Learning Strategies in Decentralized Matching Markets under Uncertain Preferences

Xiaowu Dai xwdai@berkeley.edu
University of California, Berkeley, CA 94720-1776, USA

Michael I. Jordan jordan@cs.berkeley.edu
University of California, Berkeley, CA 94720-1776, USA

Editor:

Abstract

We study two-sided decentralized matching markets in which participants have uncertain preferences. We present a statistical model to learn the preferences. The model incorporates uncertain state and the participants' competition on one side of the market. We derive an optimal strategy that maximizes the agent's expected payoff and calibrate the uncertain state by taking the opportunity costs into account. We discuss the sense in which the matching derived from the proposed strategy has a stability property. We also prove a fairness property that asserts that there exists no justified envy according to the proposed strategy. We provide numerical results to demonstrate the improved payoff, stability and fairness, compared to alternative methods.

Keywords: Decentralized Matching, Uncertain Preference, Calibration, Stability, Fairness

1. Introduction

Many real-world decision-making problems can be viewed from an economic point of view and a statistical point of view. The economic point of view focuses on scarcity of shared resources and the need to coordinate among multiple decision-makers. Thus, decision-makers must assess preferences over outcomes and those preferences need to interact in determining an overall set of outcomes. The statistical point of view recognizes that preferences are often not known a priori, but must be learned from data; moreover, agents' decisions are often influenced by latent state variables whose values must be inferred in order to determine a preferred outcome. Unfortunately, it is uncommon that these two perspectives are brought together in the literature, with economic work rarely addressing the need to learn preferences from data, and statistical machine learning rarely addressing scarcity and its consequences for decision-making. In this paper we aim to bridge this gap, studying a core microeconomic problem—two-way matching markets—in the setting in which preferences must be learned from data. Moreover, we focus on *decentralized* matching markets, reflecting several desiderata that are common in the machine-learning literature—that agents are autonomous and private, and that scalability and avoidance of central bottlenecks is a principal concern of an overall system design.

Two-sided matching markets have played an important role in microeconomics for several decades (see Roth and Sotomayor, 1990), both as a theoretical topic and as a mainstay of real-world applications. Matching markets are used to allocate indivisible "goods" to multiple decision-making agents based on mutual compatibility as assessed via sets of preferences. Matching markets are often organized in a decentralized way such that each agent makes their decision independently of others' decisions. Examples include college admissions, decentralized labor markets, and online dating.

Matching markets embody a notion of scarcity in which the resources on both sides of the market are limited. Moreover, *congestion* is a key issue in decentralized decision-making under scarcity, as participants may not be able to make enough offers and acceptances to clear the market (Roth and Xing, 1997). A major challenge for research on decentralized matching markets is to ease congestion in practice and to understand the implication of congestion for stability, fairness, and welfare.

The uncertainty of participants' preferences is ubiquitous in real-world decentralized matching markets. For instance, college admissions in the United States face applicants' uncertain preferences. The admitted students of a college may receive offers from other colleges. Students need to accept one or reject all offers, often within a *short* period. The process provides little opportunity for the college to learn students' preferences, which depends on colleges' competition and colleges' uncertain popularity in the current year. Consequently, the college may end up enrolling too many or too few students relative to its capacity (Avery et al., 2003). It is of college's interest to decide which applicants to admit, such that the entering class will meet reasonably close to its quota and be close to the attainable optimum in quality (Gale and Shapley, 1962).

This paper develops a statistical model to study participants' strategies under uncertain preferences in two-way matching markets. We use the classical college admissions market as our running example. In the proposed model, there are a set of agents (for example, colleges), each with limited capacity, and a set of arms (for example, students), each can be matched to at most one agent. Agents value two attributes of an arm: a "score" (for example, SAT/ACT score) that is common to all agents and a "fit" (for example, college-specific essay) that is agent-specific and is independent across agents. According to their score and fit, agents rank arms, but they do not observe arms' preferences, which have no restriction. The model incorporates the arm's uncertain preference into an acceptance probability, depending on both the unknown state and agents' competition. We want to learn arms' acceptance probabilities from the historical data of arms' binary choices, where arms' attributes may vary over time. Various statistical learning algorithms allow efficient learning of the acceptance probability under the proposed model. To fix ideas, we present the penalized log-likelihood method in the reproducing kernel Hilbert space (RKHS) to learn the acceptance probability. We show that the estimate achieves the minimax rate-optimality.

We focus on the single-stage decentralized matching that involves a simple timeline: agents simultaneously pull sets of arms (for example, colleges offer admissions to students). Each arm accepts one of the agents (if any) that pulled it. We derive an optimal strategy called calibrated decentralized matching (CDM), which maximizes agents' expected payoffs. We calibrate the unknown state by perturbing the state and balancing the marginal utility and the marginal penalty. The proposed calibration procedure takes the opportunity costs into account. The CDM can perform the calibration in both the average-case and worst-case scenarios, depending on maximizing the averaged or minimal expected payoff concerning the unknown true state.

We show that asymptotically CDM makes it safe for agents to act straightforwardly on their preferences. That is, CDM ensures incentive compatibility for agents. We also show that CDM yields a stable outcome for the market. Our notion of stability is similar to that of Liu et al. (2014), which is designed to study decentralized matching with incomplete information. This stability notion extends the classical notion of stability due to Gale and Shapley (1962) that assumes complete information of participants' preferences. We prove that CDM is asymptotically stable. Moreover, we show that CDM is asymptotically fair for arms, in the sense of no justified envy in the matching procedure according to CDM.

The CDM algorithm can serve as a recommendation engine in decentralized matching markets. Indeed, our results add a learning component to the decentralized market and help participants decide which participants on the other side of the market are the best to connect to. In particular, even with an unknown state, agents can estimate the probability of successfully pulling an arm using historical data. The prediction of match compatibility is also possible in another direction that arms can learn how much an agent may prefer them. The CDM procedure reduces congestion in the decentralized matching market via recommendation and achieves optimal expected payoffs for agents.

1.1 Related Work

We propose a statistical model for learning strategies in decentralized matching markets. In particular, we focus on many-to-one decentralized matching with uncertain participants' preferences. Our work is related to two strands of literature. First, there has been significant work in the economics literature on congestion in decentralized markets, where participants cannot make enough offers and acceptances to clear the market. Roth and Xing (1997) discussed such a decentralized market for graduating clinical psychologists, especially the market's timing aspect, which lasts over a day. They found that such a decentralized but coordinated market exhibits congestion since the interviews that a student could schedule were limited, and the resulting matching was unstable. Das and Kamenica (2005) considered that both sides of the decentralized market have uncertain preferences and presented an empirical study of the resulting matchings. Unlike these works, we provide an analytical model and study the implications for congestion. Haeringer and Wooders (2011) considered the decentralized job matching with complete information, where agents and arms are assumed to know the entire sequence of actions employed by their opponents. However, we consider incomplete information, where agents do not know their opponents' actions. Coles et al. (2013) showed how introducing a signaling device in a decentralized matching market alleviates congestion. We instead focus on the optimal strategy under uncertain preference without signaling. Liu et al. (2020) extended the multi-armed bandits framework (see Bubeck and Cesa-Bianchi, 2012) to multiple agents. They proposed an algorithm to achieve a low cumulative regret for decentralized matching. By contrast, we study the optimal strategy for single-stage matching that involves no accumulated regret.

Another closely related literature strand is the algorithmic and economic problems on college admissions (see Gale and Shapley, 1962). The general setting involves multiple colleges competing for students in decentralized markets. Recently, Epple et al. (2006)

modeled equilibrium admissions, financial aid, and enrollment. Fu (2014) studied effects of tuition on equilibria, incorporating application costs, admissions, and enrollment. Unlike these works, we emphasize students' multidimensional abilities and the uncertainty of students' acceptance probability. Chade and Smith (2006) provided an optimal algorithm for students' application decisions, characterized as a portfolio choice problem. Hafalir et al. (2018) discussed student efforts instead of colleges' response to congestion in decentralized college admissions with restricted applications. We instead consider colleges' optimal strategies. Avery and Levin (2010) studied early admissions when colleges have no enrollment uncertainty and showed that a cutoff strategy is optimal in equilibrium. Chade et al. (2014) developed a decentralized Bayesian model of college admissions for two colleges and a continuum of students under a particular preference structure. By contrast, we study multiple colleges in the face of enrollment uncertainty. Azevedo and Leshno (2016) adopted a continuum model for students in a centralized market. They found a characterization of equilibria in terms of supply and demand. This model is different from the decentralized market studied in our paper. Che and Koh (2016) considered aggregated uncertainty in college admissions and focused on two colleges and a continuum of students. By contrast, our model considers incomplete information with multiple colleges and a finite number of arms. We study a statistical model for learning strategies in the face of enrollment uncertainty using historical data.

The paper is organized as follows. Section 2 introduces the model. Section 3 studies the optimal strategy in decentralized matching under uncertain preferences. Section 4 shows the proposed strategy's properties, including incentive compatibility, stability and fairness. Section 5 presents the results of an empirical study. Section 6 concludes the paper with further research directions. Technical proofs are provided in the Appendix.

2. Model and Uncertain Preferences

In this section, we define a model of decentralized matching markets with incomplete information. Denote a set of m agents by $\mathcal{P} = \{P_1, P_2, \ldots, P_m\}$ and a set of n arms by $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$, where $\mathcal{P}$ and $\mathcal{A}$ are participants on the two sides of the market. Each agent can attempt to pull multiple arms and there are no constraints on the overlap among the choices of different agents. When multiple agents select the same arm, only one agent can successfully pull the arm, with the choice of agent made according to the arm's preferences. For example, $\mathcal{P}$ and $\mathcal{A}$ might represent colleges and students in the college admissions market, or firms and workers in the decentralized labor market. Colleges send admission offers to applicants. When multiple colleges send offers to the same applicant, the applicant can accept at most one offer. Suppose that each agent P_i has a quota of successfully pulling $q_i \geq 1$ arms, where $q_1 + q_2 + \cdots + q_m \leq n$. We denote $[m] \equiv \{1, \ldots, m\}, [n] \equiv \{1, \ldots, n\}$, and $[K] \equiv \{1, \ldots, K\}$.

In decentralized matching markets, agents and arms make their decisions independently of the decisions made by others. This feature distinguishes decentralized matching from centralized matching, which makes use of central clearinghouses to coordinate decision-making. Notable examples of centralized matching include the national medical residency matching (Roth, 1984) and public school choice (Abdulkadiroğlu and Sönmez, 2003). While centralized matching has been the major focus of the literature on matching, we see decentralized matching as having a potentially greater range of applications, and as providing a better platform on which to bring machine-learning tools to bear. Indeed, decentralized matching is natural when (i) arms have uncertain preferences that depend on an unknown state that has a local component, and (ii) agents possess incomplete information on other agents' decisions.

2.1 Running Example

Our running example is the college admissions market, where colleges match with students (see Gale and Shapley, 1962). College admissions in countries such as the United States, Korea, and Japan are organized in a decentralized way, where colleges make admission offers to applicants. The admitted students accept or reject the offers, often within a short period of time. Consider a college $P_i \in \mathcal{P}$ with a quota of enrolling q_i students. It is not satisfactory for college P_i to only make offers to the q_i best-qualified students since some students may reject offers. The enrollment uncertainty for P_i can be attributed to a lack of two types of information. First, P_i has little knowledge of which other colleges admit the applicants admitted by P_i. Second, P_i is uncertain about students' preferences: how each applicant ranks the colleges that she has applied to. Thus colleges do not know their own popularity in the current year, which is an aggregate over the uncertain rankings by each applicant. In the face of such uncertainty, college P_i needs to decide which applicants to admit such that the entering class will meet reasonably close to its quota q_i and be close to the attainable optimum in quality.

2.2 Latent Utility

We assume that agents' deterministic preferences are a function of underlying latent utilities. In particular, we consider the following latent utility model:

$$U_i(A_j) = v_j + e_{ij}, \quad \forall i \in [m] \text{ and } j \in [n], \tag{1}$$

where $v_j \in [0, 1]$ is arm A_j's *systematic score*, which is available to all agents, and $e_{ij} \in [0, 1]$ is an agent-specific *idiosyncratic fit* available only to agent P_i. For example, in college admissions, v_j can be a function of student A_j's test score on a nationwide test observed by all colleges. The e_{ij} corresponds to a function of student A_j's performance on college-specific essays or tests conducted by college P_i. In Appendix A.1, we show that a general utility function with multidimensional scores and fits can be transformed to the model (1) via the ANOVA decomposition. Thus, the separable structure in model (1) is without less of generality and allows us to characterize the pattern of competition of agents in Section 3. Similar separable structures have been used in the matching markets literature (see, e.g., Choo and Siow, 2006; Menzel, 2015; Chiappori and Salanié, 2016; Ashlagi et al., 2020). The analysis that we present in this paper will also hold if one restricts the range of e_{ij} to $[0, \bar{e}]$ with some $\bar{e} < 1$. This restriction may be necessary for some applications. For instance, suppose the idiosyncratic fit in college admissions is viewed as more important to colleges than the systematic score. In this case, enrolled students may find it unfair to have other enrolled students with significantly lower test scores than theirs. This, in turn, would result in a reputation cost for the college.

2.3 Uncertain Preferences

The arms' preferences for agents involve uncertainty. Let the parameter $s_i \in [0,1]$ represent the *state of the world* for agent P_i (Savage, 1972), such that each arm $A_j \in \mathcal{A}$ accepts P_i with probability

$$\pi_i(s_i, v_j), \quad \forall i \in [m] \text{ and } j \in [n].$$

Here $\pi_i(s_i, v_j)$ models the competition of agents through the dependence on the score v_j. Moreover, $\pi_i(s_i, v_j)$ incorporates the uncertainty of an arm's preference with respect to P_i and the competing agents via the state s_i. We assume that $\pi_i(s_i, v_j)$ is strictly increasing and continuous in s_i. Thus, a large value of s_i corresponds to the case that agent P_i is popular. The true state is unknown a priori to P_i. In college admissions, for example, the *yield* is defined as the rate at which a college's admitted students accept the offers. The yield is a proxy for the state s_i (see Che and Koh, 2016). The yield of the current year is unknown a priori to the college.

Theorem 1 *There exists a probability mass function $\pi_i(s_i, v_j)$ characterizing the probability of A_j accepting P_i. Moreover, the expected utility that agent P_i receives from pulling arm $A_j \in \mathcal{A}$ is*

$$\mathbb{E}[utility] = (v_j + e_{ij}) \cdot \pi_i(s_i, v_j), \quad i \in [m] \text{ and } j \in [n]. \tag{2}$$

For example, we show the explicit form of $\pi_i(s_i, v_j)$ for a *two-agent* model with agents P_1 and P_2. Consider that P_1 pulls an arm $A_j \in \mathcal{A}$. Denote by $\mu_1(s_1)$ the probability that an arm prefers P_1. Let $\mu_2(s_2) = 1 - \mu_1(s_1)$ be the probability that an arm prefers P_2, where the state $s_2 = 1 - s_1$. We write σ_2 as P_2's strategy, which is defined by $\sigma_2(v_j, e_{2j}) = \mathbf{1}\{P_2 \text{ pulls } A_j\}$. Since A_j would be pulled by P_1 with probability $1 - \sigma_2(v_j, e_{2j})$ and pulled by both P_1 and P_2 with probability $\sigma_2(v_j, e_{2j})$, A_j would accept P_1 with probability $1 - \sigma_2(v_j, e_{2j}) + \mu_1(s_1)\sigma_2(v_j, e_{2j})$. Since e_{2j} is unknown to A_j and σ_2 is determined by e_{2j}, the averaged probability that A_j accepts P_1 is $\pi_1(s_1, v_j) = 1 - \mathbb{E}[\sigma_2(v_j, e_{2j})] + \mu_1(s_1)\mathbb{E}[\sigma_2(v_j, e_{2j})]$, where the expectation is taken over e_{2j}. Hence, $\pi_1(s_1, v_j)$ represents the probability of A_j accepting P_1 due to the uncertain state and the competition between agents.

