Essays on Empirical Two-Sided Matching Models with Unobserved Heterogeneity

by

Yajing Jiang

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

August, 2016

Abstract

This dissertation studies two structural frameworks in empirical studies of Industrial Organization: two-sided matching models and simultaneous auction/contest models. Both models involve two disjoint sides of players: matching occurs between firms and workers, schools and students, and so forth; in an auction or contest, there always exists an auctioneer or contest designer on one side and bidders or contestants on the opposite. First, empirical studies of two-sided matching markets reveal that sorting patterns between potential employers and employees may be driven by unobserved heterogeneity on both sides and preferences over multidimensional wage contracts. Therefore, in Chapter 2, I study a generalized matching model with Non-Transferrable Utility (NTU), i.e. a two-stage model where employers firstly set wage contracts for their jobs, workers then match with the jobs in a decentralized way. I propose a strategy that exploits the variation in agent- and match-specific characteristics from finite-sized repeated markets to identify and estimate the preference primitives in the presence of two-sided unobserved heterogeneity, assuming employers share a vertical preference over workers. I further suggest a likelihood-based estimation strategy that tackles the dimensionality issue emerging from the existence of global players of the repeated game, and show its performance via Monte Carlo simulation analyses.

In Chapter 3, I apply the model and identification method developed in Chapter 2 to study recently fast-growing online labor markets that match skilled labor to short-term jobs using a contest-based mechanism. Despite the anecdotal evidence showing both firms and workers benefiting from largely decreased meeting friction and increased flexibility on the platform, it is economically substantial to quantitatively reveal the preference structure of both parties, which may include unobserved factors to researchers. I, therefore, adopt the two-stage model where firms set wage contracts for their jobs before programmers choose coding projects simultaneously. I then use the identification strategy that exploits the variation in agent- and match-specific characteristics from finite-sized repeated markets to estimate workers' latent skill levels and jobs' latent complexity levels. Using individual-level data from a leading online tournament-based labor market, TopCoder.com, which matches workers worldwide with short-term software developing tasks, I find a multidimensional preference beyond cash motives when workers consider which jobs to take. Using the results from the estimation, I further study the elements regarding market design that could leverage off the matching mechanisms to improve the total surplus generated from such markets.

While this "crowdsourcing" market can be modeled as a matching process between

ABSTRACT

firms with temporary jobs and workers, I would also like to capture the strategic interaction of workers *after* they match with an individual job. Within the same job, workers exert effort to win pre-determined cash prizes according to the rank order of their delivered work. This can be naturally modeled in a (multi-prize) contest environment. A central concern is to recover the underlying preferences of workers, which again requires the full knowledge of the unobserved heterogeneity, or types, of both workers and jobs. Chapter 4 develops a new method to identify and estimate primitives in simultaneous contests with multiple prizes. I establish a two-stage game where bidders/contestants first choose one among multiple auctions/contests, then in the second stage, they compete within each auction/contest by submitting their bids simultaneously, contributing their efforts to win over the pre-determined prize based on the rank order. I show that by observing their first-stage choice probability combined with the second-stage bidding strategy, I can nonparametrically identify the joint distribution of unobserved heterogeneity on both sides of the market. I then present an estimation strategy and show the performance of Monte Carlo experiments.

Primary Reader: Professor Yingyao Hu

Secondary Reader: Professor Elena Krasnokutskaya

Acknowledgments

I wish to sincerely thank my advisors – Yingyao Hu (chair), Elena Krasnokutskaya, Yuya Takahashi and Jian Ni – for their continual and invaluable encouragement, guidance and support throughout the project. I also acknowledge the helpful comments from Ruli Xiao, Yuya Sasaki, Jorge Balat, Yi Xin and other applied-micro seminar participants at Johns Hopkins University for their advice over the years. Here I wish to add an extra special thanks to my husband Delong Li who provided the most precious companion, patience and guidance during my whole job market process. All errors are my own.

Contents

List of Tables

List of Figures

Chapter 1

Introductory Chapter

Two-sided markets have been widely observed in real life and reached growing attention by researchers in recent decades. Examples can be found from online markets matching skilled labor force with temporary jobs to the public school admission process. Economically, it is of great importance to understand how preferences from both sides of the market drive the observable market outcomes, such as the sorting between workers and jobs, and ultimately how the welfare could be leveraged up by policy intervention alternating the market power of players. This dissertation explores two different perspectives to identify the underlying preference primitives for both sides, allowing for the existences of multi-layer unobserved heterogeneity.

First, in Chapter 2, I establish a generalized one-to-many matching model with Non-Transferrable Utility (NTU) to study such markets. Specifically, there are two sets of players, denoted as firms and workers for simplicity. They meet in the market to match with each other to form an employment relationship. Capacities from both sides are constrained so that a worker can complete at most one job in the market, and firms can hire no more than their pre-determined slot quota for the job. Initially, firms set the wage scheme for the potentially paired workers simultaneously, based upon their own preference over workers' skills and rivalry from other firms; afterward, workers match with jobs by signing up a slot for a job, commonly observing the posted wages and other job features. Compared with standard NTU matching models in the literature, I add a wage-setting stage before the matching process, to relax the exogeneity assumption of monetary transfers under NTU framework and better understand how financial motives shifts the sorting patterns. After setting up this model, I define a rational expectation equilibrium notion, in which workers acquire pairwise stable matching in the second stage, and firms maximize their expected payoff in the first stage, holding correct beliefs about second-stage matching. Additionally, by assuming firms carrying a vertical preference over workers' skills, I manage to establish critical characterizations of the equilibrium that directly links to my identification strategy. In particular, I argue that from repeatedly observing (1) a one-dimensional index that proxies workers' skill level, and (2) a match-specific outcome for each matched worker-job pair for at least two periods, I have enough variation in observables and am able to nonparametrically identify the joint distribution of job-wise and worker-wise unobserved heterogeneity in each market, using the eigenvalue-eigenvector decomposition method developed in measurement error models. Next, I recover the joint choice probability for all workers in each market, which leads to the identification of workers' utility parameters. In this step, I follow the well-studied single-agent discrete choice framework to model workers' indirect utility from choosing any job or the outside option. Lastly, from jointly observing at least three firms in each market, I recover their profit structure along with the distribution of a market-level demand shifter, as the joint observation of multiple firms links to firms' primitives in a way similar to measurement error models as well. [1] Following the identification argument, I also provide a practically viable likelihood-based estimator that performs well in Monte Carlo simulation analyses. The estimator maximizes a modified likelihood that utilizes partial information from all markets observed and can be useful especially when the number of firms and/or workers in each market grows large.

