Article

Monte Carlo Inference on Two-Sided Matching Models

Taehoon Kim [1], Jacob Schwartz [2], Kyungchul Song [3,*] and Yoon-Jae Whang [4]

[1] Department of Economics, Harvard University, Cambridge, MA 02138, USA; taehoonkim01@g.harvard.edu
[2] Department of Economics, University of Haifa, Haifa 3498838, Israel; jschwartz@econ.haifa.ac.il
[3] Vancouver School of Economics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[4] Department of Economics, Seoul National University, Seoul 08826, Korea; whang.yj@gmail.com
* Correspondence: kysong@mail.ubc.ca; Tel.: +1-604-822-2226

check for
updates

Abstract: This paper considers two-sided matching models with nontransferable utilities, with one side having homogeneous preferences over the other side. When one observes only one or several large matchings, despite the large number of agents involved, asymptotic inference is difficult because the observed matching involves the preferences of all the agents on both sides in a complex way, and creates a complicated form of cross-sectional dependence across observed matches. When we assume that the observed matching is a consequence of a stable matching mechanism with homogeneous preferences on one side, and the preferences are drawn from a parametric distribution conditional on observables, the large observed matching follows a parametric distribution. This paper shows in such a situation how the method of Monte Carlo inference can be a viable option. Being a finite sample inference method, it does not require independence or local dependence among the observations which are often used to obtain asymptotic validity. Results from a Monte Carlo simulation study are presented and discussed.

1. Introduction

Two-sided matching models have been widely used to study various interactions among people and firms. Examples are many, including medical residency match (Roth and Sotomayor (1990); Agarwal (2015), among many others), marriage/dating markets (Choo and Siow (2006); Hitsch et al. (2010); Chiappori et al. (2012)), loan markets (Chen and Song (2013)), venture capitals (Sørensen (2007)), merger markets in the mutual fund industry (Park (2013)), auction analysis (Fox and Bajari (2013)), and teacher assignments (Boyd et al. (2013)).

Large matching data pose challenges for econometric inference. Consider a matching between colleges and students. When a college and a group of students prefer to be matched more than their alternatives, this match limits the set of students available to other colleges and the set of colleges available to other students. This strategic interdependence potentially creates a nonstandard pattern of stochastic dependence among matches and makes asymptotic inference difficult, because the stochastic dependence is not in the form of weak spatial dependence or conditional independence much studied in the literature of asymptotic inference with cross-sectionally dependent data.

A popular approach to econometric modeling of matching markets is to model them as matching with transferable utility, where transfer of payoff between agents is allowed. (See Choo and Siow (2006); Galichon and Salanié (2012); Graham et al. (2014); Fox (2018); Fox and Bajari (2013).) However, there are many forms of matching markets such as marriage markets or medical residency matching, where

payoff transfer does not constitute a realistic feature. A recent body of literature develops empirical models of nontransferable utility. (See Menzel (2015) and Diamond and Agarwal (2017).)

This paper proposes a new approach of analyzing data from a large matching market. Here we consider a large, two-sided, many-to-one matching market with nontransferable utilities, where we allow agents to care about both observed and unobserved heterogeneities of the agents on the other side, but restrict one side of the market to have homogenous preferences.[1] The assumption of one-sided homogeneity of preferences in this paper is made mainly to ensure an explicit form of a unique stable matching mechanism that underlies the matching data. When the matching mechanism that is implemented in practice is known (such as in empirical examples of medical residents' matching with hospitals or students assigned to schools (e.g., Agarwal and Somaini (2018)), one can still apply the Monte Carlo inference approach of this paper without assuming one-sided homogeneity of preferences.

The main departure of this paper from the literature is that this paper develops a finite sample inference procedure for the payoff parameters in the matching market. The main idea is as follows. First, we note that the preference homogeneity on one side implies the existence of a unique stable matching which can be implemented as a form of a serial dictatorship mechanism. This mechanism has an explicit form where the student ranked first by the colleges is matched to his most preferred college and the student ranked second matched to his most preferred college among those colleges whose quota is not filled, etc. This characterization determines the exact distribution of the observed matching up to an unknown parameter, when the unobserved heterogeneities are from a parametric family of distributions. Thus, one can construct a test statistic and invert the test to perform finite sample inference on structural parameters.

The approach of Monte Carlo inference can be viewed as an extension of randomized tests of R. A. Fisher. Randomized tests achieve finite sample validity using a test statistic whose conditional distribution given data (i.e., the permutation distribution) is fully known. Similarly, the approach of Monte Carlo tests also focuses on a situation where the test statistic's distribution given the true parameter value is fully known. The Monte Carlo inference approach was developed to implement a permutation test (Dwass (1957)); Besag and Clifford (1989); Hope (1968), and introduced to econometrics and extended to various econometric models in work by Jean-Marie Dufour and his coauthors. (Dufour (2006). Also see Dufour and Khalaf (2003) for an overview of this approach in the context of econometric models.)

While this paper's approach provides a useful, alternative way to analyze matching data, there are limitations. The major limitation of this paper's approach is the assumption that we observe the entire set of players in the large matching game. This assumption is frequently used in many game-theoretic models, and hard to remove, because without this assumption, the payoff specification involves actions or characteristics of players that are not observed by the econometrician and one needs to assume a particular way the players are sampled in each game.[2] Nevertheless, assuming full observation of players in a game is restrictive in modeling the interactions among many agents. Second, as mentioned previously, this paper uses the one-sided homogeneity of preferences in this paper mainly to ensure an explicit form of a unique stable matching mechanism. It would be good to relax this requirement so that multiple stable matchings are potentially allowed in the model.

---

[1] This assumption of homogeneity in preferences of one side is certainly restrictive, and yet this asymmetry of preference heterogeneity between the two sides reflects various many-to-one matching markets in practice. For example, colleges mostly agree on who the best students are whereas many students face tradeoff between the distance from their homes to a college and the college's quality. This assumption of homogeneous preference on one side is not unprecedented in the literature either. See for example Agarwal (2015) who used this assumption in the analysis of medical residents' matching market.

[2] Canen et al. (2018) for an empirical model of linear interactions over a large network. Using a set of behavioral assumptions, they produce best responses that exhibit local dependence, and permit partial observation of the players by the econometrician for inference.

We performed Monte Carlo simulation studies using a simple many-to-one matching model between students and colleges. Due to the nature of finite sample inference, any deviation from the nominal size stems from the Monte Carlo simulation errors. With the number of Monte Carlo replications equal to 1000, the Monte Carlo inference exhibit reasonably good size properties. However, we have found that the power properties are uneven across different directions in which the parameter deviates from the true parameter. When we increase the number of the students, the power improves yet again unevenly across different directions of deviations.

The paper is organized as follows. The next section gives a general overview of the Monte Carlo inference approach. In Section 3, we introduce a two-sided many-to-one random matching model based on a college admission model in Roth and Sotomayor (1990). Section 4 presents an approach of Monte Carlo inference, explaining ways to construct test statistics and critical values. Section 5 gives results from a small-scale Monte Carlo simulation experiment and discusses them. In Section 6 we conclude. In Appendix A, we provide a simple algorithm that generates a matching based on a serial dictatorship mechanism.