2.4 Learning Arms' Uncertain Preferences

Agents decide which arms to pull based on the attributes of the arms and historical matches. In general, we do not make the assumption that an arm with the same score and fit appears in the historical data in the matching market. Indeed, no college admissions applicant likely has exactly the same attributes as the applicants in previous years. Even the "repeated applicants" will often change their records in their second applications. The repeated applicants either have no offer or reject all their offers and wait a full year to apply for colleges. The value of learning from the historical matches accrues when the data are used to estimate an untried arm's acceptance probability.

Denote by $\mathcal{A}^t = \{A_1^t, A_2^t, \ldots, A_{n_t}^t\}$ the set of arms at $t \in [T] \equiv \{1, \ldots, T\}$. Let s_i^t be the state of agent P_i at time t. The state s_i^t is *unknown* until time $t+1$, and it varies over time. For instance, the yield rate of a college may change over the years. For any arm $A_j^t \in \mathcal{A}^t$, there are an associated pair of score and fit values, (v_j^t, e_{ij}^t), obtained from Eq. (1), where $i \in [m], j \in [n_t]$. We refer to (v_j^t, e_{ij}^t) as the attributes of arm A_j^t. Define the set

$$\mathcal{B}_i^t = \{j \mid P_i \text{ pulls arm } A_j^t \text{ at time } t \text{ for } 1 \leq j \leq n_t\},$$

where $|\mathcal{B}_i^t| = n_{it}$ and $n_{it} \leq n_t$. For any $j \in \mathcal{B}_i^t$, the outcome that P_i observes is whether an arm A_j^t accepted P_i, that is, $y_{ij}^t = \mathbf{1}\{A_j^t \text{ accepts } P_i\}$. Here, the iid outcome y_{ij}^t has the likelihood $\pi_i(s_i^t, v_j^t)$ that $y_{ij}^t = 1$ and the likelihood $1 - \pi_i(s_i^t, v_j^t)$ that $y_{ij}^t = 0$.

The goal in this section is to estimate the function π_i based on historical data. Denote the training data by $\mathcal{D} = \{(s_i^t, v_j^t, e_{ij}^t, y_{ij}^t) : i \in [m]; j \in \mathcal{B}_i^t; t \in [T]\}$. Let the log odds ratio be

$$f_i(s_i, v) = \log\left(\frac{\pi_i(s_i, v)}{1 - \pi_i(s_i, v)}\right).$$

There exist a variety of methods in supervised learning that can efficiently estimate the log odds ratio (Hastie et al., 2009). For concreteness, we focus on using a penalized log-likelihood method for the log odds ratio estimation in a reproducing kernel Hilbert space (RKHS) (Wahba, 1999). To this end, we assume that f_i is from RKHS $\mathcal{H}_{K_i}$ with the reproducing kernel K_i. Then we find $f_i \in \mathcal{H}_{K_i}$ to minimize

$$\sum_{t=1}^{T} \sum_{j \in \mathcal{B}_i^t} \left[-y_{ij}^t f_i(s_i^t, v_j^t) + \log\left(1 + \exp\left(f_i(s_i^t, v_j^t)\right)\right)\right] + \frac{1}{2} \sum_{t=1}^{T} n_{it} \lambda_i \|f_i\|_{\mathcal{H}_{K_i}}^2, \tag{3}$$

where $\|\cdot\|_{\mathcal{H}_{K_i}}$ is a RKHS norm and $\lambda_i \geq 0$ is a tuning parameter. Consider a tensor product structure of the RKHS $\mathcal{H}_{K_i}$ defined $K_i((s_i, v), (s_i', v')) = K_i^s(s_i, s_i') K_i^v(v, v')$ based on kernels K_i^s and K_i^v (Wahba et al., 1995). Assume a random feature expansion of the following form: $K_i^s(s_i, s_i') = \mathbb{E}_{w_s}[\phi_i^s(s_i, w_s)\phi_i^s(s_i', w_s)]$ and $K_i^v(v, v') = \mathbb{E}_{w_v}[\phi_i^v(v, w_v)\phi_i^v(v', w_v)]$, where $\phi_i^s(\cdot, w_s)$ and $\phi_i^v(\cdot, w_v)$ are random features (Rahimi and Recht, 2008). Let $\{w_{s1}, w_{s2}, \ldots, w_{sp}\}$ and $\{w_{v1}, w_{v2}, \ldots, w_{vp}\}$ be the sets of p independent copies of w_s and w_v, respectively. Write the feature vector as $\psi_i(s_i, v) \in \mathbb{R}^p$ with the lth entry equal to $\frac{1}{\sqrt{p}}\phi_i^s(s_i, w_{sl})\phi_i^v(v, w_{vl})$, for $l = 1, \ldots, p$. Then $K_i((s_i, v), (s_i', v'))$ can be approximated by the product $\psi_i(s_i, v)^\intercal \psi_i(s_i', v')$. Let the matrix Φ_i have rows $\psi_i(s_i^t, v_j^t)^\intercal$, where $j \in \mathcal{B}_i^t$ and $t \in [T]$. By the representer theorem in Kimeldorf and Wahba (1971), the solution to Eq. (3) has the form $\widehat{f}_i(s_i, v) = \psi_i(s_i, v)^\intercal \Phi_i^\intercal c_i$ for some vector c_i. We only need to find $\theta_i = \Phi_i^\intercal c_i \in \mathbb{R}^p$ to obtain a solution to Eq. (3):

$$\widehat{f}_i(s_i, v) = \psi_i(s_i, v)^\intercal \theta_i, \quad \forall i \in [m]. \tag{4}$$

Denote by Y_i the response vector with entries y_{ij}^t, $j \in \mathcal{B}_i^t$ and $t \in [T]$. Applying the Newton-Raphson method to Eq. (3) yields the following iterative updates for vector θ_i:

$$\theta_i^{(\nu+1)} = \left(\Phi_i^\intercal W_i^{(\nu)} \Phi_i + \sum_{t=1}^{T} n_{it} \lambda_i \mathbf{I}\right)^{-1} \Phi_i^\intercal W_i^{(\nu)} \left\{\Phi_i \theta^{(\nu)} + (W_i^{(\nu)})^{-1}[Y_i - \pi_i^{(\nu)}]\right\}, \ \nu \geq 1.$$

Here, $\theta_i^{(\nu)}$ is the νth update of θ_i, and $W_i^{(\nu)} = \operatorname{diag}[\pi_i^{(\nu)}(s_i^t, v_j^t)(1 - \pi_i^{(\nu)}(s_i^t, v_j^t))]_{j \in \mathcal{B}_i^t, 1 \leq t \leq T}$ is a weight matrix with $\pi_i^{(\nu)} = [1 + \exp(-f_i^{(\nu)})]^{-1}$ and $f_i^{(\nu)} = \phi_i^\intercal \theta_i^{(\nu)}$. The tuning parameter $\lambda_i \geq 0$ can be selected using cross validation or GACV (see Wahba, 1999).

Theorem 2 *The integrated squared error of the estimate in Eq. (4) satisfies the following inequality:*

$$\mathbb{E}[(\widehat{f}_i - f_i)^2] \leq c_f \left[T(\log T)^{-1}\right]^{-2r/(2r+1)}, \quad \forall i \in [m],$$

for sufficiently large T, where we let $\lambda_i \leq c_\lambda [T(\log T)^{-1}]^{-2r/(2r+1)}$ and $p \geq c_p [T(\log T)^{-1}]^{-2r/(2r+1)}$. Here the constants $c_f, c_\lambda, c_p > 0$ are independent of T, and $r \geq 1$ denotes the smoothness of kernels such that $K_i^s(s, \cdot)$ and $K_i^v(v, \cdot)$ have squared integrable rth-order derivatives. Moreover, the estimate in Eq. (4) is minimax rate-optimal.

We make three remarks on the theorem. First, although f_i depends on both covariates s and v, the optimal rate given in the theorem is very close to the minimax rate for the one-dimensional model, with the only difference in the logarithm term. Second, it is possible to extend the theorem to allow different orders of the smoothness of the kernels K_i^s and K_i^v. Finally, we can predict the acceptance probability for a new arm. Let $\mathcal{A}^{T+1} = \{A_1, \ldots, A_n\}$ be the set of arms at time $T+1$, where each arm A_j has attributes obtained from Eq. (1). Using the estimated log odds ratio $\widehat{f}_i$ in Eq. (4) and given the state s_i at time $T+1$, we have the estimate

$$\widehat{\pi}_i(s_i, v_j) = [1 + \exp(-\widehat{f}_i(s_i, v_j))]^{-1} \tag{5}$$

as the prediction of the probability that arm A_j would accept agent P_i at time $T+1$.

3. Optimal Strategies in Decentralized Matching

We study single-stage decentralized matching that involves a simple timeline. First, Nature draws a state denoted by s_i^* for agent P_i such that the arms' preferences are realized. Next, arms simultaneously show their interests to all agents. For example, students apply to colleges. Under the assumption that students face negligible application costs, submitting applications to all colleges is the dominant strategy. The reason is that students do not know how colleges evaluate their academic ability or personal essays (Avery and Levin, 2010; Che and Koh, 2016). Then, agents simultaneously decide which arms to pull based on the arm's attributes. Finally, each arm accepts one of the agents (if any) that pulled it. See an illustration in the left plot of Figure 1.

3.1 Agent's Expected Payoff

An agent's expected payoff consists of two parts: the expected utilities of arms that the agent pulls and the penalty for exceeding the quota. Let $\mathcal{B}_i(s_i) \subseteq \mathcal{A}^{T+1}$ be the set of arms that agent P_i pulls at $T+1$ given the state s_i. By Eq. (2), agent P_i's expected payoff is

$$\mathcal{U}_i[\mathcal{B}_i(s_i)] = \sum_{j \in \mathcal{B}_i(s_i)} (v_j + e_{ij}) \cdot \pi_i(s_i, v_j) - \gamma_i \cdot \max \left\{ \sum_{j \in \mathcal{B}_i(s_i)} \pi_i(s_i, v_j) - q_i,\ 0 \right\}. \tag{6}$$

Here, γ_i denotes the marginal penalty for exceeding the quota. We assume that $\gamma_i > \max_{j \in \mathcal{A}^{T+1}} \{v_j + e_{ij}\}$; that is, the penalty is larger than an arm's latent utility. Eq. (6) excludes the situation when P_i faces extra uncertainty in receiving the reward even if P_i has successfully pulled an arm. For example, in the dating market, Eq. (6) models the agent's expected payoff from the dates, instead of a subsequent relationship that may eventually result from a date. The following theorem shows that the agent's optimal strategy is the *cutoff strategy* with respect to the arms' latent utilities.

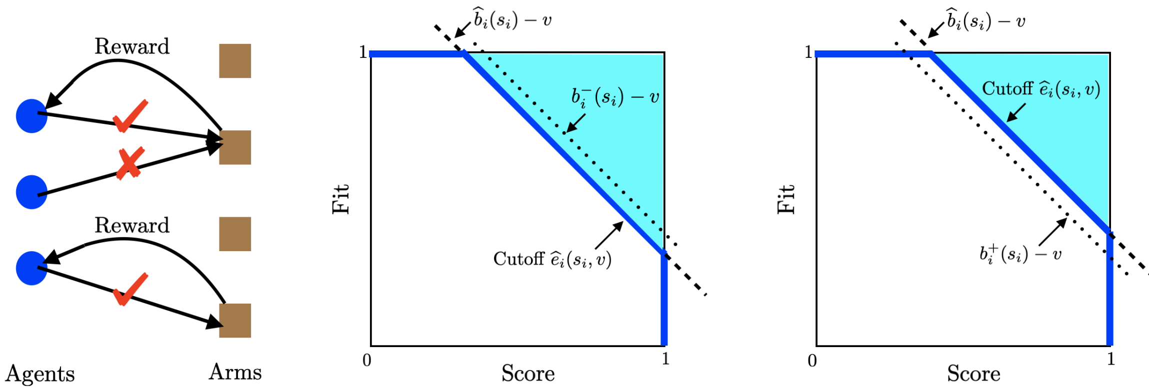

Figure 1: The left plot shows the process of single-stage matching in a decentralized market, with no centralized clearinghouse for coordination. The middle plot shows the cutoff $\widehat{e}_i(s_i, v)$ of Theorem 3 when Eq. (7) holds. The dotted line represents $b_i^-(s_i) - v$ in verifying Eq. (7). The dashed line represents $\widehat{b}_i(s_i) - v$, and if thresholding to $e_i \in [0, 1]$, it yields the cutoff $\widehat{e}_i(s_i, v)$, which is denoted by the blue solid line segments. The shaded area represents $\widehat{\mathcal{B}}_i(s_i)$. The right plot shows the cutoff $\widehat{e}_i(s_i, v)$ when Eq. (7) does not hold.

Theorem 3 *The expected payoff $\mathcal{U}_i[\mathcal{B}_i(s_i)]$ in Eq. (6) is maximized if and only if agent P_i uses a cutoff strategy with respect to arms' latent utilities, $\forall i \in [m]$. That is, P_i pulls arms from the set*

$$\widehat{\mathcal{B}}_i(s_i) = \{j \mid A_j \in \mathcal{A}^{T+1} \text{ whose attributes } (v_j, e_{ij}) \text{ satisfy } e_{ij} \geq \widehat{e}_i(s_i, v_j)\},$$

where the cutoff $\widehat{e}_i(s_i, v)$, chosen according to Eq. (8), is decreasing in $v \in [0, 1]$ and satisfies $d\widehat{e}_i(s_i, v)/dv = -1$ when $\widehat{e}_i(s_i, v) \in (0, 1)$.