Using data from an emerging online labor market that matches computer programmers to software development tasks on a weekly basis, I apply my generalized matching framework and identification strategy in Chapter 3 to unveil the market players' preferences. I have individual-level data about the website TopCoder.com, the world's online leader in accommodating programmers from worldwide to complete Information Technology (IT)-related short-term jobs remotely. It adopts the concept of "crowdsourcing", where multiple programmers usually engage in one job at the same time, and payment is delivered according to the rank order of their final prod-

[1] An example of the market-level demand shifter is the efficiency to for firms to recruit workers via this platform compare with alternative platforms.

ucts submitted to the firm. From data, I find a preliminary yet interesting pattern from regression – highly paid jobs on average get deliveries of poorer quality on this platform. As this is apparently unappealing for the long-run sustainability of the website, it is substantial to explore the underlying preference pattern of both parties using a structural model. By applying the two-stage model, I developed in Chapter 3 to estimate the parameters, I find that skilled workers, while indeed prefer to get more money out of a job, generically try to avoid more complicated jobs which pay more in equilibrium. Based on the estimates, I conduct two sets of counterfactual experiment, and find out that compared with stimulating skilled workers' passion towards demanding jobs through indirect channels, one more efficient way to boost the market is to allow firms to "discriminate" workers by using a wage menu contingent on the quality of submissions. These novel findings are not all intuitive compared with reduced-form regressions and have their credits in the empirical studies of online markets, as well as empirical matching market analyses.

The aforementioned "crowdsourcing" feature of the market I study in Chapter 3 relates to another important strand of structural IO studies, the all-pay auctions if the winner takes all and contests if prize schedule is based upon rank orders. Specifically, I shift my focus from the matching process between jobs and workers to the strategic behavior *after* workers choosing the job in Chapter 4. Workers are in fact engaged in a multi-prize contest, where they exert efforts to deliver their products to the firm, and their payment, or prize, depend on the rank order of the final products' quality.

Therefore, I simplify the wage-setting stage of firms to be exogenous and set up another two-stage model for workers afterward. [2] Workers hold a private valuation toward all jobs in the market. This can be alternatively thought of as their capacity of completing a job. They initially choose one job to take part in, then decide how much effort they exert to maximize their expected payoff. I show in this chapter that from jointly observing workers' first-stage choice probability and second-stage bidding behavior, I can nonparametrically recover the distribution of bother job-wise and worker-wise unobserved heterogeneity (or types). The intuition is that, following the seminal paper by Guerre, Perrigne and Voung (2009), I can establish a mapping from observed bids, or quality of delivered products, to the unobserved valuation of each worker. [3] On the job side, the variation in the probability of choosing a job given different numbers of competing workers in the same market provides identification power under regulatory conditions. I suggest a two-step estimator corresponding to my identification argument and present the Monte Carlo simulation performance. The estimator performs better when the job-wise unobserved heterogeneity is drawn from a discrete space rather than continuous.

[2]Firms' wage-setting behavior is exogenous, but I allow the wages or prizes to be related with their job characteristics in an arbitrary way. Thus, this model still captures some features of two-sided markets.

[3]Essentially, the valuation affects workers' profit of taking part in a job.

Chapter 2

Identification of Matching Games with Two-sided Unobserved Heterogeneity

2.1 Introduction

When a market explicitly consists of two disjoint sets (sides) of agents, and agents on both sides have preference over forming a relationship with agents on the other side, it is often referred to as two-sided matching markets. Empirical studies of matching markets have recently garnered considerable attention. Ever since the seminal work by Gale and Shapley (1962) and Shapley and Shubik (1972), researchers separately study two frameworks of matching games: Non-Transferrable Utility (NTU) framework if all the characteristics on both sides including potential monetary transfers are exogenously determined; and Transferrable Utility (TU) framework if monetary transfers are additively separable to one's utility and endogenously determined along with the matching allocation.[1] The choice between the two structures usually hinges on researchers' understanding of the institutional structure of the market. For instance, when analyzing school admission problems, it is inappropriate to adopt the TU framework, as tuition fees are usually pre-determined by the government or schools with little room for negotiation, if not completely nonnegotiable. This is the main reason why the NTU framework is adopted in college admission problems. Nevertheless, when doing so, one should be cautious when claiming tuition fees, similar to other school characteristics, are exogenous. In fact, if one believes that tuition fees are set strategically either by the government or by schools, then the underlying preferences over students will affect the observed distribution of tuition fees, leading to inconsistent estimation of model primitives if ignoring so. It is, therefore, important to study the determination process of monetary transfers even when they are not negotiable between two sides.

Another challenge in the empirical analysis of two-sided matching games is that the preferences resulting in the observed outcomes may depend on individual's features that are unobservable to the researcher. In the example of school admission problems, important factors influencing the observed matching outcomes include students' latent

[1] A recent paper by Galichon et.al. analyzes the Imperfect Transferable Utility (ITU) framework that is based on the TU structure but allows for non-additive transfers to agents' utility functions.

skills and schools' reputation, both of which are unlikely to be entirely captured by observables. Ignoring the existence of such unobserved heterogeneity will also impede the accuracy of model estimates to suggest any policy improvements further.

In this paper, I analyze a generalized Non-Transferrable Utility (NTU) matching game in which the monetary transfers are firstly determined by one side of the market (e.g. the employers), then the matching process occurs in a decentralized fashion in the second stage, taking the monetary transfers as given. I use information repeatedly observed for agent- and match-specific characteristics to nonparametrically identify both sides' unobserved heterogeneity, assuming employers share a vertical preference over employees. Preference primitives are further identified from solving this two-stage model, linking the observed distribution of market outcomes to employers' latent pricing strategies and employees' latent choices over multiple jobs. Afterward, I propose a likelihood-based estimation strategy that tackles the dimensionality issue emerging when the number of markets and/or players increases, and establish the consistency result of such an estimator.

Contribution of this paper:

First, from a theoretical perspective, the two-stage model in this paper combines the matching process with a wage-setting process beforehand, which generalizes the nontransferable utility (NTU) framework in two-sided matching games that assumes the complete exogeneity of monetary transfers in matching games. An explicit or implicit wage-setting stage is widely observed in markets without individually negotiable contracts but has been underexplored in the applied matching-theory literature. By incorporating this stage, I can explain in depth how monetary transfers are set strategically based upon jobs' features, and how it further determines the matching allocation of workers towards jobs. When some job and/or worker features are not observed by the researcher, characterizing the wage-setting process will further facilitate the identification of these unobserved types which will be discussed in details later.