2. Monte Carlo Inference: Overview

2.1. Monte Carlo Inference

In this section, we provide a general overview of the finite sample inference approach that we employ in this paper. Suppose that $\mathbf{Y}$ is an n-dimensional endogenous random vector and $\mathbf{X}$ is an exogenous random vector. Suppose further that the conditional distribution of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ follows a parametric family of distributions, say,

$$\mathcal{F}(\mathbf{x}) \equiv \{F_\theta(\cdot|\mathbf{x}) : \theta \in \Theta\}.$$

In our context, $\mathbf{Y}$ represents the matching outcomes between two sides of agents (e.g., students and colleges). Please note that such a parametric family assumption underlies maximum likelihood estimation, where the random map $\theta \to f_\theta(\mathbf{Y}|\mathbf{X})$ refers to the likelihood function, when $f_\theta(\cdot|\mathbf{x})$ denotes the conditional density function corresponding to $F_\theta(\cdot|\mathbf{x})$.

Let us first consider a finite sample inference on θ_0, where θ_0 denotes the true parameter such that $F_{\theta_0}(\cdot|\mathbf{x}) = F(\cdot|\mathbf{x})$, where $F(\cdot|\mathbf{x})$ denotes the conditional distribution function of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$. First, we construct a test statistic $T_n(\mathbf{Y}, \mathbf{X}, \theta)$. A standard way to construct a test statistic usually involves a sum of independent or locally dependent observations to facilitate asymptotic theory. However, in our case, it is hard to write the test statistic as a function of the sum of random variables for the inference procedure to exhibit finite sample validity. As we shall see, our matching data is such that each match between a student and a college involves all the other agents' payoff components nonlinearly.

A confidence set is generated in the following way. First, for each $\theta \in \Theta$, we draw by simulation

$$\mathbf{Y}_1(\theta), ..., \mathbf{Y}_R(\theta) \sim \text{i.i.d. } F_\theta(\cdot|\mathbf{X}).$$

We construct $t_1(\theta), ..., t_R(\theta)$, where $t_r(\theta) = T_n(\mathbf{Y}_r(\theta), \mathbf{X}, \theta)$, with $r = 1, ..., R$. Let $c_\alpha(\theta)$ be such that

$$c_\alpha(\theta) = \inf \left\{ c \in \mathbf{R} : \frac{1}{R} \sum_{r=1}^{R} 1\{t_r(\theta) \leq c\} \geq 1 - \alpha \right\}.$$

Then the finite sample confidence set is defined to be

$$C_\alpha = \{\theta \in \Theta : T_n(\mathbf{Y}, \mathbf{X}, \theta) \leq c_\alpha(\theta)\}.$$

By construction, the confidence set is valid in finite samples. This approach to inference is called the *Monte Carlo inference approach*. This approach is generally applicable in a set-up where observations are from a parametric family of distributions and random draws from these distributions can be obtained by simulation, i.e., a situation where one applies maximum likelihood estimation. In contrast to maximum likelihood estimation, Monte Carlo inference is valid in finite sample inference, and hence does not require assumptions that are used to ensure asymptotic validity of inference. Such asymptotic validity is obtained by assuming that the observations are independent or locally (or weakly) dependent along a certain known dependence ordering. Such local dependence is hard to verify in our context of a large matching market, because the matching outcomes involve all agents' idiosyncratic payoff components. The Monte Carlo inference approach offers a viable solution in this situation.

While the choice of a test statistic does not influence the finite sample validity of the inference, it affects the power properties. One way to construct a test statistic is to compare some features of the observed outcomes and those of predicted outcomes, where the predicted outcomes are generated by simulations. More specifically, suppose that $g_n(\mathbf{Y}, \mathbf{X})$ denotes a vector where each entry captures some aspect of the data $(\mathbf{Y}, \mathbf{X})$. Using the simulated draws $\mathbf{Y}_r(\theta), r = 1, ..., R$, we can construct

$$T_n(\mathbf{Y}, \mathbf{X}, \theta) = \frac{1}{R} \sum_{r=1}^{R} \delta(g_n(\mathbf{Y}, \mathbf{X}), g_n(\mathbf{Y}_r(\theta), \mathbf{X})), \tag{1}$$

where $\delta(g_n(\mathbf{Y}, \mathbf{X}), g_n(\mathbf{Y}_r(\theta), \mathbf{X}))$ is a scalar measure of discrepancy between $g_n(\mathbf{Y}, \mathbf{X})$ and $g_n(\mathbf{Y}_r(\theta), \mathbf{X})$. If this discrepancy gets larger fast as θ moves away from θ_0, θ will be less likely to be included in the confidence set.

The Monte Carlo confidence set is obtained by inverting the test statistic, and hence can be computationally intensive when the parameter dimension is large.

2.2. Subvector Inference

When the parameter of interest is a subvector of θ_0, we can construct a confidence set for the subvector by projecting the confidence set of θ_0 onto the subvector. More generally, suppose that our parameter of interest takes the following form:

$$\beta = \psi(\theta) \in B,$$

where ψ is a map known to the econometrician, taking values in a set B. For example, we may have $\beta = \theta_j$, where θ_j is the j-th element of θ. The projection approach suggests that we may construct a confidence set consisting of β's such that there exists θ in the confidence set of θ_0 such that $\beta = \psi(\theta)$. This way of doing subvector inference is called the projection approach.

An alternative way is a profiling approach. (See Barndorff-Nielsen (1983), Romano and Shaikh (2008), and Bugni et al. (2017)). We fix θ and let $\mathbf{Y}_1(\theta), ..., \mathbf{Y}_R(\theta)$ be given as before through simulations. Let $\gamma_\alpha(\beta)$ be such that

$$\gamma_\alpha(\beta) = \inf \left\{ c \in \mathbf{R} : \frac{1}{R} \sum_{r=1}^{R} 1 \left\{ \inf_{\theta \in \Theta: \psi(\theta) = \beta} \sup_{\tilde{\theta} \in \Theta: \psi(\tilde{\theta}) = \beta} T_n(\mathbf{Y}_r(\tilde{\theta}), \mathbf{X}; \theta) \leq c \right\} \geq 1 - \alpha \right\}.$$

Then the finite sample confidence set is defined to be

$$\tilde{C}_{\alpha,R}^{\text{PF}} = \left\{ \beta \in B : \inf_{\theta \in \Theta: \psi(\theta) = \beta} T_n(\mathbf{Y}, \mathbf{X}; \theta) \leq \gamma_\alpha(\beta) \right\}.$$

It is not hard to see that this confidence set for the subvector β is valid in finite samples. This Monte Carlo inference version of subvector inference has some differences from the standard profiling approach. First, note that the distribution of $\mathbf{Y}$ is equal to that of $\mathbf{Y}_r(\theta_0)$, but not necessarily equal to the distribution of $\mathbf{Y}_r(\theta)$ for all θ such that $\beta_0 = \psi(\theta)$. To cover this discrepancy in finite samples, we take the supremum over $\tilde{\theta} \in \Theta$ in the computation of critical values to ensure finite sample validity. Second, one of the main challenges in the subvector inference literature is that it is hard to approximate the distribution of $\inf_{\theta \in \Theta: \psi(\theta) = \beta} T_n(\mathbf{Y}, \mathbf{X}; \theta)$. This approximation often requires a careful choice of tuning parameters as well. However, being a finite sample inference method by nature, the Monte Carlo-based subvector inference does not suffer from this difficulty. Third, the computational cost of doing subvector inference through profiling may not be reduced substantially relative to the projection method, especially when the evaluation of $T_n(\mathbf{Y}_r(\tilde{\theta}), \mathbf{X}; \theta)$ is computationally costly.[3]

3. A Two-Sided Random Matching Model

3.1. A College Admissions Model

We begin with a standard college admissions model as in Chapter 5 of Roth and Sotomayor (1990) using different notation that is suitable for our purpose. Then we introduce a random preference profile, and make explicit the distribution of the observed matching.