We now specify the cutoff $\widehat{e}_i(s_i, v)$. Suppose that arms on the cutoff have the latent utility $b_i \geq 0$. The expected number of arms in $\widehat{\mathcal{B}}_i(s_i)$ that would accept P_i is:

$$\Pi_i(b_i) \equiv \sum_{j \in \mathcal{A}^{T+1}} \mathbf{1}\left(e_{ij} \geq \min\{\max\{b_i - v_j, 0\}, 1\}\right) \pi_i(s_i, v_j).$$

If there exists some $b_i \geq 0$ such that $\Pi_i(b_i) = q_i$, we let $\widehat{b}_i(s_i) = b_i$ and we have $\widehat{e}_i(s_i, v) = \min\{\max\{\widehat{b}_i(s_i) - v, 0\}, 1\}$. On the other hand, if there is no solution to $\Pi_i(b_i) = q_i$, we let

$$b_i^+(s_i) = \arg\max_{b_i \geq 0} \{\Pi_i(b_i) > q_i\} \quad \text{and} \quad b_i^-(s_i) = \arg\min_{b_i \geq 0} \{\Pi_i(b_i) < q_i\}.$$

To choose between $b_i^+(s_i)$ or $b_i^-(s_i)$, it is necessary to balance the expected utility and the expected penalty for exceeding the quota due to pulling arms on the *boundary*. Define two cutoffs $e_i^+(s_i, v) \equiv \min\{\max\{b_i^+(s_i) - v, 0\}, 1\}$ and $e_i^-(s_i, v) \equiv \min\{\max\{b_i^-(s_i) - v, 0\}, 1\}$, which correspond to arm sets $\mathcal{B}_i^+(s_i) = \{j \mid e_{ij} \geq e_i^+(s_i, v_j)\}$ and $\mathcal{B}_i^-(s_i) = \{j \mid e_{ij} \geq$

$e_i^-(s_i, v_j)\}$, respectively. Then the arms on the boundary are those in the set $\{\mathcal{B}_i^+(s_i) \setminus \mathcal{B}_i^-(s_i)\}$, whose expected utility is larger than the expected penalty if

$$\sum_{j \in \mathcal{B}_i^+(s_i) \setminus \mathcal{B}_i^-(s_i)} (v_j + e_{ij}) \cdot \pi_i(s_i, v_j) \geq \gamma_i \sum_{j \in \mathcal{B}_i^+(s_i)} \pi_i(s_i, v_j) - \gamma_i q_i. \tag{7}$$

If Eq. (7) holds, let $\widehat{b}_i(s_i) = b_i^+(s_i)$; otherwise, let $\widehat{b}_i(s_i) = b_i^-(s_i)$. Finally, we define the cutoff

$$\widehat{e}_i(s_i, v) = \min\{\max\{\widehat{b}_i(s_i) - v, 0\}, 1\}. \tag{8}$$

It is clear that $d\,\widehat{e}_i(s_i, v)/dv = -1$ when $\widehat{e}_i(s_i, v) \in (0, 1)$. Thus, P_i prefers the arms with larger latent utilities defined by Eq. (1). Figure 1 illustrates this cutoff strategy. Although arms *independently* accept or reject agents in decentralized markets, the cutoff $\widehat{e}_i(s_i, v)$ in Eq. (8) ensures that the expected number of arms accepting P_i excluding those on the boundary is bounded by the quota q_i. According to Eq. (7), the arm on the boundary is pulled if the expected utility is larger than the expected penalty.

The cutoff strategy in Theorem 3 relates to the straightforward behavior (see Fisman et al., 2006), where agents pull arms that they value more than those they do not pull. Theorem 3 shows that although agents face uncertainty with respect to acceptance of their offers in decentralized markets, the straightforward behavior suffices.

3.2 Calibrated Decentralized Matching (CDM)

Let s_i^* be the true state for agent P_i at time $T+1$. Theorem 3 shows that the cutoff strategy maximizes P_i's expected payoff if s_i^* is known. However, the true state s_i^* is generally unknown in practice. A natural question is how to calibrate the state s_i in Theorem 3. We propose a calibration method for s_i that maximizes the average-case expected payoff $\mathbb{E}_{s_i^*}\{\mathcal{U}_i[\widehat{\mathcal{B}}_i(s_i)]\}$ over the uncertain true state s_i^*. To formulate the theorem, we introduce some additional notation. Let $\partial\widehat{\mathcal{B}}_i(s_i)$ be the *marginal set*, defined as the change of set $\widehat{\mathcal{B}}_i(s_i)$ with respect to a perturbation of s_i:

$$\partial\widehat{\mathcal{B}}_i(s_i) \equiv \lim_{\delta s_i \to 0_+} \Big\{\widehat{\mathcal{B}}_i(s_i - \delta s_i) \setminus \widehat{\mathcal{B}}_i(s_i)\Big\}.$$

Theorem 4 *The average-case expected payoff, $\mathbb{E}_{s_i^*}\{\mathcal{U}_i[\widehat{\mathcal{B}}_i(s_i)]\}$, is maximized if $s_i \in (0, 1)$ is chosen as the solution to*

$$\begin{aligned} &\mathbb{P}(s_i^* \neq s_i) \sum_{j \in \partial\widehat{\mathcal{B}}_i(s_i)} (v_j + e_{ij}) \cdot \mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j) \mid s_i^* \neq s_i] \\ &\qquad = \gamma_i \left[1 - F_{s_i^*}(s_i)\right] \sum_{j \in \partial\widehat{\mathcal{B}}_i(s_i)} \mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j) \mid s_i < s_i^* \leq 1], \end{aligned} \tag{9}$$

where $F_{s_i^*}$ is the cumulative distribution function of $s_i^* \in [0, 1]$.

We refer to the calibration by Theorem 4 as the *mean calibration* since it maximizes the agent's average-case expected payoff. The key idea of the proof is to balance the tradeoff between the opportunity cost and the penalty for exceeding the quota. In particular, the left side of Eq. (9) considers the opportunity cost and the marginal utility, and the right side of Eq. (9) estimates the marginal penalty for exceeding the quota. We make four remarks on the theorem. First, the calibrated state in Theorem 4 is different from the naive mean estimate $\mathbb{E}[s_i^*]$ of the true state. The latter is inefficient in maximizing the agent's expected payoff in decentralized matching.

Second, the calibration in Theorem 4 takes agents' competition into account. Note that $\mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j)|s_i < s_i^* \le 1]$ in Eq. (9) is strictly increasing in s_i due to the monotonicity of $\pi_i(s_i, v_j)$ in s_i. Thus, it is more costly for an agent to pull an arm when the agent is popular. This result is intuitive because when an agent P_i is popular, it is more likely that an arm A_j pulled by multiple agents would accept P_i. Since P_i has a larger probability of exceeding the quota when P_i is popular, it is more costly for P_i to pull A_j compared to the case when P_i is not popular.

Third, the distribution $F_{s_i^*}$ is estimable from historical states $\{s_i^1, \dots, s_i^T\}$. For example, the kernel density method gives the estimate:

$$\widehat{f}_{s_i^*}(\cdot) = \frac{1}{T}\sum_{t=1}^{T} K_i^s\left(\frac{\cdot - s_i^t}{h}\right), \quad \text{and} \quad \widehat{F}_{s_i^*}(s_i) = \int_0^{s_i} \widehat{f}_{s_i^*}(s)ds. \tag{10}$$

Here, h is the bandwidth parameter, and K_i^s is the kernel introduced in Section 2.4. It is well-known that the estimate $\widehat{f}_{s_i^*}(\cdot)$ is rate-optimal (Silverman, 1986). Moreover, it is possible to incorporate side information into the estimation of $F_{s^*}(\cdot)$. For example, one can incorporate the belief that an agent tends to be popular at $T+1$ by overweighting the popular states. If Eq. (9) has more than one solution, then s_i is chosen as the largest one. If the distribution $F_{s_i^*}$ has discrete support, we change the objective in Theorem 4 to choosing the minimal $s_i \in [0,1]$ such that the left side of Eq. (9) is not less than the right side of Eq. (9). Here the search of s_i starts from the maximum value in the support and decreases to the minimal value.

Finally, we apply Eq. (5) to estimate the acceptance probability $\pi_i(s_i, v)$ in practice. This estimate is consistent and rate-optimal, according to Theorem 2. Besides the average-case expected payoff in Theorem 4, we also consider the worst-case expected payoff concerning s_i^*. In the following theorem, we propose a *maximin calibration*, where the calibration maximizes the minimal expected payoff $\min_{s_i^*}\{\mathcal{U}_i[\widehat{\mathcal{B}}_i(s_i)]\}$ over the uncertain true state s_i^*.

Theorem 5 *The worst-case expected payoff* $\min_{s_i^*}\{\mathcal{U}_i[\widehat{\mathcal{B}}_i(s_i)]\}$ *is maximized if* $s_i \in [0,1]$ *is chosen as the solution to*

$$\begin{aligned}\sum_{j\in\widehat{\mathcal{B}}_i(s_i)} (v_j + e_{ij})\cdot[\pi_i(1, v_j) - \pi_i(0, v_j)] - \gamma_i \sum_{j\in\widehat{\mathcal{B}}_i(s_i)} \pi_i(1, v_j) + \gamma_i q_i \\ = \sum_{j\in\widehat{\mathcal{B}}_i(1)} (v_j + e_{ij})\cdot\pi_i(1, v_j) - \sum_{j\in\widehat{\mathcal{B}}_i(0)} (v_j + e_{ij})\cdot\pi_i(0, v_j).\end{aligned}$$

Algorithm 1 summarizes the resulting algorithm, which we refer to as *calibrated decentralized matching* (CDM).

Algorithm 1 Calibrated decentralized matching (CDM)

1: **Input:** Historical data $\mathcal{D} = \{(s_i^t, v_j^t, e_{ij}^t, y_{ij}^t) : i \in [m]; j \in \mathcal{B}_i^t; t \in [T]\}$; New arm set $\mathcal{A}^{T+1}$ with attributes $\{(v_j, e_{ij}) : i \in [m]; j \in [n]\}$ at time $T+1$; Penalty $\{\gamma_i : i \in [m]\}$ for exceeding the quota.
2: **for** $i = 1, 2, \ldots, m$ **do**
3: Predict the acceptance $\widehat{\pi}_i(s_i, v_j)$ by Eq. (5);
4: Estimate the state distribution $F_{s_i^*}(\cdot)$ by the kernel density estimate in Eq. (10);
5: Calibrate the state s_i according to Theorem 4 or Theorem 5;
6: Calculate the cutoff strategy $\widehat{\mathcal{B}}_i(s_i)$ by Theorem 3.
7: **end for**
8: **Output:** The arm sets $\widehat{\mathcal{B}}_1(s_1), \widehat{\mathcal{B}}_2(s_2), \ldots, \widehat{\mathcal{B}}_m(s_m)$ for agents.

4. Properties of CDM: Incentives, Stability and Fairness

Each participant alone, on both sides of the market, knows their own preference. However, the decentralized market has no centralized system for eliciting the participants' preferences. The proposed CDM algorithm provides a mechanism to aggregate the historical data of the arms' choices. It also provides a framework for the learning of the agent strategies. This section addresses the question of whether CDM gives the agents incentives to act according to their true preferences. Such incentive-compatibility property for agents is desired for the design of matching markets (Roth, 1982).

A related problem to incentives is the stability of the matching outcomes that we obtain. We show that any pair of agent and arm has no incentive to disregard the CDM matching and seek an alternative outcome in the context of incomplete information in decentralized markets.

Finally, a different kind of problem concerning the matching procedure is fairness. For example, it is crucial to have a matching procedure that is fair for students in college admissions. We show that the CDM is fair for arms.

4.1 Incentives of Agents

We first need to define what is meant by a procedure giving agents incentives to act according to their true preferences in decentralized matching. The matching problem is a noncooperative game among the agents, whose payoffs are defined by the outcome of the matching process. A procedure that gives agents an incentive to act according to true preferences is defined as one that aggregates participants' actions so that, in the resulting noncooperative game, it is a dominant strategy for each agent to act on their preferences honestly. In such a procedure, no matter what strategies other players may play, an agent who deviates from the true preference can achieve no better outcome than if the agent had acted on their preferences honestly.

Theorem 6 *As $T \to \infty$, the CDM gives agents an incentive to act straightforwardly on their true preferences. That is, agents pull arms according to the arms' latent utilities in Eq. (1).*

Arms pulled by multiple agents self-select based on preference. The CDM algorithm provides a mechanism to aggregate the historical matches from arms' preference-based sorting in decentralized markets. Theorem 6 affirms an incentive-compatibility property for agents even with arbitrary uncertain preference profiles of arms. This theorem also distinguishes CDM from strategies according to arms' expected utilities in Eq. (2) (see Das and Kamenica, 2005).

4.2 Stability

An agent-arm pair (P_i, A_j) is referred to as *matched* if P_i pulls A_j, and A_j accepts P_i. A *blocking pair* is an agent-arm pair that is not matched, but both the agent and arm prefer to be matched together. Gale and Shapley (1962) defined a matching of arms to agents to be *stable* if there is no blocking pair. This definition was motivated by the concern that following a matching process, some agent-arm pairs may deviate together, thus hindering the implementation of the intended matching outcome. This form of stability is widely considered to be a key factor in a successful implementation of a centralized clearinghouse (Roth, 1990).

Gale and Shapley (1962) also proved that the set of stable matching is never empty. Among all stable matchings, the one that is the most preferred by all arms is called an *arm-optimal matching*. The one that is the most preferred by all agents is called an *agent-optimal matching*. The arm- (or agent-) optimal matching is also agents' (or arms') least-preferred matching among all stable matchings (Knuth, 1997).

The stability of a centralized matching relies on the stringent assumption of complete information on the preferences of the participants. In our decentralized matching formulation, it is more natural to assume that agents only have incomplete information on preferences. Thus we require a modification of the stability notion in Gale and Shapley (1962). In particular, it is necessary to specify the uninformed agents' beliefs that might block a candidate stable matching (cf. Liu et al., 2014). We formulate the following notion of *individual rationality* to specify the beliefs of uninformed agents. Let $\mathcal{N}_i$ be the expected number of arms that would accept an agent P_i under P_i's current strategy. An additional arm A_j with attributes (v_j, e_{ij}), which has a smaller latent utility than those currently pulled by P_i, is acceptable to P_i if and only if individual rationality holds:

$$(v_j + e_{ij}) \cdot \pi_i(s_i, v_j) \geq \gamma_i \cdot \max\left\{\mathcal{N}_i + \pi_i(s_i, v_j) - q_i, 0\right\}. \tag{11}$$

In other words, under P_i's current strategy, A_j is acceptable to P_i if and only if A_j's expected utility is at least the expected penalty for exceeding the quota.

A matching in a decentralized market is defined as stable if there is no blocking pair under the individual rationality condition (11). This stability notion is motivated by considering feasibility constraints (for example, single-stage interaction) in decentralized markets. In practice, participants are often compelled to accept the outcomes in such markets. For instance, the procedure in which some high school athletes are matched to colleges involves signing "letters of intent," which prohibits athletes from further negotiating with other colleges. Suppose that an arm, A_j, does not satisfy Eq. (11). Then P_i finds that A_j only has a small latent utility but pulling A_j may incur a large penalty due to exceeding the quota. It is individually rational that P_i does not pull A_j in order to receive at least zero payoff. After the matching procedure, suppose P_i finds that A_j is acceptable. That is, P_i has remaining capacity, and A_j prefers to match to P_i. Still, P_i has no incentive to disregard the matching and seek an alternative outcome due to the feasibility constraints of decentralized markets.

Theorem 7 *As $T \to \infty$, the CDM procedure yields a stable matching. That is, agent and arm has no incentive to disregard the CDM matching and seek an alternative outcome in decentralized markets.*

There exists a lattice structure of stable matchings yielded by the CDM algorithm. If CDM allows each agent to pull all arms, the outcome is the arm-optimal stable matching. If CDM allows each agent to pull arms up to their quota, and those pulled arms are distinct, the outcome is the agent-optimal stable matching (Roth, 2008). In practice, the stable matching yielded by CDM is generally between the agent-optimal and arm-optimal matchings due to the competition of agents and a lack of coordination. Indeed, agents are incentivized to pull more arms than their quotas in decentralized markets in order to hedge against the uncertainty of whether the pulled arms will accept them. On the other hand, arms may benefit from this uncertainty. For example, suppose the college admissions market uses a clearinghouse to form a centralized market. All students apply to top colleges. In that case, the deferred acceptance algorithm yields the college-optimal matching (Gale and Shapley, 1962; Knuth, 1997). Compared to the CDM outcomes, students are worse off in the centralized market.

In contrast to CDM, the strategies of pulling arms according to the expected utilities in Eq. (2) may result in an unstable matching. For instance, if the agent P_i decides according to A_j's expected utility, $\pi_i(s_i, v_j)(v_j + e_{ij})$, then P_i will not pull A_j if the acceptance probability $\pi_i(s, v_j)$ is small enough such that the expected utility of A_j is the smallest among all arms, while the latent utility $(v_j + e_{ij})$ of A_j is the largest. As a result, (P_i, A_j) will form a blocking pair if A_j is unmatched, and the matching is unstable.