Second, my model allows market players' preferences over the opposite side to be driven by both observed and unobserved features. Workers are furthermore allowed to have heterogeneous preferences over the types of jobs they favor. Heterogeneity in types and preferences is mostly detected in the empirical literature of online marketplaces but has rarely been modeled explicitly in literature, as identification issues are a central concern. In this paper, I obtain point identification through a multistep method in which the unobserved type distribution of both sides are initially nonparametrically identified from jointly observing worker- and match-specific characteristics for multiple periods, using a modified version of eigenvalue-eigenvector decomposition techniques based on measurement error models. Then workers' heterogeneous preference primitives are recovered from the market-level choice probability using the widely-studied discrete choice model. Lastly, firms' profit primitives are identified through the symmetric monotone strategy of their equilibrium wages against their job types under incomplete information. This identification strategy is novel in the structural literature estimating matching games with unobservables, and could be applied to various empirical settings with finite-sized markets. [2]

When estimating the generalized matching game using maximum likelihood, a general concern is the dimensionality issue when the number of players and/or markets increases. In the model, one important feature is that some players from one side of the market (i.e. workers) are repeatedly observed over time, which makes them "global players" of the repeated game. When this set of players increase, or the number of market increases, it requires a high-dimensional integration over global players' types to construct the likelihood function, which can be very computationally impractical using real data. In this paper, I suggest a likelihood-based estimator that maximizes a reconstructed likelihood function to reduce the dimensionality significantly, and it performs well in Monte Carlo simulation. This estimator could be adopted in other similar models (e.g. single agent discrete choice models) with a high dimension of unobserved heterogeneity and/or a significant number of players within a market.

Related literature:

From a methodological point of view, this paper contributes to the recently growing empirical literature analyzing preferences in two-sided matching markets when

[2] Alternatively, researchers assume an infinite number of players in one market (market being thick) and develop identification strategies accordingly. The thickness of the market is usually assumed in the current empirical literature, but this is not ensured: In online labor markets, transactions usually take the form of spot tasks, which can be sensitive both temporally and geographically. Examples of temporally sensitive tasks are the coding jobs at Topcoder.com. An example of a geographically sensitive work is a babysitting job on TaskRabbit.com.

unobserved heterogeneity. Modeling the role of monetary transfers in these markets is crucial for econometric specification and counterfactual experiments. When it is believed that monetary transfers are determined simultaneously with the equilibrium match, it falls into the TU framework, in which the researchers are only able to identify the one-dimensional match-specific output/surplus from data on observed matches. One strand of the literature manages to transform the matching problem into a general equilibrium problem by assuming agents have preferences over finite types of their counterparts and that the market is thick on both sides (Choo and Siow, 2006; Chiappori et al., 2006; Graham, 2013; Galichon and Salanie, 2012; Sinha, 2014).

Meanwhile, the *rank-order property* developed in Fox (2010) is a major econometric tool when preferences are over individuals, and requires that across a population of observationally equivalent identical markets, the matching allocation that yields a higher (deterministic part of the) surplus will be more frequently observed. While this assumption is intuitive economically, it is challenging to write an exact data-generating process under which it holds. Moreover, this property no longer holds when unobserved heterogeneity is introduced into the model. Fox and Yang (2012) show that without the rank-order property, in markets with match-specific unobserved heterogeneity, the distribution of unobserved heterogeneity could be identified following the special regressor method in the multinomial choice literature when researchers are able to observe markets where no agents are matched. In this paper, I adopt a similar intuition by assuming preferences are over individuals rather than types, and translate the matching problem to a well-studied discrete-choice problem by assuming firms admit a vertical preference over workers, but I do not depend on observing completely unmatched markets to achieve identification. Instead, by exploiting the distribution of endogenous wage contracts, along with repeatedly observing agents on one side across markets, I can uniquely recover the joint distribution of unobserved types of both sides using the eigenvalue-eigenvector decomposition. A more fundamental difference from Fox and Yang(2012) is that I adopt the NTU framework in my second stage model, wherein monetary transfers are treated as given and cannot be negotiated between the two parties.

Although there has been a stream of papers that estimate the preferences over various characteristics (see for example Logan et al., 2008; Boyd et al., 2013), two recent papers have shown that utility primitives are typically not identified from merely observing the joint distribution of characteristics of one-to-one matched pairs: Menzel (2015) shows that under the parametric assumption on idiosyncratic match-specific tastes, only the sum of individual surplus from a match is identified in a single large market. Agarwal and Diamond (2014) show that double-vertical preference in a large single market could be identified when econometricians can observe the joint distribution of (at least) two-to-one matching characteristics. To achieve identification of distributions of unobserved heterogeneity (or partially identifying some function of unobserved heterogeneity), both papers assume the unobserved components are independent of observables, inducing no systematic correlations in preferences among market players; furthermore, they rely crucially on market size going to infinity. This paper complements their work by showing that, by observing a one-dimensional match-specific outcome and worker-specific outcome, one could obtain point identification even with limited market size and one-to-one matching, as long as preferences on one side are vertical. This assumption – or, more generally, the aligned-preference assumption – ensures uniqueness of the equilibrium and is widely used in the literature. For instance, Agarwal (2014) makes the same one-sided vertical preference assumption when studying the national medical-resident placement market, a labor market with almost nonnegotiable wage contracts. Sorensen (2007) adopts a fixed sharing rule between matched pairs, which also leads to aligned preference, and identifies utility primitives from exclusion restrictions across different markets with match-specific idiosyncratic components. In this paper, I explicitly model the contract setting stage before the matching process, which is not captured in any of the papers listed above.

An alternative strand of empirical studies dodges the discussion of point identification of utility primitives. Instead, it estimates a set of parameters consistent with the pairwise-stability notion (Baccara et al., 2012; Uetake and Watanabe, 2014). In this paper, I make a stronger assumption on the preference structure to obtain a unique equilibrium outcome, as point identification facilitates the computation of counterfactuals while the identified set may not be sharp.

The rest of the paper is organized as follows: Section 2 establishes the structural model and characterize the equilibrium for the Generalized Non-Transferable Utility matching game. Next, in Section 3, I present the main identification and estimation results that nonparametrically recover the unobserved heterogeneity and parametrically recovers the utility primitives. Also, I provide a likelihood-based estimation procedure implementing my identification method. Following that, Section 4 presents the Monte Carlo results that confirm the validity of my estimation strategy. Lastly, I conclude this paper in Section 5.

2.2 A Structural Model for the Generalized NTU Matching Games

I now establish the structural model for the generalized matching game. Each market consists of two sides of players, denoted as firms and workers seeking for a job. They gather to match and form an employment relationship. In the market, each firm carries one job needed to be fulfilled but can attract multiple applications, and each potential employee can be matched with at most one job in a market; across different markets, however, there is no restriction on the number of jobs a worker can take. As the researcher, I observe three sets of information: on one side of the market, I observe the monetary transfers proposed by each employer to their potential employees; on the other side, I observe a one-dimensional "index" for each potential employee, and this index provides partial information about their ability levels. Lastly, I observe the allocation of which person ends up with which employer. Conditional on other observed information, the complexity level of each job carried by an employer, along with the latent ability level of each employee and the market-wise demand condition, is unobservable to me. I will focus on these layers of unobserved heterogeneity and abstract away observable heterogeneity of both sides of the market. It could be viewed as the model is built for each subpopulation that shares the same observable characteristics.

2.2.1 The Timeline

The market game consists of two stages in each market t. At Stage 1, a finite number J_t of employers randomly sign up to be present in the market. They have common prior about the distribution of peer jobs' complexity and potential employees' utility profiles, but cannot observe other employers' types nor the actual identity of participating employees. Besides, they all observe the market-wise demand condition, ω_t, that commonly affects their profits.[3] By the end of this stage, employers simultaneously determine the monetary transfers they will grant to their potential employees. They are allowed to recruit multiple people but are faced with a budget constraint that is commonly known to all peer employers.