Suppose that we have a set of students indexed by $N_s = \{1, ..., n_s\}$ and that of colleges indexed by $N_c = \{1, ..., n_c\}$. In many situations, the colleges are capacity-constrained for various reasons. For each college j, let q_j be a positive integer that represents the quota of college j. To accommodate the possibility of students or colleges unmatched, define $N'_s = N_s \cup \{0\}$ and $N'_c = N_c \cup \{0\}$, so that unmatched student (or college) is viewed as being matched to 0. (When we need to view N'_s and N'_c as an ordered set, according to the ordering of natural numbers, we take 0 to be the last element of N'_s and N'_c.) A *(many-to-one) matching under (the capacity constraint)* $q = (q_j)_{j \in N_c}$ is defined as a (point-valued) map $\mu : N'_s \to N'_c$ such that $|\mu^{-1}(j) \cap N_s| \leq q_j$ for each $j \in N_c$, i.e., the number of the students assigned to each college does not exceed its capacity.

The matching result depends on a preference profile of agents. In this paper, we allow for only strict preferences so that each student is never indifferent between choices from N'_c and the same with each college. It is convenient if we represent each preference ordering with a permutation of the agent indices. Let Π_c and Π_s be the collections of permutations over N'_s and N'_c respectively, so that each permutation is associated with a preference ordering. (A word of caution with subscripts: Π_c represents the set of preferences of colleges over students.) For example, suppose that $N'_c = (0, 1, 2, 3, 4)$. If a student has *preference ordering* $\pi \in \Pi_s$ over colleges N'_c such that $\pi = (3, 2, 0, 4, 1)$ (or equivalently, $\pi(1) = 3$, $\pi(2) = 2$, $\pi(3) = 0$, etc., this means the student ranks college 3 as highest, and college 1 as lowest (even lower than being unmatched.)

More formally, for given $\pi \in \Pi_s$, we write $i_1 \succ_\pi i_2$ if $\pi^{-1}(i_1) < \pi^{-1}(i_2)$, i.e., i_1 is ranked higher than i_2 by preference π. Given preference $\pi \in \Pi_s$ of student i over colleges, we say that college j is *acceptable* by student i if $j \succ_\pi 0$. We make similar definitions for college preferences. For two sets S_1 and S_2 of students and a preference ordering π of a college, we write $S_1 \succ_\pi S_2$, if for all $i_1 \in S_1$ and $i_2 \in S_2$, $i_1 \succ_\pi i_2$.

The collection of preference ordering profiles, $\boldsymbol{\pi} = (\boldsymbol{\pi}_s, \boldsymbol{\pi}_c)$, is given by

$$\mathbf{\Pi} = \Pi_s^{n_s} \times \Pi_c^{n_c},$$

---

[3]　Schwartz (2018) used the Monte Carlo subvector inference approach following this paper's proposal. However, his setting permits using standard inference on part of the parameter vector as a first step, applying Monte Carlo inference for the remaining parameters. This two-step approach no longer ensures finite sample validity. Nevertheless, it sharpens the inference results and reduces the computational costs.

where $\pi_s \in \Pi_s^{n_s}$ and $\pi_c \in \Pi_c^{n_c}$. Given a quota vector q, we call any map $\mu : N_s' \times \Pi \to N_c'$ a *matching mechanism under q*, if for each $\pi \in \Pi$, $\mu(\cdot; \pi)$ is a matching under q.[4]

In modeling a predicted outcome of matching in economics, it is standard in the literature to focus on (pairwise) stable matchings (Roth and Sotomayor (1990)).[5]

Definition 1.
(i) A matching $\mu : N_s' \to N_c'$ is stable with respect to $\pi = (\pi_s, \pi_c) \in \Pi$, if the following two conditions are satisfied.

　　(a) There is no $i \in N_s$ such that $0 \succ_{\pi_i} \mu(i)$ and no $j \in N_c$ such that $0 \succ_{\pi_j} i'$ for some $i' \in \mu^{-1}(j)$.

　　(b) There is no pair $(i, j) \in N_s \times N_c$ such that both $j \succ_{\pi_i} \mu(i)$ and $i \succ_{\pi_j} i'$ for some $i' \in \mu^{-1}(j)$.

(ii) A matching mechanism $\mu : N_s' \times \Pi \to N_c'$ is stable if $\mu(\cdot; \pi)$ is stable with respect to each $\pi \in \Pi$.

Definition 1 says that a matching is stable with respect to π, if (a) each student matched with a college prefers to be matched with the college than to remain unmatched, and each college matched with a student prefers to be matched with the student than to remain unmatched, and (b) there is no unmatched pair of a student and a college both of whom prefer to be matched with each other than with their current matches.

3.2. Random Preferences

3.2.1. Students' Heterogeneous Preferences

Let us introduce random preferences for students with a view to an econometric modeling. It is convenient to introduce notation that turns a vector of real numbers into a permutation according to the ordering of the numbers.

Definition 2. *Given $a = (a_i)_{i \in N_c'} \in \mathbf{R}^{n_c+1}$, let $p_s(a) \in \Pi_s$ be such that $p_s^{-1}(a)(i)$ denotes the rank of a_i. For example, suppose that $a = (-0.9, 0.2, 0.7, -0.2)$. Then $p_s(a) = (2, 1, 3, 0)$. Hence the third entry of a (we start with zero so 2 represents the third entry) is ranked first. We define similarly $p_c \in \Pi_c$.*

The students' preferences are drawn in the following way. First, each student i is given a heterogeneous single index over colleges

$$v_s(i, j; \varepsilon_{i,j}, \beta_0) = f_s(x_{s,i}, x_{c,j}, \varepsilon_{i,j}; \beta_0),$$

which represents student i's "score" of college j, where $x_{s,i}$ denotes the characteristics of student i and $x_{c,j}$ those colleges j that are observed by the econometrician, $\varepsilon_{i,j}$, unobserved taste component associated with the match between student i and college j, and $f_s(\cdot, \cdot, \cdot; \beta_0)$ denotes a certain parametric function known up to a true parameter vector β_0.