4.3 Fairness

We consider fairness in terms of "no justified envy" (Balinski and Sönmez, 1999; Abdulkadiroğlu and Sönmez, 2003). Here an arm A_j has justified envy if A_j prefers an agent $P_{i'}$ to another agent P_i that pulls A_j, even though $P_{i'}$ pulls an arm $A_{j'}$ which ranks below A_j according to the true preference of $P_{i'}$. A matching procedure in decentralized markets is fair if no arm has justified envy.

Theorem 8 *As $T \to \infty$, the CDM is fair for arms. That is, no arm has justified envy in the matching outcome realized by CDM.*

We now compare CDM with the oracle arm set, where the latter maximizes agents' average-case expected payoff by assuming complete information on opponent agents' strategies. For any arm set $\mathcal{B}_i \subseteq \mathcal{A}^{T+1}$, let $O_{\mathcal{B}_i}$ be the set of true states under which agent P_i's strategy of pulling $\mathcal{B}_i$ results in the over-enrollment:

$$O_{\mathcal{B}_i} \equiv \left\{ s_i^* \in [0,1] \;\middle|\; \sum_{j \in \mathcal{B}_i} \pi_i(s_i^*, v_j) > q_i \right\}, \quad \forall i \in [m].$$

Under the true state s_i^*, agent P_i's expected payoff of pulling $\mathcal{B}_i$ is

$$\mathcal{U}_i[\mathcal{B}_i] = \sum_{j\in\mathcal{B}_i} (v_j + e_{ij}) \cdot \pi_i\left(s_i^*, v_j\right) - \gamma_i \cdot \max\left\{\sum_{j\in\mathcal{B}_i} \pi_i(s_i^*, v_j) - q_i,\ 0\right\}.$$

Then P_i's average-case expected payoff $\mathbb{E}_{s_i^*}\{\mathcal{U}_i[\mathcal{B}_i]\}$ is

$$\sum_{j\in\mathcal{B}_i} (v_j + e_{ij}) \cdot \mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j)] - \gamma_i \cdot \mathbb{P}\left(s_i^* \in O_{\mathcal{B}_i}\right) \left\{\sum_{j\in\mathcal{B}_i} \mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j) \mid s_i^* \in O_{\mathcal{B}_i}] - q_i\right\}.$$

Hence, the oracle arm set $\mathcal{B}_i^*$ for maximizing $\mathbb{E}_{s_i^*}\{\mathcal{U}_i[\mathcal{B}_i]\}$ becomes

$$\mathcal{B}_i^* = \left\{ j \in \mathcal{A}^{T+1} \;\middle|\; v_j + e_{ij} \geq \gamma_i \cdot \mathbb{P}(s_i^* \in O_{\mathcal{B}_i^*}) \frac{\mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j) \mid s_i^* \in O_{\mathcal{B}_i^*}]}{\mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j)]} \right\}. \tag{12}$$

Che and Koh (2016) discussed a particular case of the oracle arm set in a two-agent model. Eq. (12) generalizes the oracle set to multiple agents. Although $\mathcal{B}_i^*$ has a closed-form expression in Eq. (12), $\mathcal{B}_i^*$ is not estimable from training data. The reason is that the probability $\mathbb{P}(s_i^* \in O_{\mathcal{B}_i^*})$ requires the knowledge of opponent agents' strategies, which are unknown in decentralized markets. Moreover, we show that the matching procedure according to the strategy of P_i pulling arms from $\mathcal{B}_i^*$ will be unfair for arms.

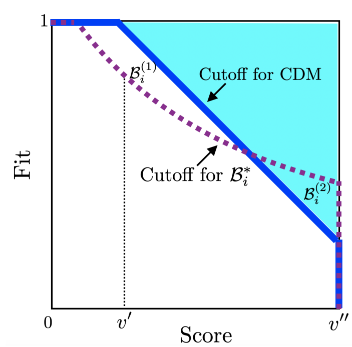

Figure 2: Fairness of CDM compared to the unfairness of oracle arm set $\mathcal{B}_i^*$. The dotted curve represents the cutoff for $\mathcal{B}_i^*$ in Eq. (12). The solid line segments denote the cutoff for CDM. The arms in $\mathcal{B}_i^{(2)}$ have justified envy towards arms in $\mathcal{B}_i^{(1)}$.

Theorem 9 *The strategy corresponding to the oracle set in Eq. (12) is unfair if for at least one agent P_i, there exists an interval $(v', v'') \subset [0, 1]$ such that*

$$\mathbb{E}\left[\frac{\partial \pi_i(s_i^*, v)}{\partial v} \;\middle|\; s_i^* \notin O_{\mathcal{B}_i^*}\right] < \mathbb{E}\left[\frac{\partial \pi_i(s_i^*, v)}{\partial v} \;\middle|\; s_i^* \in O_{\mathcal{B}_i^*}\right] < 0, \quad \forall v \in (v', v''). \tag{13}$$

(a) Latent utility

	A_1	A_2	A_3
P_1	2	3	2.5
P_2	2	2.5	3
P_3	2.5	2	3

(b) Arm's preference

	P_1	P_2	P_3
A_1	3	2	1
A_2	2	3	1
A_3	1	3	2

(c) Expected utility

	A_1	A_2	A_3
P_1	0.52	1.99	2.5
P_2	0.67	0	0
P_3	2.5	2	1.05

Table 1: (a) Arm's latent utilities for each agent, which corresponds to Eq. (1). For example, P_1 receives utility 2.5 when it successfully pulls A_3. (b) Arms' preferences with the number indicating the arms' ranking of agents. For example, A_1 ranks P_3 first, P_2 second, P_1 third. These preferences are unknown to agents. (c) Expected utilities, which corresponds to Eq. (2). For example, P_1 expects to receive utility 0.52 if it pulls A_1.

Here, Eq. (13) means that the probability of arms with score $v \in (v', v'')$ accepting P_i decreases less when P_i is popular ($s_i^* \in O_{\mathcal{B}_i^*}$) than that when P_i is not popular ($s_i^* \notin O_{\mathcal{B}_i^*}$). This condition holds for many decentralized matching examples, such as the two-agent example in Section 2.3.

Figure 2 illustrates the unfairness of the strategy corresponding to the oracle set. The proof of Theorem 9 shows that the slope of $\mathcal{B}_i^*$'s cutoff curve is in the interval $(-1, 0)$, for any $v \in (v', v'')$. Hence, there are arms not selected to $\mathcal{B}_i^*$ (for example, those in $\mathcal{B}_i^{(2)}$). However, they rank higher than some selected arms (for example, those in $\mathcal{B}_i^{(1)}$) according to agent P_i's true preference. Thus, arms in $\mathcal{B}_i^{(2)}$ have justified envy towards arms in $\mathcal{B}_i^{(1)}$. On the contrary, Theorem 3 shows that the slope of CDM's cutoff equals -1 and hence the agent prefers arms with larger latent utilities.

5. Numerical Experiments

In this section, we provide a numerical investigation of the fairness and stability properties of CDM. We also study the payoffs achieved by CDM compared to alternative methods.

5.1 Stability and Fairness of CDM

Suppose that there are three agents, P_1, P_2, P_3, and three arms, A_1, A_2, A_3. The latent utilities and the arms' true preferences are given in Table 1. Arms have scores and fits as follows: $v_1 = v_2 = v_3 = 2$, and $e_{11} = e_{21} = e_{32} = 0, e_{13} = e_{22} = e_{31} = 0.5, e_{12} = e_{23} = e_{33} = 1$, respectively. Each agent has quota $q = 1$ and the penalty for exceeding the quota is $\gamma = 10$. Agents have to make decisions on which arms to pull without knowing the arms' true preferences. We compare the CDM procedure with the *greedy action*, which chooses arms with maximum expected utilities and total expected acceptance up to the quota (Das and Kamenica, 2005).

The training data are simulated by having each agent pull a random number of arms according to its latent utilities. Figure 3 shows arms' acceptance probabilities $\pi_i(s_i^*, v_j)$ based

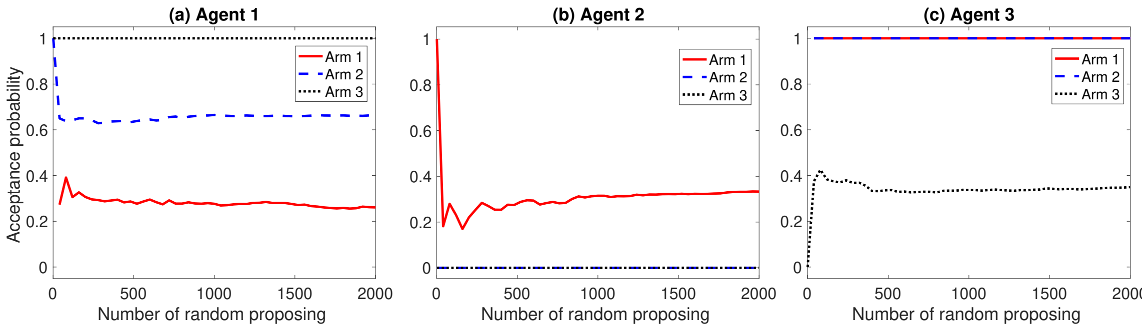

Figure 3: Arms' acceptance probabilities for three agents. (a) P_1, (b) P_2, (c) P_3.

on a total of 2000 rounds of random proposals. Since there is a unique true state, the estimates of acceptance probabilities converge. The CDM procedure suggests that P_1 pulls A_2, P_2 pulls A_1, A_2, A_3, and P_3 pulls A_3, and it gives the stable matching $(A_1, P_2), (A_2, P_1), (A_3, P_3)$. On the contrary, the greedy action suggests that P_1 pulls A_3, P_2 pulls A_1, A_2, A_3, and P_3 pulls A_1, and it yields the following matching: $(A_1, P_3), (A_2, P_2), (A_3, P_1)$. We note three differences between the two matchings. First, the greedy action is *unfair* since A_2 has justified envy towards A_3 in the sense that A_2 prefers P_1 to P_2. However, P_1 pulls A_3 that ranks below A_2 according to the true preference of P_1. Second, the greedy action also yields an *unstable* matching since (A_2, P_1) is a blocking pair. On the other hand, CDM is fair and yields a stable matching. The stable matching is both agent-optimal and arm-optimal in this example. Finally, the total payoff that agents receive using CDM equals $3 + 2 + 3 = 8$, which is larger than the total payoff that agents receive from the greedy action, $2.5 + 2.5 + 2.5 = 7.5$.

5.2 Lattice Structure for the Stability of CDM

We consider the decentralized matching with four different preference structures: S1, S2, S3, S4. The market consists of three agents and three arms. Each agent has the same quota $q = 1$ and penalty $\gamma = 5$. Table 2 gives arms' latent utilities and true preferences. The training data are simulated by having agents pull random numbers of arms according to their latent utilities. The last column of Table 2 gives estimates of arms' acceptance probabilities, $\pi_i(s_i^*, v_j)$, after 2000 rounds of random proposals and under each of the structures S1—S4. These acceptance probabilities are evaluated at convergence.

We find that CDM gives stable matchings under all of the structures S1—S4. However, these stable matchings have different optimality in terms of agents' and arms' welfare. In particular, the matching in S1 is $(A_1, P_1), (A_2, P_3), (A_3, P_2)$, which is both agent-optimal and arm-optimal; the matching in S2 is $(A_1, P_2), (A_2, P_1), (A_3, P_3)$, which is arm-optimal but not agent-optimal; the matching in S3 is $(A_1, P_2), (A_2, P_3)$, (A_3, P_1), which is not agent-optimal or arm-optimal; the matching in S4 is (A_1, P_1), (A_2, P_2), (A_3, P_1), which is not agent-optimal or arm-optimal. This lattice structure corroborates the results in Section 4.2. We further make three remarks on these matchings. First, some arms benefit from the decentralized matching compared with the centralized matching using the agent-proposing DA algorithm.

Latent utility of S1

	A_1	A_2	A_3
P_1	2.5	3	2
P_2	3	2.5	2
P_3	3	2.5	2

Arm's preference of S1

	P_1	P_2	P_3
A_1	1	2	3
A_2	2	3	1
A_3	1	2	3

Acceptance probability of S1

	A_1	A_2	A_3
P_1	1	0.34	1
P_2	0.35	0	0.65
P_3	0	1	0.45

Latent utility of S2

	A_1	A_2	A_3
P_1	3	2.5	2
P_2	2.5	3	2
P_3	2.5	2	3

Arm's preference of S2

	P_1	P_2	P_3
A_1	3	1	2
A_2	1	2	3
A_3	2	3	1

Acceptance probability of S2

	A_1	A_2	A_3
P_1	0.10	1	0
P_2	1	0.35	0
P_3	0.32	0	1

Latent utility of S3

	A_1	A_2	A_3
P_1	2	3	2.5
P_2	2.5	2	3
P_3	3	2.5	2

Arm's preference of S3

	P_1	P_2	P_3
A_1	1	2	3
A_2	3	1	2
A_3	2	3	1

Acceptance probability of S3

	A_1	A_2	A_3
P_1	1	0.22	0.67
P_2	0.66	1	0.24
P_3	0.22	0.66	1

Latent utility of S4

	A_1	A_2	A_3
P_1	3	2	2.5
P_2	2	2.5	3
P_3	2	2.5	3

Arm's preference of S4

	P_1	P_2	P_3
A_1	3	2	1
A_2	2	1	3
A_3	1	3	2

Acceptance probability of S4

	A_1	A_2	A_3
P_1	0.43	0.29	1
P_2	0.69	1	0
P_3	1	0.22	0.35

Table 2: The left column shows arms' latent utilities for each agent. The middle column shows arms' preferences, where the number indicates the arms' ranking of agents. The right column shows arms' acceptance probabilities.

For example, the CDM matching in S2 is arm-optimal. The agent-proposing DA would give the agent-optimal matching, that is, the arms' least-preferred stable matching.

Second, no strategy in decentralized markets guarantees yielding the agent-optimal matching, mainly due to the competition of agents and a lack of coordination. Consider the S3 structure as an example, where CDM suggests that P_1 pulls A_1, A_2, P_2 pulls A_2, A_3 and P_3 pulls A_1, A_3. In this example, the strategy, according to the CDM, is a subgame perfect equilibrium since each agent's action is the best response against other agents' actions. For instance, if P_1 changes to the strategy by only pulling A_2 while other agents' strategies do not change, this gives the following matching: (A_1, P_2), (A_2, P_3), (A_3, P_2), where P_1 is worse off due to the unfilled quota. If P_1 changes to pull all arms A_1, A_2, A_3 while other agents' strategies do not change, then the resulting matching is (A_1, P_1), (A_2, P_3), (A_3, P_1), where P_1 is also worse off due to exceeding the quota. A similar argument applies to P_2 and P_3. Hence, no agent has an incentive to change its strategy. In contrast, the agent-proposing DA permits each agent to pull one arm each time according to its latent utility. Such coordination results in the agent-optimal matching. However, if there is no coordination, the competition of agents generates uncertainty on the acceptance of the pulled arms. Hence, agents have the incentive to pull more arms than their quotas to combat the uncertainty.