At Stage 2, a finite number I_t of workers are randomly drawn to enter the market.

[3] An example that determines ω_t is the cost efficiency of recruiting people via alternative platforms.

Perfectly observing the characteristics and utility profiles of every job and every peer job-hunter, they simultaneously choose the job they would like to work on. Due to capacity constraint, they are allowed to sign up for at most one job in a market; on the other side, each firm can hire up to Q_{jt} workers to work on the same project. After signing-up, employees simultaneously work on the project and deliver their completed job by the due date. Firms then hire a third-party reviewer board to rank order all submissions based on their quality. The market game finishes when workers get cash payment according to their rank order. In this model, I assume away strategic entry and exit decisions on the firm side. This is a significant simplification that enables us to focus on the strategic pricing behavior of firms and discrete choices of workers in the market, which I believe are essential to recover for policy implications.

Next, I define the preference structure of players from both sides. There is an abuse of notation in the subsequent discussion: I_t (resp.J_t) denotes both the number of employees (resp. employers) in market t, and the set of employees (resp. employers) in market t.

2.2.2 Preferences of Firms

Firm j's *ex post* payoff from recruiting a set of workers $\mu_t(j)$ to complete their posted job in market t is modeled as:

$$\pi^{post}_{j,\mu_t(j),t} = \mathcal{R}(\omega_t, Z^*_j, \sum_{i\in\mu_t(j)} Z_{ij}) - \sum_{i\in\mu_t(j)} P_{ij} \tag{2.1}$$

where $\mathcal{R}$ is the revenue function that depends on three elements: (i) the job-wise complexity Z^*_j; (ii) the sum of Z_{ij}, which denotes the quality of worker i's submission to firm j, [4] and (iii) the demand-side shifter, such as the cost efficiency of recruiting via outside options, that affects the overall willingness to pay of all firms in the market, and is denoted by ω_t.

In particular, the quality of a submission is determined by the equation below:

$$Z_{ij} = g_2(X^*_i, Z^*_j, b_j), \tag{2.2}$$

where g_2-function is increasing in worker i's ability, X^*_i, and decreasing in the complexity of a job, Z^*_j. The idiosyncratic term b_j is independent of (X^*_i, Z^*_j) and is realized only after the employees complete their job. [5] A possible factor included in b_j is the bias when the third-party reviewer board evaluates the submissions.

[4] Here I assume all submissions enter firms' profit function uniformly; instead, I can assume non-winning submissions add partial credits to the firm's profit. The core idea is that, without considering cost, firm strictly prefers having extra submissions.

[5] This is to exclude the case where firms know *ex ante* the value of b_j when they decide the cash prizes at Stage 1.

In this paper, I assume the revenue function is known to the researcher up to a one-dimensional coefficient, c:

$$\mathcal{R}(\omega_t, Z_j^*, \sum_{i \in \mu_t(j)} Z_{ij}) = c \cdot \mathcal{R}_0(\omega_t, Z_j^*, \sum_{i \in \mu_t(j)} Z_{ij}) \tag{2.3}$$

where $\mathcal{R}_0$ is a known function. Lastly, the rule of how each employee gets paid according to their rank-order is also observed by the researcher. Specifically,

$$P_{ij} = \begin{cases} P_j, & \text{if } i \text{ is the first place} \\ \phi_1 P_j, & \text{if } i \text{ is the second place} \\ \phi_2 P_j, & \text{if } i \text{ is the third place} \\ \cdots & \\ \phi_{Q_{jt}} P_j, & \text{if } i \text{ is the } Q_{jt} \text{ place} \end{cases}$$

where $\phi_k > \phi_m, \forall k < m, k, m = 1, 2, \cdots, Q_{jt}$; $\phi_k, \phi_m \in [0, 1)$ and is known to the researcher.

2.2.3 Preferences of Workers

Following the single-agent discrete choice literature, I assume the indirect utility of employee i choosing firm j in market t is determined by the following equation:

$$u_{ijt} = \beta P_{ij} + \tilde{u}_{it}(X_i^*, Z_j^*) + \nu_{ijt} \tag{2.4}$$

where β captures one's vertical preference over cash prizes, and $\beta \geq 0$. The function $\tilde{u}_{it}$ is a worker's heterogeneous preference over the complexity of a job, and may interact with their own skills. Lastly, ν_{ijt} is the idiosyncratic taste shock that is independent of other variables. I use $j = 0$ to denote the choice of staying outside the market.

Formally, this Extensive Form game is defined as follows:

Definition 1 *The two-stage market game is described by* $(I_t, J_t, \mathbb{Z}, \mathbb{X}, \mathbb{P}, \mu_t, \Omega, F_\omega, \pi, u)$, *where*

1. I_t *is the set of workers, and* J_t *is the set of jobs posted by employers.*
2. $\mathbb{Z}$ *is the support of the complexity of jobs;* $\mathbb{X}$ *is the support of employees' skill levels.*
3. $\mathbb{P}$ *is the support of cash payments of all jobs.*
4. μ_t *is the matching allocation such that if employee* i *chooses job* j, *then* $\mu(i) = j$ *and* $\mu(j) = i$. *If employee* i *remains unmatched, then* $\mu(i) = \emptyset$*; if no one chooses*

job j, then $\mu(j) = \emptyset$.

5. *Ω is the support of market unobservables that affects all firms' profits.*

6. *π is the firms' profit function and is defined in equation (3.1).*

7. *u is the workers' indirect utility function and is defined in equation (2.4).*

2.2.4 Equilibrium Notion and Characterization

After defining the game and players' preferences, I now define a rational expectation equilibrium notion for this extensive-form game.

Definition 2 *The rational-expectation equilibrium (δ_t^*, μ_t^*) is such that: At Stage 2, for any observed $(\boldsymbol{P}_t, \boldsymbol{Z}_t^*)$, the matching allocation $\mu_t^* : I \cup J \to I \cup J \cup \emptyset$ is pairwise stable; At Stage 1, given the rational expectation about the stable matching function and the knowledge of distribution of $(\boldsymbol{X}_t^*, \boldsymbol{Z}_t^*)$, firms play the mixed-strategy pricing function $\delta_t^* : \mathbb{Z} \to \Sigma$ which is Bayesian Nash Equilibrium strategy.*

The equilibrium concept defined in the second-stage market is the well-known *pairwise stability* notion introduced by Gale and Shapley (1962) and generalized in Roth and Sotomayor (1989). An observed μ is said to be *pairwise stable* if it satisfies:

1. (Individual Rationality) $u_{i,\mu(i),t} \geq u_{i,\emptyset,t}, \forall i \in I_t$ and $|\mu_t(j)| \leq Q_t$, $\pi^{post}_{j,\mu_t(j),t} \geq \pi^{post}_{j,\mu_t(j)\backslash i,t}, \forall j \in J_t, i \in \mu_t(j)$.