For example, one may consider the following single-index specification with an additive error:[6]

$$f_s(x_{s,i}, x_{c,j}, \varepsilon_{i,j}; \beta_0) = x_{c,j}' \beta_{0,1} + x_{s,i} x_{c,j}' \beta_{0,2} + \varepsilon_{i,j}.$$

---

[4]　By the definition of a matching mechanism as a map on $N_s' \times \Pi$, it is *anonymous* in the sense that the matching mechanism remains invariant to the relabeling of the agents' indices.

[5]　A proper development will require defining preferences over *sets of students* by colleges, and defining stability of a matching in terms of these preferences. When the preferences are so-called responsive, the group stability is equivalent to pairwise stability. As we make use of pairwise stability for econometric inference, we refer the reader to Chapter 5 of Roth and Sotomayor (1990) for further details.

[6]　Adding an additive term whose covariate depends only on $x_{s,i}$ instead of the interaction term $x_{c,j}$ is superfluous for "identification", because variations in $x_{s,i}$ do not change the ranking of the colleges by the student i.

However, this paper's procedure does not require this particular structure; all that is required is that the function f_s is assumed to be known to the econometrician.

To characterize the preference for being unmatched (i.e., an outside option), we also define

$$v_s(i, 0; \varepsilon_{i,0}, \bar{\beta}_0) = f_{s,0}(x_{s,i}, \varepsilon_{i,0}; \bar{\beta}_0),$$

for some parameter $\bar{\beta}_0$. Throughout this paper, we often write simply $v_s(i, j; \varepsilon_{i,j})$ suppressing the notation for β_0. Also define $v_s(i; \varepsilon_i) = (v_s(i, j; \varepsilon_{i,j}))_{j \in N_c'}$, i.e., the vector of the scores given to the colleges (by student i).

Let us assume that the preference profile for the students is given as follows:

$$\boldsymbol{\pi}_s(\varepsilon) \equiv (\pi_{s,i}(\varepsilon_i))_{i=1}^{n_s}, \text{ where } \pi_{s,i}(\varepsilon_i) \equiv p_s(v_s(i; \varepsilon_i)), \text{ for all } i \in N_s.$$

Therefore, for each $i \in N_s$, the random preference $\pi_{s,i}(\varepsilon_i)$ places a college j as first when $v_s(i, j; \varepsilon_{i,j})$ is largest.

3.2.2. Colleges' Homogeneous Preferences

Throughout this paper, we assume that the colleges have the same preference over the students. Again, let $x_{s,i}$ be a vector of student i's observable characteristics. Let

$$v_c(i; \eta_i, \gamma_0) = g_c(x_{s,i}, \eta_i; \gamma_0)$$

be student i's single index, where η_i represents student i's unobserved quality and g_c is a function known up to the parameter γ_0. As for an outside option, define

$$v_c(0; \eta_0, \bar{\gamma}_0) = g_{c,0}(\eta_0; \bar{\gamma}_0).$$

From here on, we often write $v_c(i; \eta_i)$ simply, suppressing the notation for γ_0. The colleges' preference over students is not solely determined by the students' observable characteristics. The common random preference of each college j for each student i is given as follows:

$$\pi_c(\eta) \equiv p_c(v_c(\eta)),$$

where $v_c(\eta) = (v_c(i; \eta_i))_{i \in N_s'}$. Thus, each college j ranks a student i as highest if $v_s(i; \eta_i)$ is highest.

3.3. The Joint Distribution of a Large Observed Matching

In this section, we obtain an explicit expression of the distribution of the observed matching when the matching arises from a stable matching mechanism. Throughout this paper, we regard $x_{s,i}$ and $x_{c,j}$ as non-stochastic, which means that all other unobserved random components such as η and ε are independently drawn from these observed characteristics. Recall that the preference profile of students and colleges are given by $\boldsymbol{\pi}(\varepsilon, \eta) = (\pi_c(\eta), \boldsymbol{\pi}_s(\varepsilon))$, where we simply write the preference profile of colleges as $\pi_c(\eta)$ because it is the same across the colleges.

Let $\mathbf{Y}$ denote the observed matching which is generated by a stable matching mechanism, say, $\mu(\cdot; \boldsymbol{\pi}(\varepsilon, \eta))$.[7] In other words,

$$\mathbf{Y}(\cdot) = \mu(\cdot; \boldsymbol{\pi}(\varepsilon, \eta)).$$

This is a *reduced form* for the observed matching. The randomness of the observed matching comes from the randomness of the students' and colleges' preferences (i.e., η and ε).

---

[7]　Our choice of notation for matching as $\mathbf{Y}$ is to emphasize that matching is an endogenous outcome.

We make the following assumptions regarding the random preferences for students and colleges.

Assumption 1. $(v_c(\eta), (v_s(i; \varepsilon_i))_{i \in N_s})$ *is a continuous random vector.*

The continuity assumption is made to generate strict preferences. Later we assume that the distributions of η_i's and $\varepsilon_{i,j}$'s are independently drawn from certain parametric families of distributions.

Let us define

$$N_s(\eta) = \{i \in N_s : v_c(i; \eta_i, \gamma_0) > v_c(0; \eta_0, \overline{\gamma}_0)\},$$

which is the set of students that colleges prefer to match than to stay unmatched with any student. The stability of matching μ requires that $\mu(i) = 0$ for all $i \notin N_s(\eta)$ that is, students not preferred by any of the colleges (to the alternative of staying unmatched) are not matched with any colleges under μ. If $N_s(\eta) = \varnothing$, there is no college-student pair that is matched under a stable matching μ. Thus, it suffices to determine the match between colleges with students in $N_s(\eta)$ when $N_s(\eta)$ is not empty. Let us enumerate the student indexes in $N_s(\eta)$ as $\{S(1), ..., S(n_s')\}$ with $n_{s'} = |N_s(\eta)|$, so that $S : \{1, ..., n_s'\} \to N_s$ is a (random) map depending on η.

As we shall see now, the stability of matching yields a useful distributional characterization of the observed matching. For each $S(i) \in N_s(\eta)$, and for any set $A \subset N_c'$, let

$$\rho_i(A; \boldsymbol{\pi_s}) = \min_{j \in A} \pi_{s,S(i)}^{-1}(j).$$

In other words, $\rho_i(A; \boldsymbol{\pi_s})$ is the ranking of a college that is most preferred among colleges j in A by student $S(i)$. Hence

$$\pi_{s,S(i)}\left(\rho_i(A; \boldsymbol{\pi_s})\right)$$

denotes the college among A that is most preferred by student $S(i)$.