Third, we discuss individual rationality in Eq. (11). Without this condition, any stable matching can be supported by a subgame perfect equilibrium, and only stable matchings can arise in equilibrium (Alcalde and Romero-Medina, 2000). However, under Eq. (11), the set of stable matchings is enlarged. It includes (but may not coincide with) the set of subgame perfect equilibria. Consider S4 as an example, where CDM suggests that P_1 pulls A_1, A_3, and P_2 pulls A_2, A_3, and P_3 pulls A_2, A_3. The strategy given by CDM is not a subgame perfect equilibrium. However, the resulting matching is stable under Eq. (11). For instance, P_3 finds A_1 having a larger expected penalty for exceeding the quota than its expected utility. That is, P_3 finds A_1 unacceptable. Hence, there is no blocking pair.

5.3 Agents' Payoffs Achieved by CDM

Consider ten agents and varying numbers of arms: $\{50, 70, 90, 110, 130, 150\}$. Each arm has a score v_j and fits e_{ij} drawn uniformly from $[0, 1]$. The agents' preferences are determined by arms' latent utilities as in Eq. (1). Each agent has the same quota $q = 5$ and the same penalty γ chosen from $\{2, 2.5, 3\}$. The simulation generates random arm preferences with 10 different states from $\{s_1, \ldots, s_{10}\} \subset [0, 1]$. The training data are simulated by having agents pull random numbers of arms according to their latent utilities. We train 20 times under each of the arms' preference structures. This training data simulates the history of 20 experiments. The testing data draws a random state from $\{s_1, \ldots, s_{10}\}$ and generates the corresponding arms' preferences. This example simulates top colleges competing for top students. Students' preferences are uncertain and depend on colleges' reputation and popularity in the current year.

We compare the agent's expected payoff achieved by CDM with that of other methods. In particular, we also consider (i) the *simple cutoff strategy* where the agent chooses q best arms; (ii) the *greedy action* where the agent chooses arms with maximum expected utilities and a total expected acceptance up to q. Figure 4 reports the agent P_1's averaged payoffs over 500 data replications. Here, all agents other than P_1 use the CDM with mean calibration and P_1 uses one of the three methods. It is seen that CDM gives the largest average payoffs compared to alternative methods, and the advantage of CDM is robust to different numbers of arms and penalty levels. We make two further remarks. First, the state's calibration is useful in improving the agent's expected utility under uncertain preferences of the arms. For example, CDM outperforms the simple cutoff strategy that has no calibration on the uncertain state. Second, CDM performs particularly well if the matching market has intense competition. In this simulated market where arms' preferences are random, a smaller number of arms corresponds to a higher competition level. It is seen that CDM is significantly better than other methods in the regime of small numbers of students. On the other hand, the simple cutoff strategy does not work well with small

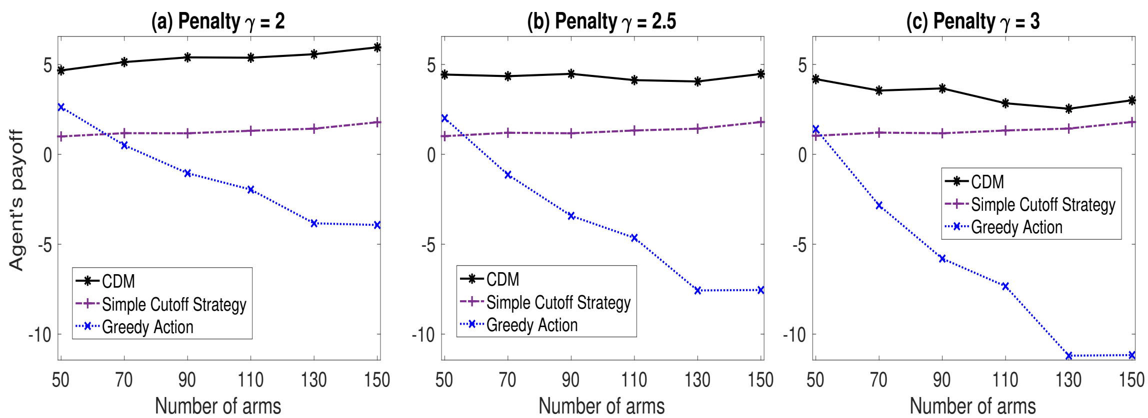

Figure 4: Performance of three strategies with varying numbers of arms. The results are averaged over 500 data replications. Penalty levels (a) $\gamma = 2$, (b) $\gamma = 2.5$, (c) $\gamma = 3$.

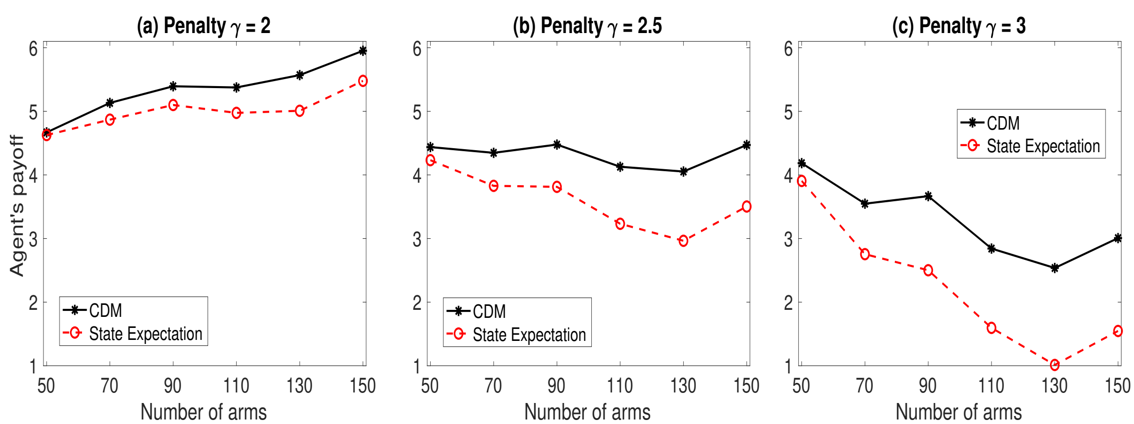

Figure 5: Performance of two calibration methods with varying numbers of arms. The results are averaged over 500 data replications. Penalty levels (a) $\gamma = 2$, (b) $\gamma = 2.5$, (c) $\gamma = 3$.

numbers of students. The reason is that under an intense competition among the agents, arms reject most offers sent according to the simple cutoff strategy.

We also compare two different calibration methods: CDM and *state expectation*. The latter calibrates the unknown state using the naive mean estimate of states. The state expectation method was discussed in Section 3.2. Figure 5 shows P_1's averaged payoffs over 500 data replications. Here, all agents other than P_1 use the CDM with mean calibration and P_1 uses one of the two methods. We observe that CDM is adaptive to different levels of penalty γ. However, the state expectation method degrades quickly as γ increases. This difference between the two methods is because CDM balances the marginal payoff and the marginal penalty for exceeding the quota. Hence CDM is sensitive to the penalty γ, while the state expectation method is not.

5.4 Simulated Graduate School Admission

This section provides a simulation of graduate school admissions, where the programs of graduate schools have limited quotas. Suppose there are a total of 50 graduate schools from three tiers of colleges: five top colleges $\{P_1, \ldots, P_5\}$, ten good colleges $\{P_6, \ldots, P_{15}\}$, and 35 other colleges $\{P_{16}, \ldots, P_{50}\}$. Each has the same quota $q = 5$ and penalty $\gamma = 2.5$. The simulation generates students' preferences with ten different states $\{s_1, \ldots, s_{10}\} \subset [0, 1]$. For any state, students' preferences for colleges from the same tier are random. However, students always prefer top colleges to good colleges, and then the other colleges. The random preferences depend on the state due to colleges' uncertain reputation and popularity in the current year. We consider varying numbers of students, ranging over $\{250, 260, 270, 280, 290, 300\}$. For each number of students, there are ten students having a score v_j chosen uniformly and independently from $[0.9, 1]$ and 100 students having score v_j drawn uniformly and independently from from $[0.7, 0.9)$. The rest of the students have score v_j chosen uniformly and independently from $[0, 0.7)$. The fits e_{ij} for all college-student pairs are drawn uniformly from $[0, 1]$. The colleges' preferences are determined by students' latent utilities according to Eq. (1). We consider a simple case that students face negligible application costs. Since students do not know how colleges evaluate their fits (e.g., personal essays), submitting applications to all colleges is students' dominant strategy.

We compare the colleges' expected payoffs achieved by three methods: CDM, simple cutoff strategy, and greedy action. The latter two methods are described in Section 5.3. The training data are simulated from colleges' random proposing, where colleges admit random numbers of students according to their latent utilities. We train 20 times under each of the students' preference structures. This training data simulates admissions over 20 years. The testing data draws a random state from $\{s_1, \ldots, s_{10}\}$, which gives the corresponding student preferences. The CDM procedure estimates the acceptance probability using Eq. (5). Figure 6 reports the averaged payoffs of three colleges P_1, P_6 and P_{16} over 500 data replications. Here colleges P_1, P_6 and P_{16} belong to the three different tiers, respectively. In Figure 6(a1) and (a2), all colleges other than P_1 use the CDM with mean calibration, and P_1 uses one of the three methods. The same setup applies to Figure 6(b1) and (b2) and Figure 6(c1) and (c2). It is seen that the CDM gives the largest average payoffs for all of P_1, P_6 and P_{16}. CDM performs particularly well for tier 2 and tier 3 colleges compared to the simple cutoff strategy. Moreover, CDM outperforms the greedy action, especially for tier 1 and 2 colleges. The reason is that greedy action does not calibrate the state which results in enrolling too many students.

6. Discussion

This paper devises a statistical model to estimate uncertain preferences and manage competition in decentralized matching markets. The proposed model serves as a first step towards bridging the statistical machine learning literature and the microeconomic objectives of de-

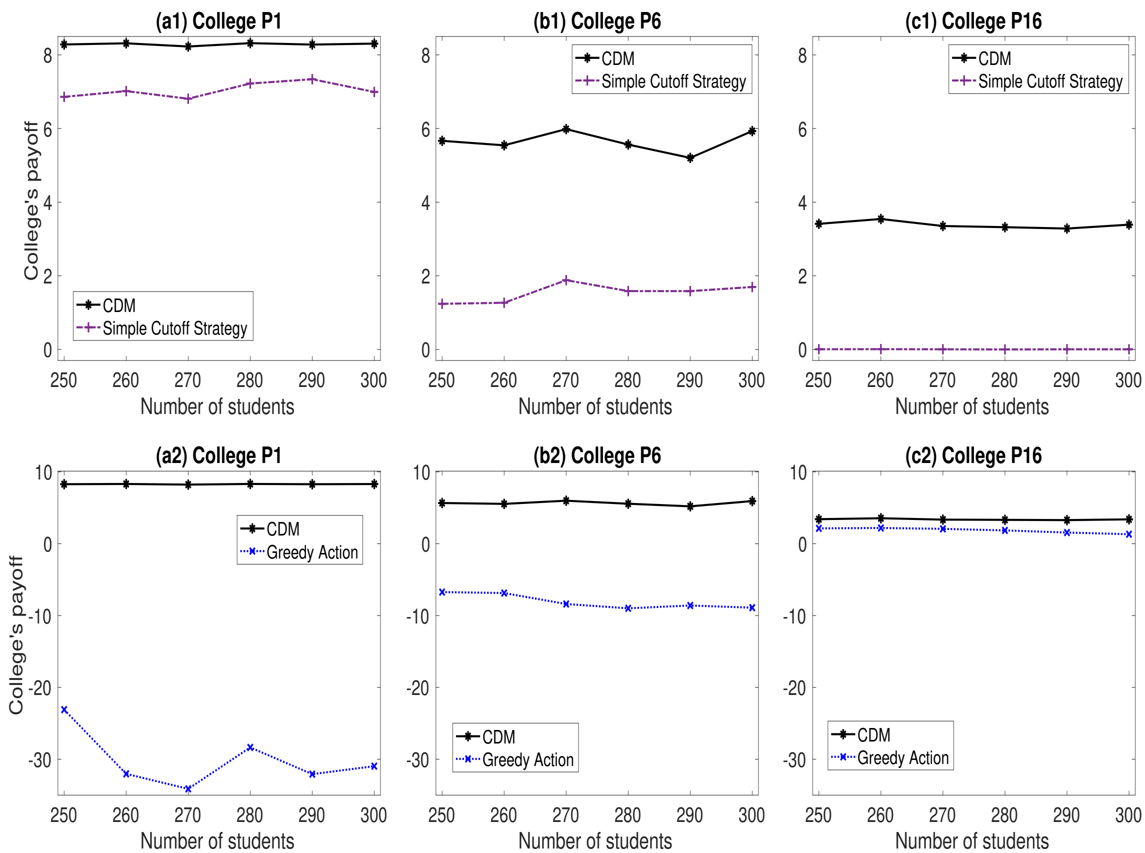

Figure 6: Performance of three strategies with varying numbers of students. The results are averaged over 500 data replications. (a1) and (a2): College P_1 from tier 1. (b1) and (b2): College P_6 from tier 2. (c1) and (c2): College P_{16} from tier 3.

centralized matching. In the model, arms have uncertain preferences that depend on the unknown state of the world. The arms' acceptance probabilities also depend on agents' competition. We propose an optimal strategy called calibrated decentralized matching (CDM) that maximizes agents' expected payoffs. Various statistical learning algorithms allow efficient learning of the acceptance probability under the proposed model. The CDM procedure calibrates the unknown state by perturbing the state and balancing the marginal utility and the marginal penalty for exceeding the capacity. This calibration procedure takes the opportunity cost into account. We show that CDM makes it safe for agents to act straightforwardly on their preferences. The CDM procedure achieves stability under incomplete information, where the formulation of the individual rationality condition models agents' beliefs. Moreover, CDM is fair for arms in the sense that arms have no justified envy.

It is possible to extend the CDM to multi-stage decentralized matching applications such as college admissions with a waiting list, and we are currently working on that extension. Another possible extension is to consider algorithmic strategies where agents' preferences exhibit complementarities in decentralized matching. For instance, some firms demand workers that complement one another in terms of their skills and roles. Another interesting extension is to consider decentralized markets with indifferent preferences. For example, many applicants may be indistinguishable for a college. Still, it is necessary to break ties since the college may have insufficient capacity to admit all applicants in the same indifference class.

References

Atila Abdulkadiroğlu and Tayfun Sönmez. School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, 2003.

José Alcalde and Antonio Romero-Medina. Simple mechanisms to implement the core of college admissions problems. *Games and Economic Behavior*, 31(2):294–302, 2000.

Itai Ashlagi, Mark Braverman, Yash Kanoria, and Peng Shi. Clearing matching markets efficiently: Informative signals and match recommendations. *Management Science*, 66 (5):2163–2193, 2020.

Christopher Avery and Jonathan Levin. Early admissions at selective colleges. *American Economic Review*, 100(5):2125–56, 2010.

Christopher Avery, Andrew Fairbanks, and Richard J Zeckhauser. *The Early Admissions Game: Joining the Elite*. Harvard University Press, Cambridge, MA, 2003.

Eduardo M Azevedo and Jacob D Leshno. A supply and demand framework for two-sided matching markets. *Journal of Political Economy*, 124(5):1235–1268, 2016.

Michel Balinski and Tayfun Sönmez. A tale of two mechanisms: Student placement. *Journal of Economic Theory*, 84(1):73–94, 1999.

Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

Hector Chade and Lones Smith. Simultaneous search. *Econometrica*, 74(5):1293–1307, 2006.

Hector Chade, Gregory Lewis, and Lones Smith. Student portfolios and the college admissions problem. *Review of Economic Studies*, 81(3):971–1002, 2014.