2. (Nonblocking Pairs) For any employee i and firm j such that $j \neq \mu_t(i)$, the following situations cannot happen simultaneously: $u_{i,j,t} > u_{i,\mu(i),t}$; and $\pi^{post}_{j,\mu_t(j)\cup i,t} > \pi^{post}_{j,\mu_t(j),t}$ if $|\mu_t(j)| < Q_t$; or $\pi^{post}_{j,\mu_t(j)\setminus i' \cup i,t} > \pi^{post}_{j,\mu_t(j),t}, \exists i' \in \mu_t(j)$ if $|\mu_t(j)| = Q_t$.

The first condition implies that the matching allocation I observe is at least as desirable for all firms and workers as staying unmatched. The second condition means that, for any worker i in the market, his/her current choice $\mu_t(i)$ is the most desirable job in his/her choice set. This choice set consists of any firms that are willing to swap their current matched employees with i, or to fulfill a vacant space with i.

At Stage 1, expecting employees will behave on the equilibrium path in the following stage, and conditional on the prior knowledge of the joint distribution of workers' abilities and peer firms' job complexity, along with the distribution of workers' idiosyncratic taste shock and the idiosyncratic shock to the submission quality, the mixed-strategy Bayesian Nash Equilibrium is defined as a mapping $\delta^*_t : \mathbb{Z} \rightarrow \Sigma$, where $\Sigma := \{\delta | \sum_{l=1}^m \delta^l = 1\}$, such that for each firm j, given other firms' equilibrium strategy, $\delta^*_{-j,t}$, and the correct belief for the second stage, the following inequality holds:

$$\begin{aligned}
&\sum_{l=1}^{m}[\sum_{\boldsymbol{Z}^*_{-j},\boldsymbol{X}^*} Pr(\boldsymbol{Z}^*_{-j},\boldsymbol{X}^*)\pi^{\text{int}}(p_l, Z^*_j, \boldsymbol{Z}^*_{-j}, \boldsymbol{X}^*, \boldsymbol{\delta}^*_{-j,t}(\cdot),\omega)] \cdot \delta^{l,*}_{j,t} \geq \\
&\sum_{l=1}^{m}\sum_{\boldsymbol{Z}^*_{-j},\boldsymbol{X}^*} Pr(\boldsymbol{Z}^*_{-j},\boldsymbol{X}^*)\pi^{\text{int}}(p_l, Z^*_j, \boldsymbol{Z}^*_{-j}, \boldsymbol{X}^*, \boldsymbol{\delta}^*_{-j,t}(\cdot),\omega)] \cdot \delta^{l}_{j,t}, \forall \delta \in \Sigma. \qquad (2.5)
\end{aligned}$$

where π^{int} denotes the interim payoff for firms j that chooses cash prize P_j and believing other firms will play the mixed strategy $\delta^*_{-j,t}$:

$$\pi^{\text{int}}(p_l, Z^*_j, \boldsymbol{Z}^*_{-j}, \boldsymbol{X}^*, \boldsymbol{\delta}^*_{-j,t}(\cdot), \omega) =$$

$$\sum_{\mu_t} Pr(\mu_t | \boldsymbol{Z}^*_{-j,t}, \boldsymbol{X}^*_t, Z^*_j, p_l, \delta_t(\boldsymbol{Z}^*_{-j,t}))\pi^{post}(Z^*_j, \omega_t, \boldsymbol{X}^*_t, \mu_t, p_l) \tag{2.6}$$

In the subsequent paragraphs, I characterize the rational-expectation equilibrium that leads us to the identification results in Section 3 in a backward fashion.

At Stage 2, the matching stage, it is well known from theory literature that a pairwise stable outcome always exists when preferences are responsive (Roth and Sotomayor, 1989). Uniqueness is ensured through the assumption that workers have strict preferences over slots, and all slots agree upon the vertical and strict ranking over workers (Clark, 2006; Niederle and Yariv, 2009). The formal result is presented in the following lemma.

Lemma 1 *If firm j's profit under matching allocation μ_t is defined as in equation (3.1), then firms' preference over set of employees is responsive to the preference over individual employee. Furthermore, when preferences are strict on both sides, the stable matching μ exists and is unique.*

The proof of Lemma 1 uses the well-known Gale-Shapley Deferred Acceptance Algorithm(DAA) to find the unique matching allocation. Nonetheless, as this game proceeds without multiple rounds of making, holding and rejecting offers as suggested in DAA, the question arises naturally: would the observed outcome still obtain pairwise stability? The answer is yes. Consider the following non-cooperative game:

In the market firms simultaneously announce their jobs, each assigned with a cash payment schedule and offers multiple (but finite) slots, each worker then simultaneously apply for a job , or simply exit the market; then, each firm accepts or rejects received applications if there is any; lastly, if the application is accepted, the worker and the job slot are matched, otherwise workers exit the market remaining single, and the corresponding slots remain vacant. Matched workers then complete the job as requested.

It is shown by Proposition 2 in Niederle and Yariv (2009) that under complete information assumption, meaning that all players in the market are fully informed of the utility profile, the stable match is the unique Nash equilibrium outcome surviving iterated elimination of weakly dominated strategies. Therefore, if we restrict our attention to players' rationalizable equilibrium strategies only, the non-cooperative game's rationalizable equilibrium are all pairwise stable.

After establishing the equivalence mapping from the non-cooperative game to the matching framework, I would like to characterize pairwise stability in such a way that relates to the single-agent discrete choice problem. By the proof of Lemma 1, we can see that the skill level of a worker directly affects the cardinality of his choice set. For instance, the most skilled worker, with the knowledge of being the best in the market enjoys a very rich choice set containing the slots of all jobs in the market. On the contrary, the choice set of the least-skilled one would be very limited, as in equilibrium, those slots favored by better workers would never be available to him. In addition, only higher-ranked workers' preferences will affect worker i's choice set, but not the lower-ranked ones. As a result, for each worker i, we can characterize his choice set conditional on the stable matching μ in market t as:

$$\begin{aligned} M_i[\mu_t] := \{j^1 | j \neq \mu(l), \forall X_l^* > X_i^*, \pi_{j,i,t}^{post} \geq 0\} \cup \{j^k | \exists l_m, m = 1, 2, \cdots, k-1, \\ \text{such that } \mu(l_m) = j, X_{l_m}^* > X_i^*, k \leq Q_t, \pi_{j,l_m \cup i,t}^{post} \geq \pi_{j,l_m,t}^{post}\} \cup \emptyset \end{aligned} \tag{2.7}$$

where the first part of the RHS denotes the first-place of all jobs that are not chosen by any of the better workers and are willing to hire worker i than letting the slot remain vacant, and the second part denotes the highest slot of each job that is not fulfilled by a better worker and the firms has to be willing to hire worker i than letting the slot remain vacant. Through this characterization, I can rewrite the pairwise stability condition into a series of single-agent optimal choices over heterogeneous choice sets:

Lemma 2 *The match μ_t is pairwise stable if and only if $u_{i,\mu(i)} \geq \max_{m \in M_i[\mu_t]} u_{i,m}, \forall i \in I_t$, where $M_i[\mu_t]$ is defined in equation (3.7).* [6]

The proof is in the appendix. Lemma 2 implies that as long as we know the actual skills of all potential employees in market t, the pairwise stable outcome degenerates to a discrete choice problem, in which workers make their discrete choices sequentially. Firms' preferences are embedded in the skill ranking of workers, which affects the

[6] Menzel (2015) characterizes the equilibrium using a similar argument here. His specification is more general, though, as he includes all possible preference structure for both sides.

sequence of workers' moves. This lemma holds as long as one side of the market admits a vertical preference over the other; therefore, a symmetric result could be obtained if workers have vertical preference over jobs yet firms hold heterogeneous preferences over workers.