Suppose that the colleges' homogeneous preference is given by $\pi_c(\eta)(1) = S(1)$, $\pi_c(\eta)(2) = S(2),...,\pi_c(\eta)(n_s') = S(n_s')$. In other words, the colleges prefer student $S(i)$ to $S(j)$ in $N_s(\eta)$ if and only if $i < j$. Thus, let us simply write $\pi_c(\eta)$ as $N_s(\eta)$. In the stable matching with this $\pi_c(\eta)$, student $S(1)$ is ranked highest by all the colleges and chooses his most preferred college first. Then student $S(2)$ chooses his most preferred college among the colleges whose quota is not filled yet, etc.[8] To formalize this matching mechanism, let us define for each student $S(i) \in N_s(\eta)$,

$$\tau\left([i]; \boldsymbol{\pi_s}\right) \equiv \pi_{s,S(i)}\left(\rho_i(N_{[i-1]}(\boldsymbol{\pi_s}); \boldsymbol{\pi_s})\right),$$

where $N_{[0]}(\boldsymbol{\pi_s}) = N_c'$, (setting $q_j = \infty$ if $j \in N_c'$ is 0),[9]

$$N_{[i-1]}(\boldsymbol{\pi_s}) \equiv \{j \in N_c' : \tilde{q}_j([i-1]; \boldsymbol{\pi_s}) < q_j\},$$

and

$$\tilde{q}_j([i-1]; \boldsymbol{\pi_s}) = |\{i' \in N_s : \tau([i'-1]; \boldsymbol{\pi_s}) = j, \ S(i') \leq S(i)\}|. \tag{2}$$

---

[8] In the literature of mechanism design, this mechanism is called a serial dictatorship mechanism. (See Satterthwaite and Sonnenschein (1981)).

[9] Using the notation $[i]$ instead of i is meant as a reminder that the quantity depends on the "history" $[i] = (1, 2, ..., i-1, i)$, rather the current agent index i.

The college $\tau([1]; \boldsymbol{\pi}_s)$ is one that student $S(1)$ prefers most among all the colleges in N'_c. Then $\tau([2]; \boldsymbol{\pi}_s)$ is a college that student $S(2)$ prefers most among all the colleges whose quota is not yet filled once student $S(1)$ is assigned to college $\tau([1]; \boldsymbol{\pi}_s)$. Now $\tau([3]; \boldsymbol{\pi}_s)$ is a college that student $S(3)$ prefers most among all the colleges whose quota is not yet filled once students $S(1)$ and $S(2)$ are assigned to colleges $\tau([1]; \boldsymbol{\pi}_s)$ and $\tau([2]; \boldsymbol{\pi}_s)$ respectively.

Please note that the assignment of each student $S(i)$ to a college $\tau([i]; \boldsymbol{\pi}_s)$ is fully known up to the students' preference profile $\boldsymbol{\pi}_s$. Thus, the matching of each student $S(i)$ to a college $\tau([i]; \boldsymbol{\pi}_s)$ is a unique stable matching that is explicitly represented as a function of random preferences. In other words, for each $S(i) \in N_s(\eta)$,

$$\mu(S(i); \boldsymbol{\pi}_s(\varepsilon), N_s(\eta)) = \tau([i]; \boldsymbol{\pi}_s) \tag{3}$$

Please note that $\tau([i]; \boldsymbol{\pi}_s)$ depends on the history of choices of students $S(1), S(2), ..., S(i)$, not merely that of the last student $S(i)$.

It is straightforward to extend this result to a general case where $\pi_c(\eta)(i) \neq S(i)$ for some $i = 1, ..., n'_s$. We begin with student $\pi_c^{-1}(\eta)(1)$ (i.e., ranked highest by the colleges according to $\pi_c(\eta)$) and let him choose his best choice among all the colleges. Then we move onto student $\pi_c^{-1}(\eta)(2)$ (ranked second highest by the colleges).

To see its major implication for the distribution of the observed matching, note that

$$\begin{aligned} \mathbf{Y}(S(i)) &= \mu\left(S(i); \boldsymbol{\pi}_s(\varepsilon), \pi_c(\eta)\right) \\ &= \mu\left((S \circ \pi_c^{-1}(\eta) \circ S)(i); \boldsymbol{\pi}_s(\varepsilon, \eta), N_s(\eta)\right), \end{aligned} \tag{4}$$

where $\boldsymbol{\pi}_s(\varepsilon, \eta)$ is a profile of students' preferences over colleges with the $S(i)$-th student's preference over the colleges equal to

$$\pi_{s,(S \circ \pi_c^{-1}(\eta) \circ S)(i)}(\varepsilon),$$

for each $S(i) \in N_s(\eta)$. The second equality in (4) follows because the matching mechanism is anonymous. In other words, college $\mathbf{Y}(S(i))$ matched by student $S(i)$ should be the same college matched by the same student after relabeling the student $S(i)$ as $S(\pi_c^{-1}(\eta)(S(i)))$. (Recall that $\pi_c^{-1}(\eta)(S(i))$ represents the ranking of student $S(i)$.) After this relabeling, the colleges' preference ordering over the students becomes: $S(i) \succ_{\pi_c} S(j)$ if and only if $i < j$. Then from (3), we obtain the following result.

Theorem 1. *Suppose that the matching mechanism μ is stable, and that Assumption 2.1 holds. Then for each $S(i) \in N_s(\eta)$,*

$$\mathbf{Y}(S(i)) = \tau\left([(\pi_c^{-1}(\eta) \circ S)(i)]; \boldsymbol{\pi}_s(\varepsilon, \eta)\right),$$

where $[(\pi_c^{-1}(\eta) \circ S)(i)] = (1, 2, ..., (\pi_c^{-1}(\eta) \circ S)(i))$.

The expression in the theorem provides an explicit reduced form for the matching $\mathbf{Y}$. It shows how the randomness of the observed matching $\mathbf{Y}$ depends on the random preferences. It shows that each single match $\mathbf{Y}(S(i))$ of student $S(i)$ is a complex function of the preferences of *all the students and colleges*. While incorporating this interdependence is crucial to properly take into account the inherent endogeneity of observed matching, it hampers the use of standard asymptotic theory. Thus, this paper pursues an approach of finite sample inference. Indeed, Theorem 1 shows that once we parametrize the distribution of $\varepsilon_{i,j}$ and η_i, we can obtain the full joint distribution of the observed matching up to a parametrization. Comparing the sorting pattern of observed characteristics implied

by this predicted distribution and the observed sorting pattern, we seek to 'perform inference on the structural parameters in finite samples.

4. Monte Carlo Inference

Let us formalize the data generating process. First, the nature draws $\varepsilon_{i,j}$'s, i.i.d., $i = 1, ..., n_s$, and $j = 1, ..., n_c$, from a parametric distribution, say, F_θ, and η_i's, i.i.d., from a parametric distribution, say, G_θ. Please note that the inference procedure also allows the error term to be involved in the single indices nonseparably, as in the case of random coefficient models.