Yeon-Koo Che and Youngwoo Koh. Decentralized college admissions. *Journal of Political Economy*, 124(5):1295–1338, 2016.

Pierre-André Chiappori and Bernard Salanié. The econometrics of matching models. *Journal of Economic Literature*, 54(3):832–61, 2016.

Eugene Choo and Aloysius Siow. Who marries whom and why. *Journal of Political Economy*, 114(1):175–201, 2006.

Peter Coles, Alexey Kushnir, and Muriel Niederle. Preference signaling in matching markets. *American Economic Journal: Microeconomics*, 5(2):99–134, 2013.

Sanmay Das and Emir Kamenica. Two-sided bandits and the dating market. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, volume 5, page 19, 2005.

Dennis Epple, Richard Romano, and Holger Sieg. Admission, tuition, and financial aid policies in the market for higher education. *Econometrica*, 74(4):885–928, 2006.

Raymond Fisman, Sheena S Iyengar, Emir Kamenica, and Itamar Simonson. Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics*, 121(2):673–697, 2006.

Chao Fu. Equilibrium tuition, applications, admissions, and enrollment in the college market. *Journal of Political Economy*, 122(2):225–281, 2014.

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

Guillaume Haeringer and Myrna Wooders. Decentralized job matching. *International Journal of Game Theory*, 40(1):1–28, 2011.

Isa E Hafalir, Rustamdjan Hakimov, Dorothea Kübler, and Morimitsu Kurino. College admissions with entrance exams: Centralized versus decentralized. *Journal of Economic Theory*, 176:886–934, 2018.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, 2009.

George Kimeldorf and Grace Wahba. Some results on tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, 33(1):82–95, 1971.

Donald Ervin Knuth. *Stable Marriage and Its Relation to Other Combinatorial Problems: An Introduction to the Mathematical Analysis of Algorithms*, volume 10. CRM Proceedings and Lecture Notes. American Mathematical Society., 1997.

SangMok Lee. Incentive compatibility of large centralized matching markets. *The Review of Economic Studies*, 84(1):444–463, 2016.

Yi Lin. *Tensor Product Space ANOVA Models in Multivariate Function Estimation*. PhD thesis, University of Pennsylvania, Philadelphia, Pennsylvania, 1998.

Lydia T Liu, Horia Mania, and Michael I Jordan. Competing bandits in matching markets. In *International Conference on Artificial Intelligence and Statistics*, pages 1618–1628. PMLR, 2020.

Qingmin Liu, George J Mailath, Andrew Postlewaite, and Larry Samuelson. Stable matching with incomplete information. *Econometrica*, 82(2):541–587, 2014.

Konrad Menzel. Large matching markets as two-sided demand systems. *Econometrica*, 83 (3):897–941, 2015.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems*, pages 1177–1184, 2008.

Alvin E Roth. The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7(4):617–628, 1982.

Alvin E Roth. The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, 92(6):991–1016, 1984.

Alvin E Roth. New physicians: a natural experiment in market organization. *Science*, 250 (4987):1524–1528, 1990.

Alvin E Roth. Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory*, 36(3-4):537–569, 2008.

Alvin E Roth and Marilda Sotomayor. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, volume 18. Econometric Society Monographs. Cambridge University Press, Cambridge, 1990.

Alvin E Roth and Xiaolin Xing. Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists. *Journal of Political Economy*, 105(2):284–329, 1997.

Alessandro Rudi and Lorenzo Rosasco. Generalization properties of learning with random features. In *Advances in Neural Information Processing Systems*, pages 3215–3225, 2017.

Leonard J Savage. *The Foundations of Statistics.* Courier Corporation, MA, 1972.

Bernard W Silverman. *Density Estimation for Statistics and Data Analysis*, volume 26. Chapman and Hall, London, 1986.

Grace Wahba. Support vector machines, reproducing kernel hilbert spaces and the randomized gacv. *Advances in Kernel Methods-Support Vector Learning*, pages 69–87, 1999.

Grace Wahba, Yuedong Wang, Chong Gu, Ronald Klein, and Barbaras Klein. Smoothing spline anova for exponential families, with application to the wisconsin epidemiological study of diabetic retinopathy. *Annals of Statistics*, 23(6):1865–1895, 1995.

Appendix A. Proofs

A.1 ANOVA Decomposition for General Utility Function

Suppose that each arm $A_j \in \mathcal{A}$ is described by x_j and $(\epsilon_{1j}^\intercal, \epsilon_{2j}^\intercal, \dots, \epsilon_{mj}^\intercal)^\intercal$, where x_j is a multidimensional vector available to all agents and ϵ_{ij} is a multidimensional vector that is only available to agent P_i. In college admissions, x_j can be student A_j's high school record and test score on a nationwide exam such as SAT/ACT, and ϵ_{ij} represents student A_j's college-specific essays or test scores.

Proposition 10 *For any agent-specific latent utility function U_i, the ANOVA decomposition allows the separable form*

$$U_i(A_j) = v_j + e_{ij}, \quad \forall i \in [m] \text{ and } j \in [n].$$

Here, $v_j \in \mathbb{R}$ is a function of x_j and is common to all agents. The agent-specific $e_{ij} \in \mathbb{R}$ is a function of x_j and ϵ_{ij} and is considered only by agent P_i. Thus, the separable structure of the utility function in Eq. (1) can be assumed without loss of generality.

Proof Denote the utility function $U_i(A_j) \equiv U_i(x_j, \epsilon_{ij}) \in \mathbb{R}$. By the analysis of variance (ANOVA) decomposition, we have that for $P_i \in \mathcal{P}, A_j \in \mathcal{A}$,

$$\begin{aligned} & v_j \equiv \frac{1}{m}\sum_{i=1}^m \mathbb{E}_{\epsilon_{ij}}[U_i(x_j, \epsilon_{ij})] \quad \text{and} \\ & e_{ij}^\dagger \equiv \mathbb{E}_{\epsilon_{ij}}[U_i(x_j, \epsilon_{ij})] - \frac{1}{m}\sum_{i=1}^m \mathbb{E}_{\epsilon_{ij}}[U_i(x_j, \epsilon_{ij})], \quad e_{ij}^\ddagger \equiv U_i(x_j, \epsilon_{ij}) - \mathbb{E}_{\epsilon_{ij}}[U_i(x_j, \epsilon_{ij})]. \end{aligned} \tag{14}$$

Here, v_j represents the average utility of x_j and is common to all agents. The $e_{ij}^\dagger$ is agent P_i's adjustment for the utility of x_j. The $e_{ij}^\ddagger$ is the utility of ϵ_{ij} received by agent P_i. Thus, $e_{ij}^\dagger$ and $e_{ij}^\ddagger$ are agent-specific and they are only known to agent P_i. Moreover, letting $e_{ij} \equiv e_{ij}^\dagger + e_{ij}^\ddagger$, then Eq. (14) implies the desired result that $U_i(x_j, \epsilon_{ij}) = v_j + e_{ij}$. We refer to Figure 7 for an illustration on the ANOVA decomposition.

The utility function $U_i(x_j, \epsilon_{ij})$ is generally assumed to be strictly increasing in x_j and nondecreasing in ϵ_{ij} (Che and Koh, 2016; Lee, 2016). Then v_j in Eq. (14) is strictly increasing in x_j, and e_{ij} is nondecreasing in ϵ_{ij}, for any $i \in [m]$. ■

We demonstrate the ANOVA decomposition in Eq. (14) through the college admission example. The score v_j in Eq. (14) represents the "public valuation" of a high school's quality, high school GPA, and SAT/ACT score. Although most colleges place the highest importance on academic achievement in evaluating applications, each factor's specific weight can differ from college to college. Moreover, "holistic admission"—such that a high SAT/ACT score and a high GPA is no guarantee of admission—is not rare in college admissions, especially for top colleges. The decomposition in Eq. (14) incorporates the college-specific weight for students' academic performance and extracurricular activity to terms $e_{ij}^\dagger$ and $e_{ij}^\ddagger$. The term $e_{ij}^\dagger$ in Eq. (14) can represent college-specific adjustment, which is "private valuation" of a student's high school record and SAT/ACT score, in addition to special talent,

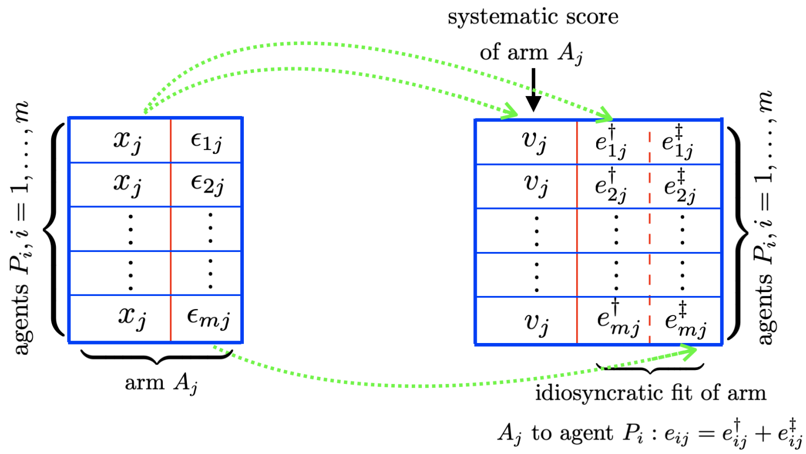

Figure 7: ANOVA decomposition of the utility function corresponding to Eq. (14).

grades of challenging college preparatory curriculum and work experience. For example, some colleges may place a larger weight on the SAT/ACT score than other colleges. The term $e^\ddagger_{ij}$ in Eq. (14) is a college-specific "private valuation" of students' writing skills and compelling personal stories.

A.2 Proof of Theorem 1: Acceptance Probability

Proof The uncertainty of the arm's acceptance comes from two parts: state of the world and agents' strategies. We prove the existence of an acceptance probability mass function (PMF) $\pi_i(s_i, v_j)$ by a three-step construction. First, Nature draws the state ω and arms' preferences for agents that are characterized by ω. Then agent-specific states are determined as a function of ω: $s_i = s_i(\omega)$ for $i \in [m]$.

Second, we derive an arm's acceptance probability. Suppose now agent P_i pulls arm A_j. Since A_j would accept its most preferred agent among those who have pulled it, A_j's acceptance of P_i depends on other agents' strategies. Let $I \subseteq [n]$ and $P_I \equiv \{P_i, i \in I\}$. We define

$$\mu_{i,I\cup\{i\}}(\omega, v_j, \mathbf{e}_j) \equiv \mathbb{P}(A_j \text{ accepts } P_i \mid P_{I\cup\{i\}} \text{ pulls } A_j).$$

That is, $\mu_{i,I\cup\{i\}}(\omega, v_j, \mathbf{e}_j)$ is the probability mass function (PMF) that an arm with the same score and fits as A_j would accept P_i conditional on agents $P_{I\cup\{i\}}$ have pulled the arm. The $\mu_{i,I\cup\{i\}}(\omega, v_j, \mathbf{e}_j)$ depends on arms' preferences as characterized by ω, and also on agents' strategies determined by the score v_j and the vector of fits $\mathbf{e}_j = (e_{1j}, \ldots, e_{mj})$. Moreover, $\mu_{i,I\cup\{i\}}(\omega, v_j, \mathbf{e}_j)$ is a valid PMF as it satisfies

$$\sum_{k\in I\cup\{i\}} \mu_{k,I\cup\{i\}}(\omega, v_j, \mathbf{e}_j) = 1, \quad \forall j \in [n]. \tag{15}$$

Third, we derive an arm's acceptance probability from an agent's perspective. Suppose again that agent P_i pulls arm A_j. Let $\mathcal{I}_{-i} \equiv \{I : I \subseteq [n] \setminus \{i\}\}$ be the family of subsets of

$[n] \setminus \{i\}$. Then by Tonelli's theorem,

$$\begin{aligned}
&\mathbb{P}(A_j \text{ accepts } P_i \mid P_i \text{ pulls } A_j) \\
&= \mathbb{E}\left[\mathbb{P}(A_j \text{ accepts } P_i \mid P_{I\cup\{i\}} \text{ pulls } A_j, P_{[n]\setminus(I\cup\{i\})} \text{ do not pull } A_j)\right] \qquad (16)\\
&= \mathbb{E}_{I\in\mathcal{I}_{-i}}\left[\mathbb{E}_{\mathbf{e}_{-i,j}}[\mu_{i,I\cup\{i\}}(\omega, v_j, \mathbf{e}_j)\cdot \mathbf{1}(P_I \text{ pulls } A_j)\cdot \mathbf{1}(P_{[n]\setminus(I\cup\{i\})} \text{ do not pull } A_j)]\right],
\end{aligned}$$

where $\mathbf{e}_{-i,j} = (e_{1j}, \ldots, e_{i-1,j}, e_{i+1,j}, \ldots, e_{n,j})$. Conditional on v_j, $\mathbf{1}(P_I \text{ pulls } A_j)$ only depends on $\{e_k, k \in I\}$, and similarly, $\mathbf{1}(P_{[n]\setminus(I\cup\{i\})} \text{ do not pull } A_j)$ only depends on $\{e_k, k \in [n] \setminus (I \cup \{i\})\}$. By definition,

$$\pi_i(s_i, v_j) = \mathbb{P}(A_j \text{ accepts } P_i \mid P_i \text{ pulls } A_j).$$

By Eqs. (15) and (16), it is clear that $\pi_i(s_i, v_j)$ is a valid marginal PMF of acceptance. In particular, $\pi_i(s_i, v_j)$ is averaged over other agents' strategies except P_i's. Note that $\pi_i(s_i, v_j)$ does not depend on e_{ij} since π_i is defined by conditioning on P_i's strategy, that is, P_i pulls A_j. In other words, the probability $\pi_i(s_i, v_j)$ represents the uncertainty of A_j accepting P_i.