Next, to characterize first-stage Bayesian Nash Equilibrium (BNE), recall that according to Harsanyi (1967), the incomplete information game is equivalent to a complete information game with $k \cdot J$ "shadow firms," and these shadow firms belonging to the same job will be randomly selected by Nature to be present at Stage 1 after they set their equilibrium strategy. The existence of a mixed-strategy BNE is therefore ensured by Nash's theorem. Formally,

Lemma 3 *If both the support of submission qualities $\mathbb{Z}$ and the support of cash prizes $\mathbb{P}$ are finite, then the game defined above has at least one mixed-strategy Bayesian Nash Equilibrium.*

The proof is provided in the appendix. The next proposition summarizes the existence result for the complete two-stage game:

Proposition 1 *There exists a rational-expectation equilibrium (μ_t^*, δ_t^*) for the game defined in Definition (1).*

I already proved the uniqueness of the second-stage subgame equilibrium; Nonetheless, the uniqueness of the whole game cannot be assured with further restrictions. For simplicity, I assume:

Assumption 1 (Equilibrium Selection Rule) *The same symmetric equilibrium mixed-strategy δ^* is uniquely played in the data.*

Although due to the complexity of firms' interim payoff functional form at Stage 1 [7], I cannot derive the closed-form solution for the mixed-strategy equilibrium cash prize, there are some characterizations that suffice for identification in the subsequent section. In particular, I show the following lemma holds:

Lemma 4 *In markets where all firms fill up their vacancy, the probability of observing a firm setting the cash prize $p_m = \max \mathbb{P}$ monotonically increases with the market-level unobservable ω and the cost coefficient c.*

Proof can be found in the appendix. So far, I have established results regarding the existence and characterization of the rational expectation equilibrium, and the next section will move to the econometric discussion for recovering the primitives.

2.3 Identification

In this section, I discuss under which conditions I can use the observed data patterns from many finite-sized markets to fully recover of the underlying utility and profit parameters through firstly identifying the unobserved worker-wise and job-wise heterogeneity. Before that, let me clarify the data generating process for this game here:

[7] The complexity arises as the interim payoff integrates out distribution of all potential workers' types preferences and taste shocks, therefore difficult to solve.

2.3.1 Data-Generating Process and Notation

Suppose, as the econometrician, I can observe T markets with I_t workers and J_t firms in market t. Without loss of generality, assume each firm j offers Q_{jt} slots such that $\sum_{j_t} j_t \times Q_{jt} = I_t$. Once a firm is randomly drawn in market t, it leaves the population and will not be selected in future markets. On the other hand, workers enjoy a positive (and exogenous) probability of re-entering a new market in the game.

Every worker carries a latent ability level, X_t^* that is discrete and evolves over time according to some underlying exogenous law, $Pr(X_t^*|X_{t'}^*)$ if the worker appears in markets t' and t consecutively. Each firm carries one job with a difficulty level Z^*, also discrete. Additionally, each market t carries a one-dimensional unobserved heterogeneity, ω_t that affects all firms' willingness to pay. All Q_{jt} slots within a job j share the same characteristics, $\boldsymbol{W}_j$, including the unobserved type Z_j^*, but grant different cash payment $[P_j, \phi_1 P_j, \phi_2 P_j, \cdots, \phi_{Q_{jt}} P_j]$, respectively.

To summarize, for each market t, I observe a $I_t \times 1$ vector of worker characteristics, X_t; a $J_t \times 1$ vector of firm/job characteristics, W_t (including the cash payments, P_t); and a matching allocation μ_t such that if worker i choose slot q in firm j's job, then $\mu(i) = k_j^q$ and $\mu(k_j^q) = i$. The goal of identification consists of three items: (1) The joint distribution of workers' skill levels and jobs' complexity levels, $(\boldsymbol{X}_t^*, \boldsymbol{Z}_t^*)$; (2) workers' utility primitives in $u_{i,j,t}$; and (3) firms' *ex post* profit primitives in $\pi_{j,i,t}^{\text{post}}$ and market-level unobserved heterogeneity, ω_t.

In the following discussion, I suppress the market subscript t for ease of notation.

Also, I use bold letters to denote market-wise vectors of characteristics, and regular letters to indicate individual-level characteristics. The letter f is used to denote the probability mass/density function of any distribution, with a little abuse of notation; and M is used to indicate the matrix representation for any discrete distribution.

The identification strategy I have developed includes three steps: in the first step, I derive the nonparametric identification result for the two-folded unobserved heterogeneity; following that, I identify the utility parameters on the worker side; lastly, the firm side primitives are identified in a semi-parametric fashion. The subsequent sections provide detailed discussion on each step.

2.3.2 Step 1: Identification of $(\boldsymbol{X}^*, \boldsymbol{Z}^*)$

In the first step, within each market, conditional on observing pairwise stable matches μ, I identify the joint distribution of $(\boldsymbol{X}_t^*, \boldsymbol{Z}_t^*)$ from jointly observing worker-specific, firm-specific and match-specific characteristics. The identification strategy is based on the eigenvalue-eigenvector decomposition technique developed in Hu (2008). The basic idea is that, to fully recover the distribution of a latent variable, I need at least three sets of useful information, all of which are correlated with the latent variable in any arbitrary ways, but are independent of each other conditional on the latent variable. In particular, one of them can be as simple as a binary variable. In this model, the complication lies in the fact that the latent factor of interest is a vector consisting of two parts: worker skills and job complexity. One extra feature unique to the matching model is that the distribution of observables from both sides is correlated with the market matching allocation. For instance, if the econometrician observes a significant portion of workers having relatively good ratings for their skills, but submit jobs of very poor quality, how could he infer the underlying distribution of skills among the population labor force? One possibility is that: posted skill ratings are very noisy, and it is in fact very rare for a worker to be highly skilled. However, one other possibility is that ratings are informative, and the main reason why we see poor quality submissions are due to the complexity of jobs matched to the workers. Therefore, without conditioning on the matching allocation, the econometrician can never truly identify the distribution of workers' skills. We can make a similar argument on the job-side unobserved heterogeneity. Next, I show in details how one can construct the set of information sufficient for identification.