4.1. Test Statistics, Critical Values and Confidence Sets

We explain a general method of constructing a test statistic and a confidence set. We define $\theta = (\beta, \bar{\beta}, \gamma, \bar{\gamma})$. First, for $b = 1, ..., B$, we let $\eta_{i,b}^*$ be the b-th simulated draw from G_θ and $\varepsilon_{i,j,b}^*$ from F_θ. We construct simulated matchings: for $N_s(\eta_b^*)$ with $\eta_b^* = (\eta_{i,b}^*)_{i \in N_s}$, and $S(i) \in N_s(\eta_b^*)$,

$$\mathbf{Y}_b^*(S(i); \theta) = \tau \left([\pi_c(\eta_b^*)^{-1}(S(i))]; \pi_s(\varepsilon_b^*, \eta_b^*) \right),$$

where $\varepsilon_b^* = (\varepsilon_{i,j,b}^*)_{(i,j) \in N_s \times N_c'}$. We also draw for $r = 1, ..., R$, $\eta_{i,r}$'s and $\varepsilon_{i,j,r}$'s i.i.d. from G_θ and from F_θ, respectively, independently of $\eta_{i,b}^*$'s and $\varepsilon_{i,j,b}^*$. We construct simulated matchings:

$$\mathbf{Y}_r(S(i); \theta) = \tau \left([\pi_c(\eta_r)^{-1}(S(i))]; \pi_s(\varepsilon_r, \eta_r) \right),$$

where $\eta_r = (\eta_{i,r})_{i \in N_s}$ and $\varepsilon_r = (\varepsilon_{i,j,r})_{(i,j) \in N_s \times N_c'}$. Hence conditional on $\mathbf{X}$, $\mathbf{Y}_b^*(s_i; \theta)$'s and $\mathbf{Y}_r(i; \theta)$'s are independent. We set $\mathbf{Y}_b^*(i; \theta) = 0$ for all $i \notin N_s(\eta^*)$ and $\mathbf{Y}_r(s_i; \theta) = 0$ for all $i \notin N_s(\eta_r)$. An algorithm for computing the matching this way is provided in the appendix. Let us define

$$\begin{aligned} \mathbf{Y}_r(\theta) &\equiv \{\mathbf{Y}_r(i; \theta) : i \in N_s\}, \text{ and} \\ \tilde{\mathbf{Y}}_B^*(\theta) &\equiv \{\mathbf{Y}_b^*(i; \theta) : i \in N_s, b = 1, ..., B\}. \end{aligned}$$

Then we can construct a test statistic as a function of the observed matching $\mathbf{Y}$, the simulated matchings $\tilde{\mathbf{Y}}_B^*(\theta)$, and the observed characteristics , $\mathbf{X}$, say,

$$T(\theta) = f_n(\mathbf{Y}, \tilde{\mathbf{Y}}_B^*(\theta), \mathbf{X}),$$

for some function f_n. (Some examples of the test statistics are given below.)

As for the critical values, we simulate the test statistic using simulated matching $\mathbf{Y}_r(\theta)$ in place of observed matching $\mathbf{Y}$:

$$T_r(\theta) = f_n(\mathbf{Y}_r(\theta), \tilde{\mathbf{Y}}_B^*(\theta), \mathbf{X}), r = 1, ..., R,$$

where the simulations of $\mathbf{Y}_r(\theta)$ and take $c_\alpha(\theta)$ to be the $1 - \alpha$ percentile of the empirical distribution of the simulated test statistics $T_r(\theta), r = 1, ..., R$. Then the confidence set is given by

$$C_\alpha = \{\theta \in \Theta : T(\theta) \leq c_\alpha(\theta)\}.$$

The finite sample validity of the confidence set (up to a simulation error) immediately follows from Theorem 1.

One may choose R and B differently. Choosing a large R will reduce the Monte Carlo error in the coverage probabilities, and choosing a large B will improve the power properties (i.e., shrink the size of the confidence interval), not affecting the finite sample validity of the inference. This power improvement will be attenuated after a large enough B. When the computational cost of matching

$\mathbf{Y}_b^*(i; \theta)$ is substantial, it is practical to use only minimal B as long as it ensures decent power properties of the inference.

4.2. Constructing Test Statistics

In this section, we discuss ways to construct a test statistic which is based on comparing features of observed matching outcomes and those of predicted ones. More specifically, we define

$$f_n(\mathbf{Y}, \tilde{\mathbf{Y}}_B^*(\theta), \mathbf{X}) = m_n(\mathbf{Y}, \tilde{\mathbf{Y}}_B^*(\theta), \mathbf{X}), \tag{5}$$

where

$$m_n(\mathbf{Y}, \tilde{\mathbf{Y}}_B^*(\theta), \mathbf{X}) = \frac{1}{B} \sum_{b=1}^{B} \max_{\mathbf{x}_s, \mathbf{x}_c} \left| \hat{P}(\mathbf{x}_s, \mathbf{x}_c; \mathbf{Y}, \mathbf{X}) - \hat{P}(\mathbf{x}_1, \mathbf{x}_2; \mathbf{Y}_b^*(\theta), \mathbf{X}) \right|,$$

and

$$\hat{P}(\mathbf{x}_s, \mathbf{x}_c; \mathbf{Y}, \mathbf{X}) = \frac{1}{n_s} \sum_{i \in N_s} 1\{\mathbf{X}_{i,s} = \mathbf{x}_s, \mathbf{X}_{\mathbf{Y}(i),c} = \mathbf{x}_c\}. \tag{6}$$

Here $\hat{P}(\mathbf{x}_s, \mathbf{x}_c; \mathbf{Y}, \mathbf{X})$ measures the proportion of students with characteristic $\mathbf{x}_s$ which are matched with colleges with characteristic $\mathbf{x}_c$ through matching $\mathbf{Y}$. (See Diamond and Agarwal (2017) and Schwartz (2018) for the use of similar test statistics.) One may use some other features of the matchings. For example, following Diamond and Agarwal (2017), we may consider

$$\tilde{m}_n(\mathbf{Y}, \tilde{\mathbf{Y}}_B^*(\theta), \mathbf{X}) = \frac{1}{B} \sum_{b=1}^{B} \left| \hat{\Delta}(\mathbf{Y}, \mathbf{X}) - \hat{\Delta}(\mathbf{Y}_b^*(\theta), \mathbf{X}) \right|,$$

where

$$\hat{\Delta}(\mathbf{Y}, \mathbf{X}) = \frac{1}{n_s} \sum_{i \in N_s} \left(\mathbf{X}_{i,s} - \overline{\mathbf{X}}_{\mathbf{Y}^{-1}(\mathbf{Y}(i)),s} \right)^2$$

and

$$\overline{\mathbf{X}}_{\mathbf{Y}^{-1}(\mathbf{Y}(i)),s} = \frac{1}{|\mathbf{Y}^{-1}(\mathbf{Y}(i))|} \sum_{\ell \in \mathbf{Y}^{-1}(\mathbf{Y}(i))} \mathbf{X}_{\ell,s}.$$

The quantity $\hat{\Delta}(\mathbf{Y}, \mathbf{X})$ measures the dispersion of the observed characteristics of the students who are matched with the same college as student i. Then, we may take

$$f_n(\mathbf{Y}, \tilde{\mathbf{Y}}_B^*(\theta), \mathbf{X}) = m_n(\mathbf{Y}, \tilde{\mathbf{Y}}_B^*(\theta), \mathbf{X}) + \tilde{m}_n(\mathbf{Y}, \tilde{\mathbf{Y}}_B^*(\theta), \mathbf{X}). \tag{7}$$

This combination of two criteria attempts to capture potential deviation of θ from θ_0 on two fronts: comparison between observed characteristics of matches and predicted characteristics and comparison between observed within-match characteristic dispersion and its predicted version. It is important to note that it is not *a priori* ensured that using the two criteria instead of one improves the power properties of the inference. However, Diamond and Agarwal (2017) demonstrated through simulations that using a combination of criteria such as in (7) can sharpen the accuracy of inference.

