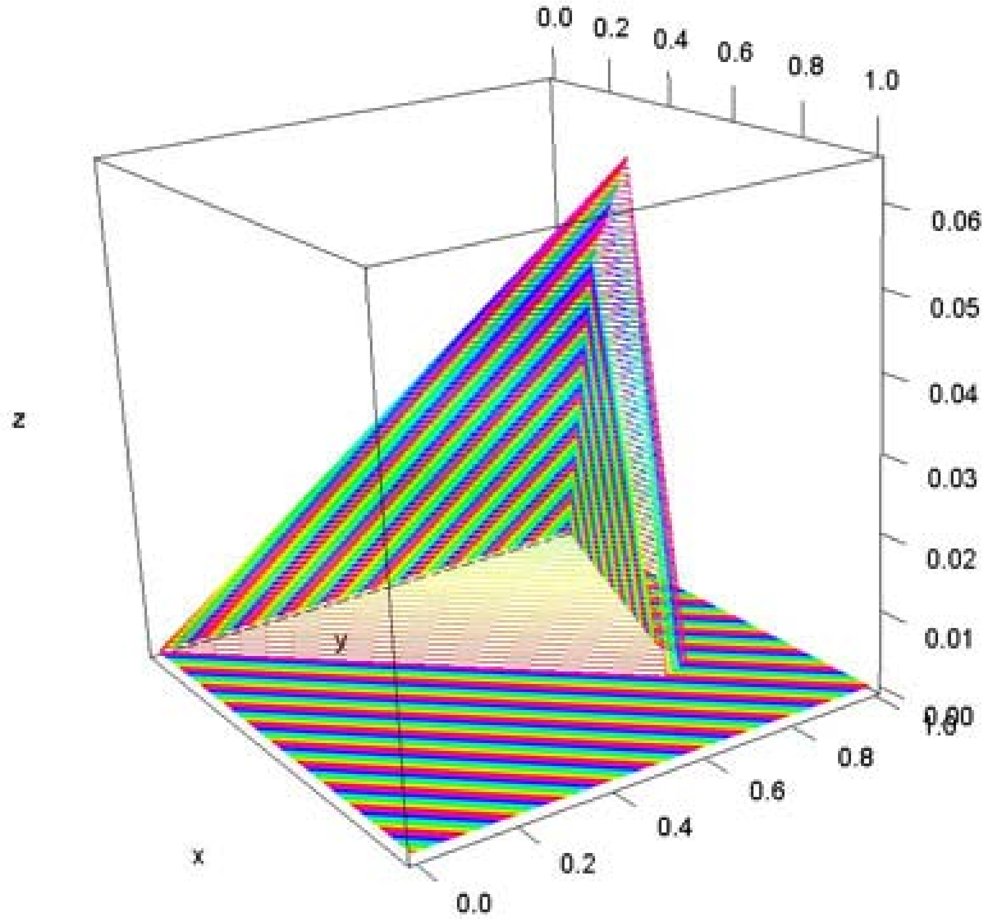


Figure 7: The Distribution of Education Level

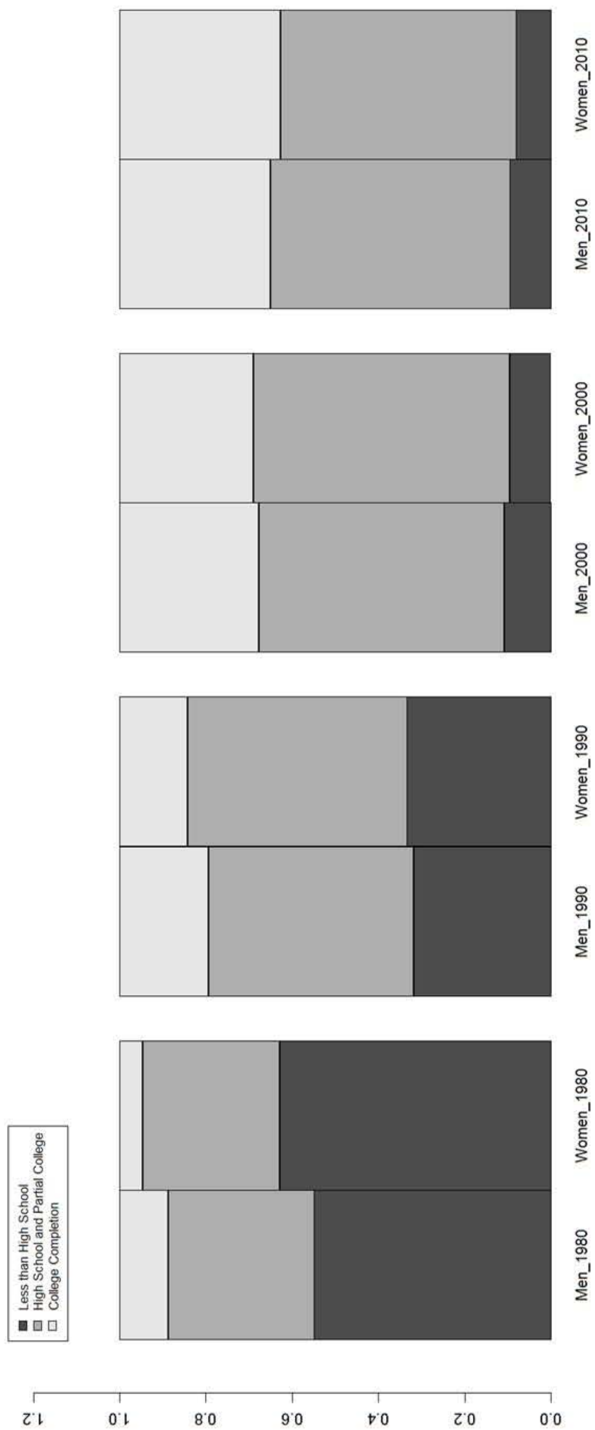


Figure 8: Average Education Cost, Assuming M-Optimal

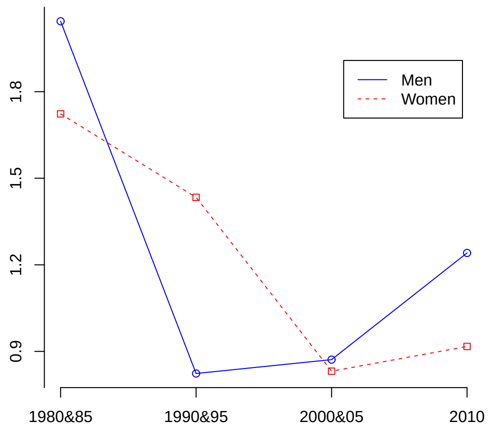


Figure 9: Average Education Cost, Assuming W-Optimal