2.3.2.1 The Job Evaluation System: A Match-Specific Outcome

First, like mentioned before, one important slice of information observed by the econometrician is the quality of each completed job, or more broadly, the match-specific outcome. In empirical studies, the match-specific outcome is often observable and can take many forms. An example in labor markets is the job evaluation system. In the matching between venture capitalists and companies for investment, this outcome could be the IPO stock price. The evaluation result provides a cardinal measurement to (1) rank order workers' performance within the same job, thereby determining the monetary transfers; and it also reflects (2) the productivity of any matched pair.

I already modeled the quality of submission in the previous section using equation (2.2); here in order to capture the two features above, I model it more specifically with regard to how the match-specific outcome are determined:

Assumption 2 *Let $Z^t_{i,j}$ represent the match-specific outcome generated from a matched pair (i, j) in market t, then,*

$$Z^t_{i,j} = g_2(X^*_{it}, Z^*_{jt}, b_{jt}),$$

where,

1. *g_2-function is continuous and monotonically increasing with respect to its first argument ($\partial g_2/\partial X^* > 0$).*

2. *the idiosyncratic part b_{jt} is independent of (X^*_{it}, Z^*_{jt}).*

3. *Conditional on observed matches μ_t,*

    (a) *For any two firms j and l in market t, $b_{jt} \perp b_{lt}$.*

    (b) *For any worker i that is matched with j in market t, $b_{jt} \perp (X^*_{it}, a_{it})$.*

    (c) *For any firm j in market t, $b_{jt} \perp (Z^*_{jt}, \omega_t)$.*

(d) For any two markets t and t', $\boldsymbol{b}_t \perp \boldsymbol{b}_{t'}$.

Assumption 2 (i) implies that the more skilled a worker is, the better the match-specific outcome would be, fixing other factors. This is a reasonable assumption in real life. Assumption 2 (ii) -(iii) implies that idiosyncratic shocks to the match-specific outcome are independent of other observable and unobservable characteristics. This is a crucial assumption to ensure identifiability, and is more general than it seems to be – I do not restrict the functional form of g_2 except for Assumption 2 (i), so that the idiosyncratic shock could interact with other factors in an arbitrary way.

Next, I exploit observed one extra slice of information on the worker side to construct more moments related to the latent types from two sides.

2.3.2.2 The One-Dimensional Index for Workers

As mentioned in the modeling section, one set of observable information is a one-dimensional index for each worker in each market, aiming to proxy their skill levels. For instance, in school admission problems, this index can be students' SAT or other standardized test scores. The major difference between this index and the match-specific outcome is that, the former is revealed *before* the matching process whereas the latter is generated *after* the matching. I denote this index for worker i in market t as X_{it}, and assume it is initially generated from a finite space for each worker, and is updated in subsequent markets by the following equation:

$$X_{it} = g_1(X_{i,t-1}, Z^*_{\mu_t(i),t}, X^*_{it}, \boldsymbol{X}_{\mu_t(\mu_t(i)),t-1}, a_{it}), \tag{2.8}$$

if he participates in both market $t-1$ and t consecutively. If not, the index remains unchanged until the next time he appears in a market. In the equation above, $X_{i,t-1}$ is the last-period index; $Z^*_{\mu_t(i),t}$ is the complexity of the job he completes in the market; X^*_{it} is the latent skill level that evolves according to an exogenous rule: $Pr(X^*_{it}|X^*_{i,t-1})$. Next, $\boldsymbol{X}_{\mu_t(\mu_t(i)),t-1}$ are the peer workers' previous skill indexes within a job – including this factor allows one's skill index to be contingent on the relative performance to others in the same market. Lastly, a_{it} is an idiosyncratic shock specific to the individual worker in the market and is assumed to be independent of the other factors. Formally, I assume:

Assumption 3 *The function g_1 is invertible with respect to its first argument. Furthermore, it also satisfies that $\frac{\partial Pr(X_{it}<X_{i,t-1})}{\partial Z^*_{\mu_t(i),t}} > 0$ when fixing other factors in g_1 function. The idiosyncratic shock a_{it} is realized after workers submit their jobs, and satisfies: (i) $a_{it} \perp X^*_i|\mu_t$; (ii) $a_{it} \perp a_{i,t'}$ for any two different markets t and t'; (iii) $a_{it} \perp a_{kt}|\mu_t$ for any two workers i and k in market t; and (iv) $a_{it} \perp \omega_t$.*

Compared with the worker side, it is sometimes difficult to observe any explicit information to index the types (more precisely, the complexity levels) of jobs. Good news is that I do not need as much information on the firm side as for the worker side. In fact, a binary indicator that carries some variation in job types suffices for identification.

By revisiting Assumption (3), we can see that the pattern of a worker's ability index over time (i.e. markets) can tell us how complex the job he completes in the current market might be. In particular, the binary indicator: $\mathbb{1}(X_{it} < X_{i,t-1})$ is monotonically increasing with the complexity level of the job he completes. This, therefore, serves as a binary indicator related with job complexity levels. [8]

Specifically, let Y_j denotes the binary indicator, equal to one if the (first-place) worker's index decreases in market t and zero otherwise. Then, I make one more assumption here:

Assumption 4 *Given Z^*, the distribution of Y is independent of both firms' profit shocks and shocks to the match-specific outcome. Furthermore, Y_j is excluded from workers' utility functions.*

This assumption restricts the potential correlation between the evolution of worker's skill index and the idiosyncratic shock in match-specific outcomes in its first part. This is reasonable if we believe that *ex post* match-specific outcome do not affect the evolution of one's ability index directly, but only through the underlying types of workers and jobs, along with other observables. Furthermore, I exclude the situation where workers explicitly care whether their indexes will increase or not after they play the game in market t on top of all other characteristics of the jobs. This is a

[8]Note that the direction could be the other way – $\frac{\partial Pr(X_{it}<X_{i,t-1})}{\partial Z^*_{\mu_t(i),t}} < 0$. The inequality sign depends on the researcher's understanding of how workers' ability indexes can be affected by the type of jobs they complete.

strong exogeneity assumption but is often assumed in the literature.

Given all the information we have explored, the next lemma formally establishes the conditional independence result regarding observed characteristics. To be more accurate, we return to the explicit market subscript t.

Lemma 5 *Under Assumptions 3–4, from observing many markets repeatedly, the following condition holds for the market-level observables.*

$$(\boldsymbol{X}_t | \boldsymbol{Y}_t) \perp (\boldsymbol{P}_t, \tilde{\boldsymbol{Z}}_t) \perp \boldsymbol{Z}_t | (\boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t),$$

where $\boldsymbol{Y}_t$ denotes the $J_t \times 1$ vector of binary variables each indicating whether the ability index of the best worker within a job decreases; $\boldsymbol{X}_t$ represents the $I_t \times 1$ vector of workers' contemporary ability index; $\boldsymbol{P}_t$ represents the $J_t \times 1$ vector of cash payment for all jobs; $\tilde{\boldsymbol{Z}}_t$ indicates the $I_t \times 1$ vector of all workers' match-specific outcome up to their most recent participation; $\boldsymbol{Z}_t$ is the $I_t \times 1$ vector of contemporary match-specific outcomes for all workers; $(\boldsymbol{Z}_t^*, \boldsymbol{X}_t^*)$ is the $I_t \times J_t$ matrix of all jobs' and workers' types; and $\tilde{\boldsymbol{X}}_t$ is the $I_t \times 1$ vector of all workers' ability indexes up to their most recent participation. Lastly, μ_t is the observed match in market t.