5. Monte Carlo Simulation Studies

In this section, we investigate size and power properties of our simulation-based inference procedure. We wish to infer the preference parameters of agents, $\theta = (\theta_c, \theta_s)$. For simplicity,

our simulation design is such that every student and every college is acceptable to each other's side. Recall that the value of the college j to student i is given as

$$v_s(i, j; \varepsilon_{i,j}, \theta_c) = \theta_c x_{j,c} + \varepsilon_{ij},$$

and the value of a student j to the colleges' is given as

$$v_c(i; \eta_i, \theta_s) = \theta_s x_{i,s} + \eta_i.$$

In the simulations, we choose $x_{s,i}$ and $x_{c,j}$ to be discrete, scalar random variables, drawing them i.i.d. from the uniform distribution on $\{1, 2, 3\}$. The variables ε_{ij} and η_i are drawn i.i.d. from $N(0, 1)$ and independently of one another and of the covariates. We consider the case that each college has an equal number of positions, K and consider the performance of the inference procedure for $K \in \{5, 10, 20\}$. For the simulations, we set the true value of the preference parameter to $\theta_0 = (1, 1)$. The simulation numbers R and B are chosen from $\{100, 500\}$ and Monte Carlo simulation number is set to be 1000.

We consider the test statistics in (5):

$$\frac{1}{B} \sum_{b=1}^{B} \max_{x_s, x_c} \left| \hat{P}(x_s, x_c; \mathbf{Y}, \mathbf{X}) - \hat{P}(x_s, x_c; \mathbf{Y}_b^*(\theta), \mathbf{X}) \right|,$$

where

$$\hat{P}(x_s, x_c; \mathbf{Y}, \mathbf{X}) = \frac{1}{n_s} \sum_{i \in N_s} 1\{x_{i,s} = x_s, \mathbf{X}_{\mathbf{Y}(i),c} = x_c\}.$$

The test statistic compares the observed and predicted joint distribution of covariates between matched agents.

The results are reported in Tables 1 and 2. The results in Table 1 are from using $R = B = 100$ and those in Table 2 are from using $R = B = 500$. In general, the size and power properties of the inference procedure are acceptable. There is little size distortion (particularly for $K = 10$ and $K = 20$), and the rejection probabilities increase in the sample size for alternatives away from the true parameter value. It is interesting to note that the results are not very different between the two tables. This means that considering the substantial increase in the computational cost using a higher value of R, B, it appears using $R = B = 100$ is just enough for practical purposes.

Table 1. Empirical Rejection Probabilities for Monte Carlo Inference on Preferences, $R = B = 100$.

θ_s	θ_c	$K = 5$			$K = 10$			$K = 20$		
		0.5	1.0	1.5	0.5	1.0	1.5	0.5	1.0	1.5
0.5	$n = 200$	0.9870	0.8720	0.8180	0.9700	0.8340	0.7620	0.9470	0.7840	0.6680
	$n = 400$	1.0000	0.9990	0.9920	1.0000	1.0000	0.9960	1.0000	0.9980	0.9810
	$n = 600$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.0	$n = 200$	0.2320	0.0640	0.1040	0.2430	0.0620	0.0830	0.2500	0.0470	0.1150
	$n = 400$	0.4170	0.0340	0.0770	0.4360	0.0450	0.0970	0.5100	0.0500	0.1010
	$n = 600$	0.6170	0.0580	0.0910	0.6290	0.0460	0.1040	0.6590	0.0510	0.1110
1.5	$n = 200$	0.0700	0.6820	0.9340	0.0790	0.6380	0.9210	0.0840	0.5890	0.9100
	$n = 400$	0.1040	0.9660	0.9980	0.1350	0.9520	1.0000	0.1570	0.9150	0.9960
	$n = 600$	0.1590	0.9990	1.0000	0.1810	0.9960	1.0000	0.2300	0.9940	1.0000

Notes: This table explores the size and power properties for the inference on student and college preferences when each college has the same number of positions, K. The true value of the parameter is $\theta_{s0} = \theta_{c0} = 1$. The simulation number is 1000. In each iteration of the simulation loop, we use 100 random draws to compute the critical value.

Table 2. Empirical Rejection Probabilities for Monte Carlo Inference on Preferences: $R = B = 500$.

θ_s	θ_c	K = 5			K = 10			K = 20		
		0.5	1.0	1.5	0.5	1.0	1.5	0.5	1.0	1.5
0.5	$n = 200$	0.9860	0.8850	0.8300	0.9760	0.8370	0.7470	0.9520	0.7820	0.6670
	$n = 400$	1.0000	0.9990	0.9960	1.0000	0.9990	0.9960	1.0000	0.9980	0.9860
	$n = 600$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.0	$n = 200$	0.2280	0.0610	0.0870	0.2530	0.0510	0.0910	0.2400	0.0490	0.1110
	$n = 400$	0.4230	0.0310	0.0690	0.4470	0.0530	0.0960	0.4940	0.0510	0.0920
	$n = 600$	0.6280	0.0570	0.0880	0.6250	0.0470	0.0960	0.6640	0.0450	0.1110
1.5	$n = 200$	0.0650	0.6820	0.9510	0.0710	0.6500	0.9210	0.0910	0.5990	0.9200
	$n = 400$	0.1010	0.9650	0.9980	0.1260	0.9520	1.0000	0.1590	0.9150	0.9960
	$n = 600$	0.1540	0.9990	1.0000	0.1710	0.9950	1.0000	0.2100	0.9930	1.0000

Notes: This table explores the size and power properties for inferring student and college preferences when each college has the same number of positions, K. The true value of the parameter is $\theta_{s0} = \theta_{c0} = 1$. The simulation number is 1000. In each iteration of the simulation loop, we use 500 random draws to compute the critical value.

6. Conclusions

This paper proposes Monte Carlo inference for a large matching model. The main challenge for inference in a large matching model is to deal with a complex form of cross-sectional dependence created by strategic interdependence between agents. Being a finite sample inference method, Monte Carlo inference can be used to deal with this difficulty, when the matching mechanism is explicitly known to the econometrician. Although we do not prove the power properties of the inference as the number of the agents grows, our Monte Carlo simulation suggest that the confidence intervals will shrink as the sample sign grows, which would indicate an accumulation of information as the number of agents grows.

Monte Carlo inference exhibits some limitations. First, the inference works only when the matching process is fully parametrized. When the parameter is high dimensional, Monte Carlo inference can be computationally costly. Second, it only applies to a situation where the econometrician knows precisely the underlying matching mechanism. One may take a known stable matching mechanism such as a Deferred-Acceptance mechanism as part of econometric specifications. However, it would be desirable to pursue an empirical model which does not require a full specification of a matching mechanism. This direction of research appears promising and is left to future work.

Author Contributions: All the authors equally contributed to this work.