Finally, the expected utility that agent P_i receives from pulling arm $A_j \in \mathcal{A}$ is

$$\begin{aligned}
\mathbb{E}[\text{utility}] &= \mathbb{E}[\text{utility} \mid \text{successful pulling}] \cdot \mathbb{P}(\text{successful pulling}) \\
&= (v_j + e_{ij}) \cdot \pi_i(s_i, v_j).
\end{aligned}$$

This completes the proof. ■

Remark 11 *Although the probability $\mu_{i,I\cup\{i\}}(\omega, v_j, \mathbf{e}_j)$ defined in the proof captures the distribution of arms' preferences, it is imperfect in practice for two reasons. First, if P_i wants to estimate $\mu_{i,I\cup\{i\}}(\omega, v_j, \mathbf{e}_j)$, it requires P_i to identify other agents who are also pulling A_j, that is, to identify the set I. However, each agent's choice set for arms differs over time and P_i cannot learn which arms the other agents are pulling since the market is decentralized and communications among agents are not allowed. Second, $\mu_{i,I\cup\{i\}}(\omega, v_j, \mathbf{e}_j)$ relies on the fit vector $\mathbf{e}_{-i,j} = (e_{1j}, \ldots, e_{i-1,j}, e_{i+1,j}, \ldots, e_{n,j})$ which is unknown to agent P_i. On the contrary, $\pi_i(s_i, v_j)$ does not require the knowledge of which agents are pulling A_j besides P_i, and $\pi_i(s_i, v_j)$ is independent of the fit $\mathbf{e}_{-i,j}$. As a result, $\pi_i(s_i, v_j)$ is estimable by P_i using historical data.*

A.3 Proof of Theorem 2: Optimal Estimation of Acceptance

Proof Let $\widetilde{f}_i$ be the minimizer of Eq. (3), that is,

$$\widetilde{f}_i = \operatorname*{arg\,min}_{f_i\in\mathcal{H}_{K_i}} \left\{ \sum_{t=1}^{T} \sum_{j\in\mathcal{B}_i^t} \left[-y_{ij}^t f_i(s_i^t, v_j^t) + \log\left(1 + \exp\left(f_i(s_i^t, v_j^t)\right)\right)\right] + \frac{1}{2}\sum_{t=1}^{T} n_{it}\lambda_i \|f_i\|^2_{\mathcal{H}_{K_i}} \right\}.$$

By the results in Chapters 5 of Lin (1998), we obtain that

$$\mathbb{E}_{s_i,v}[(\widetilde{f}_i - f_i)^2] \le c_1 \left[T(\log T)^{-1}\right]^{-2r/(2r+1)} \quad \text{as } T \to \infty,$$

where $\lambda_i \le c_\lambda[T(\log T)^{-1}]^{-2r/(2r+1)}$. Here, $c_\lambda, c_1 > 0$ are constants independent of T. Moreover, the estimate $\widetilde{f}_i$ is minimax rate-optimal. By the generalization properties of random features (Rudi and Rosasco, 2017), it is known that if the number of random features satisfies

$$p \ge c_p[T(\log T)^{-1}]^{-2r/(2r+1)},$$

then there exists some constant $c_2 > 0$ such that

$$\mathbb{E}_{s_i,v}[(\widehat{f}_i - \widetilde{f}_i)^2] \le c_2 \left[T(\log T)^{-1}\right]^{-2r/(2r+1)} \quad \text{as } T \to \infty.$$

By the triangle inequality, there exists some constant $c_f > 0$ such that

$$\begin{aligned}\mathbb{E}_{s_i,v}[(\widehat{f}_i - f_i)^2] &\le \mathbb{E}_{s_i,v}[(\widehat{f}_i - \widetilde{f}_i)^2] + \mathbb{E}_{s_i,v}[(\widetilde{f}_i - f_i)^2] \\ &\le c_f \left[T(\log T)^{-1}\right]^{-2r/(2r+1)} \quad \text{as } T \to \infty.\end{aligned}$$

Therefore, the estimate $\widehat{f}_i$ in Eq. (4) is minimax rate-optimal. ■

A.4 Proof of Theorem 3: Cutoff Strategy

Proof First, we show that the cutoff strategy with respect to the fits is optimal. Suppose that arms $A_{j_1}, A_{j_2} \in \mathcal{A}^{T+1}$ have the same score $v_{j_1} = v_{j_2}$, but A_{j_1} has a worse fit than A_{j_2} for agent P_i. Now assume that A_{j_1} was pulled by P_i but A_{j_2} was not, that is, $A_{j_1} \in \widehat{\mathcal{B}}_i(s_i), A_{j_2} \notin \widehat{\mathcal{B}}_i(s_i)$. Then the expected number of arms accepting P_i is unchanged if P_i replaces A_{j_1} with A_{j_2} in $\widehat{\mathcal{B}}_i(s_i)$. On the other hand, since the P_i's expected payoff in Eq. (6) is strictly increasing in fit e_{ij}, P_i has a strictly larger expected payoff if P_i replaces A_{j_1} with A_{j_2}. Hence, P_i should pull A_{j_2} instead A_{j_1}. This argument holds regardless of strategies of other agents.

Second, we prove that the cutoff $\widehat{e}_i(s_i, v)$ in Eq. (8) is well-defined. If the boundary $\{\mathcal{B}_i^+(s_i) \setminus \mathcal{B}_i^-(s_i)\}$ is not empty, the expected penalty due to exceeding the quota is

$$\gamma_i \sum_{j \in \mathcal{B}_i^+(s_i)} \pi_i(s_i, v_j) - \gamma_i q_i.$$

The expected utility of pulling arms on the boundary is

$$\sum_{j \in \mathcal{B}_i^+(s_i) \setminus \mathcal{B}_i^-(s_i)} (v_j + e_{ij}) \cdot \pi_i(s_i, v_j).$$

Agent P_i pulls an arm if the expected utility is at least the expected penalty, which justifies the condition specified by Eq. (7). Moreover, since $\widehat{e}_i(s_i, v) \in [0, 1]$, we conclude that the cutoff is well-defined.

Third, we prove that the cutoff $\widehat{e}_i(s_i, v)$ is the unique optimal cutoff. Let $\widetilde{e}_i(s_i, v) \in [0, 1]$ be any other cutoff. We define a mixed strategy

$$\sigma_i(s_i, v, e_i; t) \equiv t \cdot \mathbf{1}\{e_i \ge \widetilde{e}_i(s_i, v)\} + (1 - t) \cdot \mathbf{1}\{e_i \ge \widehat{e}_i(s_i, v)\}, \quad \text{for } t \in [0, 1].$$

The expected payoff of using the mixed strategy σ_i is

$$\mathbb{U}_i(t) \equiv \sum_{j \in \mathcal{A}^{T+1}} (v_j + e_{ij}) \cdot \pi_i(s_i, v_j) \cdot \sigma_i(s_i, v_j, e_{ij}; t) - \gamma_i \cdot \max \left\{ \sum_{j \in \mathcal{A}^{T+1}} \pi_i(s_i, v_j) \cdot \sigma_i(s_i, v_j, e_{ij}; t) - q_i, 0 \right\}.$$

It is clear that $\mathbb{U}_i(t)$ is concave in t. We discuss the local change $d\mathbb{U}_i(0)/dt$ in three cases.

Case (I). Consider removing a single arm from $\widehat{\mathcal{B}}_i(s_i)$. If the arm is from the non-empty boundary $\{\mathcal{B}_i^+(s_i) \setminus \mathcal{B}_i^-(s_i)\}$, the condition specified in Eq. (7) implies that P_i's expected payoff will decrease if not pulling the arm. Moreover, since removing any other arms from $\widehat{\mathcal{B}}_i(s_i)$ results in a greater loss of P_i's expected payoff compared to removing the arms on the non-empty boundary, we have $d\mathbb{U}_i(0)/dt < 0$ in this case.

Case (II). Consider adding a new arm to $\widehat{\mathcal{B}}_i(s_i)$, where the new arm has attributes $\{v_{j'}, e_{ij'}\}$ and it is not from the set $\mathcal{B}_i^+(s_i)$. Denote by $\mathcal{B}_i'(s_i)$ the new arm set with the added arm. Note that P_i pulls a new arm only if the arm has a larger expected utility than the expected penalty due to exceeding the quota, that is,

$$(v_{j'} + e_{ij'}) \cdot \pi(s_i, v_{j'}) \geq \gamma_i \sum_{j \in \mathcal{B}_i'(s_i)} \pi_i(s_i, v_j) - \gamma_i q_i. \tag{17}$$

Since the new arm is not in $\mathcal{B}_i^+(s_i)$, we have

$$\begin{aligned} &\sum_{j \in \mathcal{B}_i'(s_i)} \pi_i(s_i, v_j) - q_i \\ &= \sum_{j \in \mathcal{B}_i'(s_i)} \pi_i(s_i, v_j) - \sum_{j \in \mathcal{B}_i^+(s_i)} \pi_i(s_i, v_j) + \sum_{j \in \mathcal{B}_i^+(s_i)} \pi_i(s_i, v_j) - q_i \\ &\geq \sum_{j \in \mathcal{B}_i'(s_i)} \pi_i(s_i, v_j) - \sum_{j \in \mathcal{B}_i^+(s_i)} \pi_i(s_i, v_j) \\ &\geq \pi_i(s_i, v_{j'}). \end{aligned} \tag{18}$$

Because that $\gamma_i > \sup_{j \in \mathcal{A}^{T+1}} \{v_j + e_{ij}\}$, Eq. (18) is contradictory to Eq. (17). Hence, adding a new arm to $\widehat{\mathcal{B}}_i(s_i)$ induces a loss in P_i's expected payoff. Hence, $d\mathbb{U}_i(0)/dt < 0$ in this case.

Case (III). Consider removing an arm with attributes (v_j, e_{ij}) from $\widehat{\mathcal{B}}_i(s_i)$ and simultaneously adding new arms to $\widehat{\mathcal{B}}_i(s_i)$. Suppose that the new arms are from the arm set $\mathcal{B}_i''(s_i)$ with attributes $(v_{j''}, e_{ij''})$. Using the argument similar to Eq. (18), we know that the following condition must hold:

$$\sum_{j'' \in \mathcal{B}_i''(s_i)} \pi_i(s_i, v_{j''}) \leq \pi_i(s_i, v_j).$$

Since $(v_{j''}, e_{ij''}) \notin \mathcal{B}_i^+(s_i)$, by definition, $v_{j''} + e_{ij''} < v_j + e_{ij}$. Hence, the expected utility of added arms satisfies

$$\sum_{j'' \in \mathcal{B}_i''(s_i)} (v_{j''} + e_{ij''}) \cdot \pi_i(s_i, v_{j''})$$
$$< (v_j + e_{ij}) \sum_{j'' \in \mathcal{B}_i''(s_i)} \pi_i(s_i, v_{j''}) \leq (v_j + e_{ij}) \cdot \pi_i(s_i, v_j).$$

Therefore, exchanging an arm in $\widehat{\mathcal{B}}_i(s_i)$ with arms not in $\widehat{\mathcal{B}}_i(s_i)$ results in a smaller expected payoff for P_i. Hence, $d\mathbb{U}_i(0)/dt < 0$ in this case.

Combining the cases (I), (II), (III), we conclude that $d\mathbb{U}_i(0)/dt < 0$. By the concavity of $\mathbb{U}_i(t)$ in t, we obtain

$$\mathbb{U}_i(1) = \mathbb{U}_i(0) + \frac{d\mathbb{U}_i(0)}{dt}(1 - 0) < \mathbb{U}_i(0),$$

which implies that $\widehat{e}_i(s_i, v)$ is the unique optimal cutoff. This completes the proof. ■

A.5 Proof of Theorem 4: Mean Calibration for CDM

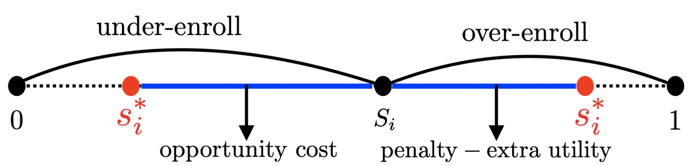

Figure 8: Cost of a strategy in the face of uncertain true state s_i^*.

Proof By the proof of Theorem 3, $\widehat{\mathcal{B}}_i(s_i) \subseteq \widehat{\mathcal{B}}_i(s_i - \delta s_i)$ for any $\delta s_i \in (0, s_i)$. Hence, the marginal set $\partial\widehat{\mathcal{B}}_i(s_i)$ is well-defined. Let $V_i(s_i^*, \widehat{\mathcal{B}}_i(s_i))$ be the expected utility that P_i receives by pulling arms from $\widehat{\mathcal{B}}_i(s_i)$ and under the true state s_i^*. That is,

$$V_i(s_i^*, \widehat{\mathcal{B}}_i(s_i)) \equiv \sum_{j \in \widehat{\mathcal{B}}_i(s_i)} (v_j + e_{ij}) \cdot \pi_i(s_i^*, v_j).$$

Let $\mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(s_i))$ be the expected number of arms in $\widehat{\mathcal{B}}_i(s_i)$ accepting P_i under s_i^*. That is,

$$\mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(s_i)) \equiv \sum_{j \in \widehat{\mathcal{B}}_i(s_i)} \pi_i(s_i^*, v_j).$$

Similarly, we define $V_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i))$ and $\mathcal{N}_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i))$ for the marginal set $\partial\widehat{\mathcal{B}}_i(s_i)$. Let the marginal utility be

$$u_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i)) \equiv \frac{V_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i))}{\mathcal{N}_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i))}.$$

Since the true state s_i^* is unknown, the cost of pulling the arm set $\widehat{\mathcal{B}}_i(s_i)$ consists of two parts. See an illustration in Figure 8. The first is the *over-enrollment* cost (OE), which occurs if the calibration parameter $s_i < s_i^*$. Then the realized number of arms in $\widehat{\mathcal{B}}_i(s_i)$ accepting P_i will be greater than the expected number of arms in $\widehat{\mathcal{B}}_i(s_i)$ accepting P_i. That is, $\mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(s_i)) > \mathcal{N}_i(s_i, \widehat{\mathcal{B}}_i(s_i)) = \mathcal{N}_i(s^*, \widehat{\mathcal{B}}_i(s_i^*)) = q_i$. Thus, OE depends on s_i and can be written as

$$\begin{aligned}
&\mathrm{OE}(s_i) \\
&\equiv \mathbb{E}_{s_i^*}\left[\gamma_i \left\{\mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(s_i)) - \mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(s_i^*))\right\} - \left\{V_i(s_i^*, \widehat{\mathcal{B}}_i(s_i)) - V_i(s_i^*, \widehat{\mathcal{B}}_i(s_i^*))\right\} \;\middle|\; s_i < s_i^* \le 1\right] \\
&= \mathbb{E}_{s^*}\left[\int_{t=s_i}^{s_i^*} [\gamma_i - u_i(s_i^*, \partial\widehat{\mathcal{B}}_i(t))] \cdot \mathcal{N}_i(s_i^*, \partial\widehat{\mathcal{B}}_i(t))dt \;\middle|\; s_i < s_i^* \le 1\right].
\end{aligned}$$

Here, $\mathrm{OE}(s_i)$ equals the penalty of the arms in $\widehat{\mathcal{B}}_i(s_i)$ which would accept P_i exceeding the quota, and deducts the utility of these arms.

The second part of the cost is *under-enrollment* (UE), which occurs if the calibration parameter $s_i > s_i^*$. Then $\mathcal{N}_i(s^*, \widehat{\mathcal{B}}_i(s_i)) < \mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(s_i^*)) = q_i$. The $\mathrm{UE}(s_i)$ equals the opportunity cost in the sennse that P_i could have successfully pulled more arms:

$$\begin{aligned}
\mathrm{UE}(s_i) &\equiv \mathbb{E}_{s_i^*}[V_i(s_i^*, \widehat{\mathcal{B}}_i(s_i^*)) - V_i(s_i^*, \widehat{\mathcal{B}}_i(s_i)) \mid 0 \le s_i^* < s_i] \\
&= \mathbb{E}_{s_i^*}\left[\int_{t=s_i^*}^{s_i} u_i(s_i^*, \partial\widehat{\mathcal{B}}_i(t)) \cdot \mathcal{N}_i(s_i^*, \partial\widehat{\mathcal{B}}_i(t))dt \;\middle|\; 0 \le s_i^* < s_i\right].
\end{aligned}$$

Therefore, the goal of finding s_i to maximize the P_i's average-case expected payoff can be written as:

$$\operatorname*{arg\,max}_{s_i \in (0,1)} \left\{\mathbb{E}_{s_i^*}\left[V_i(s_i^*, \widehat{\mathcal{B}}_i(s_i)) - \gamma_i \max\{\mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(s_i)) - q_i, 0\}\right]\right\}.$$

This goal is equivalent to finding s_i to minimize the weighted sum of $\mathrm{OE}(s_i)$ and $\mathrm{UE}(s_i)$ with the occurrence probabilities as the weights:

$$\operatorname*{arg\,min}_{s_i \in (0,1)} \left\{(1 - F_{s_i^*}(s_i))\mathrm{OE}(s_i) + (F_{s_i^*}(s_i) - \mathbb{P}(s_i^* = s_i))\mathrm{UE}(s_i)\right\}.$$

By the first-order condition, the minimizer $s_i \in (0,1)$ satisfies

$$\begin{aligned}
&[1 - \mathbb{P}(s_i^* = s_i)]\mathbb{E}_{s_i^*}[V_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i)) \mid s_i^* \ne s_i] \\
&\qquad = \gamma_i(1 - F_{s_i^*}(s_i))\mathbb{E}_{s_i^*}[\mathcal{N}_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i)) \mid s_i < s_i^* \le 1].
\end{aligned} \tag{19}$$

This result proves Eq. (9). Note that there always exists a solution to Eq. (9) since when $s_i \to 0_+$, $F_{s_i^*}(s_i) \to 0, \gamma_i > u_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i))$; and when $s_i \to 1_-$, $F_{s_i^*}(s_i) \to 1, V_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i)) > 0$; and the right side of Eq. (9) is strictly decreasing in s_i.