Appendix A. Serial Dictatorship Algorithm

In this section, we provide a MATLAB code for serial dictatorship algorithm which returns a serial dictatorship matching of students to college positions given heterogeneous student preferences and homogeneous college preferences. Let us introduce the definition of variables.

- `x_s, x_c`: $n_s \times 1$ vector of student characteristics and $n_c \times 1$ vector of college characteristics.
- `theta_s, theta_c`: scalar student and college preference parameters.
- `N_s, N_c`: number of students and colleges.
- `student_score`: $n_s \times 1$ vector of college preferences over students, based on `theta_s`, `x_s` and standard normal random variables.

- `coll_rank`: The indices associated with the preferred students of colleges. For example, `coll_rank(1)` is the index of the most preferred student according to colleges.
- `val_ji`: $n_c \times n_s$ matrix whose (j, i) element is associated with the value that student i places on college j based on `theta_c`, `x_c` and standard normal random variables.
- `pos_vec`: $n_s \times 1$ vector whose ith element says the college associated with the ith college position. For example, if `pos_vec` $= [1, 1, 2, 2, 3]'$, it means that there are three colleges, where college 1 and 2 each have two positions and college 3 has 1.
- `val_ji_pos` is an $n_s \times n_s$ matrix whose (j, i) element is the value that student i has for position j.
- `val_fv`: $n_s \times n_s$ matrix whose (j, r) value is the value that the student ranked r highest according to coll_rank has for college position j.
- `matching`: $n_s \times 1$ vector, where `matching(i)` $= j$ means student $i \in \{1, ..., n_s\}$ is matched with college position $j \in \{1, ..., n_s\}$.

The following MATLAB code of function `serial_dictatorship` returns `matching` given `theta_s` `theta_c`, `x_c`, `x_s`, `N_s`, `N_c`, and `pos_vec` when the colleges' ranking of students is based on students' score generated from a model with additive normal errors. One can change this specification in the code for alternative specification of the way colleges rank students.

```
function matching =serial_dictatorship(theta_s, theta_c, x_c, x_s, N_s, N_c,pos_vec)

student_score = theta_s*x_s + normrnd(0,1,N_s,1);
[~,coll_rank] = sort(student_score,'descend');

val_ji=repmat((theta_c*x_c),[1 N_s]) + normrnd(0,1,N_c,N_s);
val_ji_pos=val_ji(pos_vec,:);
val_fv=val_ji_pos(:,coll_rank);
inverse_matching=zeros(N_s,1);
matching=inverse_matching;
id=1:N_s;

for i=1:N_s
    [~,index]=max(val_fv(:,i));
    inverse_matching(coll_rank(i))=index;
    val_fv(index,:)=-inf;
end

matching(inverse_matching)=id;
end
```

References

Agarwal, Nikhil, and Paulo Somaini. 2018. Demand analysis using strategic reports: An application to a school choice mechanism. *Econometrica* 86: 391–444. [CrossRef]

Agarwal, Nikhil. 2015. An empirical model of the medical match. *American Economic Review* 105: 1939–78. [CrossRef]

Barndorff-Nielsen, Ole. 1983. On a formula for the distribution of the maximum likelihood estimator. *Biometrika* 70: 343–65. [CrossRef]

Besag, Julian, and Peter Clifford. 1989. Generalized monte carlo significance tests. *Biometrika* 76: 633–42. [CrossRef]

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2013. Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics* 31: 83–117. [CrossRef]

Bugni, Federico A., Ivan A. Canay, and Xiaoxia Shi. 2017. Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics* 8: 1–38. [CrossRef]

Canen, Nathan, Jacob Schwartz, and Kyungchul Song. 2018. Estimating local interactions among many agents who observe their neighbors. *arXiv*. arXiv:1704.02999v3.

Chen, Jiawei, and Kejun Song. 2013. Two-sided matching in the loan market. *International Journal of Industrial Organization* 31: 145–52. [CrossRef]

Chiappori, Pierre-André, Sonia Oreffice, and Climent Quintana-Domeque. 2012. Anthropometric and socioeconomic matching on the marriage-market. *Journal of Political Economy* 120: 659–95. [CrossRef]

Choo, Eugene, and Aloysius Siow. 2006. Who marries whom and why. *Journal of Political Economy* 114: 175–201. [CrossRef]

Diamond, William, and Nikhil Agarwal. 2017. Latent indices in assortative matching models. *Quantitative Economics* 8: 685–728. [CrossRef]

Dufour, Jean-Marie, and Lynda Khalaf. 2003. Monte carlo test methods in econometrics. In *A Companion to Theoretical Econometrics*. Hoboken: Blackwell Publishing, pp. 494–519.

Dufour, Jean-Marie. 2006. Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics* 133: 443–77. [CrossRef]

Dwass, Meyer. 1957. Modified randomization tests for nonparametric hypotheses. *American Mathematical Statistics* 57: 181–87. [CrossRef]

Fox, Jeremy T., and Patrick Bajari. 2013. Measuring the efficiency of an fcc spectrum auction. *American Economic Journal: Microeconomics* 5: 100–46. [CrossRef]

Fox, Jeremy T. 2010. Estimating Matching Games with Transfers. *Quantitative Economics* 9: 1–38. [CrossRef]

Galichon, Alfred, and B. Salanié. 2012. Cupid's Invisible Hand: Social Surplus and Identification in Matching Models. Working paper. Available online: https://hal-sciencespo.archives-ouvertes.fr/hal-01053710/ (accessed on 26 March 2019).

Graham, Bryan S., Guido W. Imbens, and Geert Ridder. 2014. Complementarity and aggregate implications of assortative matching: A nonparametric analysis. *Quantitative Economics* 5: 29–66. [CrossRef]

Hitsch, Günter J., Ali Hortacsu, and Dan Ariely. 2010. Matching and sorting in online dating. *American Economic Review* 100: 130–63. [CrossRef]

Hope, Adery C. A. 1968. A simplified monte carlo significance test procedure. *Journal of the Royal Statistical Society* 1968: 582–98. [CrossRef]

Menzel, K. 2015. Large matching markets as two-sided demand systems. *Econometrica* 83: 897–941. [CrossRef]

Park, Minjung. 2013. Understanding merger incentives and outcomes in the us mutual fund industry. *Journal of Banking and Finance* 37: 4368–80. [CrossRef]

Romano, Joseph P., and Azeem M. Shaikh. 2008. Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference* 138: 2786–807. [CrossRef]

Roth, Alvin E., and Marilda A. OliveiraSotomayor. 1990. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Econometric Society Monograph Series. Cambridge: Cambridge University Press.

Satterthwaite, Mark A., and Hugo Sonnenschein. 1981. Strategy-proof allocation mechanisms at differentiable points. *Review of Economic Studies* 48: 587–97. [CrossRef]

Schwartz, Jacob. 2018. Schooling Choice, Labour Market Matching, and Wages. *arXiv*. arXiv:1803.09020.

Sørensen, Morten. 2007. How smart is smart money? A two-sided matching model of venture capital. *Journal of Finance* 62: 2725–62. [CrossRef]