Here, we assume the calibration parameter $s_i \in (0,1)$ in the definitions of $\mathrm{OE}(s_i)$ and $\mathrm{UE}(s_i)$. We prove that this assumption is without less of generality by showing that if $s_i = 1$, P_i can pull more arms to obtain a larger expected payoff, and if $s_i = 0$, P_i can pull less arms to obtain a larger expected payoff. Consider that if $s_i = 1$ and P_i pulls an additional arm A which is not pulled currently, that is, $A \notin \widehat{\mathcal{B}}_i(1)$. Then the expected number of arms that accept agent P_i is

$$\mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(1) \cup \{A\}) = \mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(1)) + \mathcal{N}_i(s_i^*, \{A\})$$

Let $\widetilde{s}_i$ satisfy $\mathcal{N}_i(\widetilde{s}_i, \widehat{\mathcal{B}}_i(1) \cup \{A\}) = q_i$. Since $\mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(1) \cup \{A\}) > \mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(1))$, we have $\widetilde{s}_i < 1$. Let A be the arm such that $\widetilde{s}_i > 1 - \epsilon_s$ for some sufficiently small $\epsilon_s > 0$. Then the difference of average-case expected payoffs from pulling two arm sets $\widehat{\mathcal{B}}_i(1) \cup \{A\}$ and $\widehat{\mathcal{B}}_i(1)$ is

$$\begin{aligned}
&\mathbb{E}_{s_i^*}\left[V_i(s_i^*, \{A\})\right] - \gamma_i \mathbb{E}_{s_i^*}\left[q_i - \mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(1) \cup \{A\}) \;\middle|\; \widetilde{s}_i < s_i^* \le 1\right] \\
&= \mathbb{E}_{s_i^*}\left[V_i(s_i^*, \{A\})\right] - \gamma_i \mathbb{E}_{s_i^*}\left[\mathcal{N}_i(s_i^*, \{A\}) \mid \widetilde{s}_i < s_i^* \le 1\right] \\
&\qquad\qquad\qquad\qquad + \gamma_i \mathbb{E}_{s_i^*}\left[q_i - \mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(1)) \;\middle|\; \widetilde{s}_i < s_i^* \le 1\right] \\
&> \mathbb{E}_{s_i^*}\left[V_i(s_i^*, \{A\})\right] - \gamma_i \mathbb{E}_{s_i^*}\left[\mathcal{N}_i(s_i^*, \{A\}) \mid \widetilde{s}_i < s_i^* \le 1\right] \\
&= U_i(A)\mathbb{E}_{s_i^*}[\mathcal{N}_i(s_i^*, \{A\}) \mid 0 \le s_i^* \le \widetilde{s}_i] - [\gamma_i - U_i(A)]\mathbb{E}_{s_i^*}[\mathcal{N}_i(s_i^*, \{A\}) \mid \widetilde{s}_i < s_i^* \le 1] \\
&> 0,
\end{aligned}$$

where $U_i(A)$ is the latent utility of arm A defined in Eq. (1). The last step holds for sufficiently small $\epsilon_s > 0$. Similarly, if $s_i = 0$, P_i can benefit by pulling less arms. Thus, the assumption that $s_i \in (0, 1)$ is without less of generality.

If $F_{s_i^*}(\cdot)$ has discrete support, we require that $\mathrm{UE}(s_i)$ is at least $\mathrm{OE}(s_i)$. By the first-order condition similar to Eq. (19), we find the minimal $s_i \in [0, 1]$ such that

$$\begin{aligned}
&[1 - \mathbb{P}(s_i^* = s_i)]\mathbb{E}_{s_i^*}[V_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i)) \mid s_i^* \neq s_i] \\
&\qquad \ge \gamma_i(1 - F_{s_i^*}(s_i))\mathbb{E}_{s_i^*}[\mathcal{N}_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i)) \mid s_i < s_i^* \le 1],
\end{aligned} \tag{20}$$

where the search of s_i starts from the maximum value in the support to the minimal value. We note that the calibration in Eq. (20) is a *conservative* counterpart as compared with the calibration such that $\mathrm{OE}(s_i)$ is at least $\mathrm{UE}(s_i)$:

$$\begin{aligned}
&[1 - \mathbb{P}(s_i^* = s_i)]\mathbb{E}_{s_i^*}[V_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i)) \mid s_i^* \neq s_i] \\
&\qquad \le \gamma_i(1 - F_{s_i^*}(s_i))\mathbb{E}_{s_i^*}[\mathcal{N}_i(s_i^*, \partial\widehat{\mathcal{B}}_i(s_i)) \mid s_i < s_i^* \le 1].
\end{aligned} \tag{21}$$

The calibration in Eq. (20) is preferred to that in Eq. (21) since we want the calibration to be sensitive to the penalty γ_i. This completes the proof. ∎

A.6 Proof of Theorem 5: Maximin Calibration for CDM

Proof We use the notations $V_i(s_i^*, \widehat{\mathcal{B}}_i(s_i))$ and $\mathcal{N}_i(s_i^*, \widehat{\mathcal{B}}_i(s_i))$ defined in Appendix A.5. The maximum over-enrollment cost for any $s_i \in [0, 1]$ is

$$\max_{s_i^* \in [0,1]} \{\mathrm{OE}(s_i)\} = \gamma_i\{\mathcal{N}_i(1, \widehat{\mathcal{B}}_i(s_i)) - \mathcal{N}_i(1, \widehat{\mathcal{B}}_i(1))\} - \{V_i(1, \widehat{\mathcal{B}}_i(s_i)) - V_i(1, \widehat{\mathcal{B}}_i(1))\}.$$

Since $\gamma_i > \sup_{j \in \mathcal{A}^{T+1}}\{v_j + e_{ij}\}$, $\max_{s_i^*}\{\mathrm{OE}(s_i)\}$ is strictly decreasing in s_i. The maximum under-enrollment cost for any $s_i \in [0,1]$ is

$$\max_{s_i^* \in [0,1]}\{\mathrm{UE}(s_i)\} = V_i(0, \widehat{\mathcal{B}}_i(0)) - V_i(0, \widehat{\mathcal{B}}_i(s_i)),$$

which is strictly increasing in s_i. Hence, the goal of maximizing the minimal total expected payoff $\max_{s_i}\min_{s_i^*}\{U_i^{\mathrm{S}}(\widehat{\mathcal{B}}_i(s_i))\}$ is equivalent to minimizing the larger one between $\mathrm{OE}(s_i)$ and $\mathrm{UE}(s_i)$:

$$\min_{s_i \in [0,1]} \max\left\{\max_{s_i^*}\{\mathrm{OE}(s_i)\}, \max_{s^*}\{\mathrm{UE}(s_i)\}\right\}.$$

This objective amounts to finding s_i such that

$$\max_{s_i^* \in [0,1]}\{\mathrm{OE}(s_i)\} = \max_{s_i^* \in [0,1]}\{\mathrm{UE}(s_i)\}. \tag{22}$$

This proves Theorem 5. Moreover, there exists a unique solution to Eq. (22) since when $s_i = 0$, $\max_{s_i^*}\{\mathrm{OE}(0)\} > \max_{s_i^*}\{\mathrm{UE}(0)\} = 0$, and when $s_i = 1$, $\max_{s_i^*}\{\mathrm{UE}(1)\} > 0 = \max_{s_i^*}\{\mathrm{OE}(1)\}$, and together with the fact that $\max_{s_i^*}\{\mathrm{OE}(s_i)\}$ and $\max_{s_i^*}\{\mathrm{UE}(s_i)\}$ are monotonic continuous functions of s_i.

If $F_{s_i^*}(\cdot)$ has discrete support, we requires that the maximal $\mathrm{UE}(s_i)$ is at least the maximal $\mathrm{OE}(s_i)$. Hence, we need to change the goal in Eq. (22) to finding the minimal $s_i \in [0,1]$ such that $\max_{s_i^*}\{\mathrm{UE}(s_i)\} \geq \max_{s_i^*}\{\mathrm{OE}(s_i)\}$. This completes the proof. ■

A.7 Proof of Theorem 6: Incentives of Agents

Proof First, by Theorem 3, CDM uses the cutoff strategy to arms' latent utilities in Eq. (1). Since arms' latent utilities determine agents' true preferences, agents act according to their true preferences under CDM.

Second, Theorem 2 proves the consistency of the acceptance probability estimate using historical data. Theorems 4 and 5 show that the CDM maximizes the agent's expected payoff, given the population acceptance probability. Here the expected payoff is measured in either average-case or worst-case concerning the uncertain true state. Thus, it is a dominant strategy for each agent to act according to the CDM.

Combining these two observations, we conclude that as $T \to \infty$, CDM is a procedure that gives agents the incentives to act according to true preferences. ■

A.8 Proof of Theorem 7: Stability of CDM

Proof Suppose that an agent-arm pair (P_i, A_j) forms a blocking pair. Then it implies one of the following two cases:

(I). P_i prefers A_j to some of its matched arms.

(II). P_i has unfilled quota and A_j is unmatched.

For case (I), P_i must have pulled A_j according to the cutoff strategy of CDM by Theorem 3. However, P_i must have been subsequently rejected by A_j in favor of some agent that A_j liked better. Hence, A_j must prefer its currently matched agent to P_i and there is no instability.

For case (II), P_i did not pull A_j since otherwise, A_j would have accepted P_i. Note that Theorem 2 proves the consistency of acceptance probability estimate using historical data. The individual rationality in Eq. (11) implies that P_i would find A_j unacceptable in the decentralized market, given the population acceptance probability. Thus, there is no instability.

Combining these two cases, we conclude that as $T \to \infty$, CDM yields a stable matching in decentralized markets. ∎

A.9 Proof of Theorem 8: Fairness of CDM

Proof By Theorem 6, the CDM gives agents the incentives to act on their true preferences, as $T \to \infty$. Hence, each agent pulls arms according to its true preference for arms. If an arm A_j prefers an agent $P_{i'}$ to another agent P_i that pulls A_j, then all arms pulled by $P_{i'}$ must rank above A_j according to the true preference of $P_{i'}$. By definition, we conclude that the matching procedure according to CDM is fair. ∎

A.10 Proof of Theorem 9: Unfairness of the Oracle Set

Proof Denote by $e_i^*(v)$ the cutoff curve corresponding to the oracle arm set $\mathcal{B}_i^*$ in Eq. (12):

$$e_i^*(v_j) = \min\Big\{\max\Big\{\gamma_i \cdot \mathbb{P}(s_i^* \in O_{\mathcal{B}_i^*})\frac{\mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j) \mid s_i^* \in O_{\mathcal{B}_i^*}]}{\mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j)]} - v_j, 0\Big\}, 1\Big\}, \quad \forall j.$$

The average-case expected utility of an arm from the cutoff curve is

$$\mathcal{U}_i(v_j, e_i^*(v_j)) = (v_j + e_i^*(v_j)) \cdot \mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j)] - \gamma_i \cdot \mathbb{P}(s_i^* \in O_{\mathcal{B}_i^*})\mathbb{E}_{s_i^*}[\pi_i(s_i^*, v) \mid s_i^* \in O_{\mathcal{B}_i^*}].$$

Hence, for $e_i^*(v_j) \in (0,1)$,

$$\mathcal{U}_i(v_j, e_i^*(v_j)) = 0.$$

Since the acceptance probability $\pi_i(s_i, v)$ is strictly increasing in s_i, $\mathbb{E}_{s_i^*}[\pi_i(s_i^*, v) \mid s_i^* \in O_{\mathcal{B}_i^*}] > \mathbb{E}_{s_i^*}[\pi_i(s_i^*, v)]$. Thus, for $e_i^*(v_j) \in (0,1)$,

$$\begin{aligned} v_j + e_i^*(v_j) &= \gamma_i \cdot \mathbb{P}(s_i^* \in O_{\mathcal{B}_i^*})\frac{\mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j) \mid s_i^* \in O_{\mathcal{B}_i^*}]}{\mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j)]} \\ &> \gamma_i \cdot \mathbb{P}(s_i^* \in O_{\mathcal{B}_i^*}). \end{aligned} \tag{23}$$

By Eq. (13), we can derive that

$$\mathbb{E}_{s_i^*}\left[\frac{\partial \pi_i(s_i^*, v_j)}{\partial v}\right] < \mathbb{E}_{s_i^*}\left[\frac{\partial \pi_i(s_i^*, v_j)}{\partial v} \;\middle|\; s_i^* \in O_{\mathcal{B}_i^*}\right] < 0, \quad \forall v_j \in (v', v'').$$

This inequality together with Eq. (23) yield that for $v_j \in (v', v'')$ and $e_i^*(v_j) \in (0,1)$,

$$\begin{aligned}\left.\frac{\partial \mathcal{U}_i(v, e_i^*(v))}{\partial v}\right|_{v=v_j} &= \mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j)] + (v_j + e_i^*(v_j))\mathbb{E}_{s_i^*}\left[\frac{\partial \pi_i(s_i^*, v_j)}{\partial v}\right] \\ &\qquad - \gamma_i \cdot \mathbb{P}(s_i^* \in O_{\mathcal{B}_i^*})\mathbb{E}_{s_i^*}\left[\frac{\partial \pi_i(s_i^*, v_j)}{\partial v} \;\middle|\; s_i^* \in O_{\mathcal{B}_i^*}\right] \\ &< \mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j)].\end{aligned}$$

Thus, by the implicit function theorem,

$$\left|\frac{de_i^*(v_j)}{dv}\right| = \left|\frac{\partial \mathcal{U}_i(v_j, e_i^*(v_j))/\partial v}{\partial \mathcal{U}_i(v_j, e_i^*(v_j))/\partial e_i^*}\right| < \frac{\mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j)]}{\mathbb{E}_{s_i^*}[\pi_i(s_i^*, v_j)]} = 1, \quad \forall v_j \in (v', v''),$$

Therefore, we can find two arm sets $\mathcal{B}_i^{(1)}, \mathcal{B}_i^{(2)} \subseteq (v', v'') \times [0,1]$ and a constant c_0 such that, for all $(v_j, e_{ij}) \in \mathcal{B}_i^{(1)}, e_{ij} > e_i^*(v_j)$ and $v_j + e_{ij} < c_0$, and for all $(v_j, e_{ij}) \in \mathcal{B}_i^{(2)}, e_{ij} < e_i^*(v_j)$ and $v_j + e_{ij} > c_0$. Hence, the arms in $\mathcal{B}_i^{(2)}$ have justified envy toward arms in $\mathcal{B}_i^{(1)}$. We refer to Figure 2 for an illustration. By definition of fairness, the strategy corresponding to the oracle set in Eq. (12) is unfair. ■