Intuitively, for any individual worker i, we have three conditionally independent pieces of information related to their ability level: current index X_t, current and previous match-specific outcomes $(Z_t, \tilde{Z}_t))$. On the job/firm side, we are unable to observe information for the same job across different markets as each firm appears in only one market. Instead, we must rely on the information observed from the matching outcomes in *one* market to invoke identification. So far within a market, there are three slices of useful information related to job-wise heterogeneity: firms' cash payment, the match-specific outcome, and whether the skill indices of the best workers in each job decreases or not. To see this, first, let us focus on the match-specific outcomes, as they not only reflect workers' innate abilities, X^*, but also the complexity of jobs, Z^*. We must be cautious here, though, as the observed outcomes are driven by the matching allocation, through which jobs in the same market are implicitly correlated. Another bit of useful information concerns the cash payment decided by firms: They are the major movements firms make during the extensive-form game and are equilibrium outcomes. Again, cash payment from different firms is correlated through the market-level unobservable, ω, and further correlated through the matching allocation, μ.[9] Consequently, we cannot separately identify its distribution without looking at other jobs in the same market. Instead, the conditional independence could only be built upon observed (market-level) matching allocations.

The major implication of Lemma 5 is that, suppose we could observe infinitely many markets. Then, fixing the market size (I, J, Q) that is suppressed from the following equations, the matching allocation, and workers' previous performance, $(\mu, \tilde{\boldsymbol{X}})$, we could decompose the joint distribution of $(\boldsymbol{X}, \boldsymbol{P}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z}|\boldsymbol{Y})$ that is directly

[9]To see this, note that the probability of observing a certain match μ depends on the joint distribution of $(\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{P})$. Thus, the distribution of $\boldsymbol{P}$ is variant with different values of μ.

computable from data as

$$
\begin{aligned}
& f(\boldsymbol{X}, \boldsymbol{P}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z} | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \mu, I, J, Q) \\
&= \sum_{\boldsymbol{Z}^*, \boldsymbol{X}^*} f(\boldsymbol{P}, \tilde{\boldsymbol{Z}} | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}, \mu) \cdot f(\boldsymbol{Z} | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}, \mu) \cdot f(\boldsymbol{X}, \boldsymbol{Z}^*, \boldsymbol{X}^* | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \mu) \qquad (2.9)
\end{aligned}
$$

as well as,

$$
\begin{aligned}
& f(\boldsymbol{P}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z} | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \mu_t, I, J, Q) \\
&= \sum_{\boldsymbol{Z}^*, \boldsymbol{X}^*} f(\boldsymbol{P}, \tilde{\boldsymbol{Z}} | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}, \mu) \cdot f(\boldsymbol{Z} | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}, \mu) \cdot f(\boldsymbol{Z}^*, \boldsymbol{X}^* | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \mu) \qquad (2.10)
\end{aligned}
$$

Next, I formalize the distributional assumptions with respect to both observed and latent variables in the market:

Assumption 5 *1. Each job omplexity is drawn from the finite space* $\{z_1, z_2, \cdots, z_m\}$.

2. *Each worker's, i, ability is drawn from the finite space* $\{x_1, x_2, \cdots, x_l\}$.

3. *The match-specific outcome,* Z_{ij}*, is drawn from a bounded atomless support* $[0, 100]$.

4. *The cash payment is drawn from the finite space* $\{p_1, p_2, \cdots, p_M\}$ *with* $M \geq m$.

5. *The ability index,* X_{it}*, is drawn from an arbitrarily large but discrete support* $\{\widetilde{x_1}, \widetilde{x_2}, \cdots, \widetilde{x_L}\}$.

As all latent types are discrete, we would like to write equations (3.25) and (3.26) in matrix forms. The following assumption is crucial.

Assumption 6 *For any observed* $(\mu, \widetilde{\boldsymbol{X}} = \boldsymbol{x}, I, J, Q)$,

1. *There exists a mapping* $\phi_X : \{\widetilde{x_1}, \widetilde{x_2}, \cdots, \widetilde{x_L}\} \to \{x_1, x_2, \cdots, x_l\}$ *such that for any job* j*:* $\forall \boldsymbol{x}, \boldsymbol{x'} \in \{\widetilde{x_1}, \widetilde{x_2}, \cdots, \widetilde{x_L}\}^{|\mu(j)|}, \boldsymbol{x} \neq \boldsymbol{x'}$, *and for any observed* $\tilde{\boldsymbol{X}}_{\mu(j)}$,

$$Pr(\boldsymbol{X}^d_{\mu(j)} = \boldsymbol{x} | \boldsymbol{X}^*_{\mu(j)} = \boldsymbol{x}, \tilde{\boldsymbol{X}}_{\mu(j)}) > Pr(\boldsymbol{X}^d_{\mu(j)} = \boldsymbol{x} | \boldsymbol{X}^*_{\mu(j)} = \boldsymbol{x'}, \tilde{\boldsymbol{X}}_{\mu(j)}),$$

   where, for any worker i, $X^d_i = \phi_X(X_i)$.

2. *There exists a mapping* $\phi_P : \{p_1, p_2, \cdots, p_M\} \to \{1, 2, \cdots, m\}$.

3. *There exists a mapping* $\phi_{z1} : [0, 100] \to \{1, 2, \cdots, l \cdot m\}$ *for each match-specific outcome,* $Z_{j,1}$*, of the best workers within each job*

4. *There exists a mapping* $\phi_{z2} : [0, 100] \to \{1, 2, \cdots, l\}$ *for each match-specific outcome of the non-winning workers* $Z_{j,-1}$ *and a vector of values* $\boldsymbol{y}$*, such that the following matrix is of full rank* $m^J \cdot l^I$*:*

$$M_{\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d, \boldsymbol{Z}^d | \boldsymbol{Y} = \boldsymbol{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu_t, I, J, Q}. \tag{2.11}$$

   where for each non-winning worker i, $Z^d = \phi_{z2}(Z_i)$ *and* $\tilde{Z}^d = \phi_{z2}(\tilde{Z}_i)$*; for each winning worker* i, $Z^d = \phi_{z1}(Z_i)$ *and* $\tilde{Z}^d = \phi_{z1}(\tilde{Z}_i)$*; and for each job* j, $P^d = \phi_p(P_j)$.

The intuition behind Assumptions 12 and 13 is that, when (X^*, Z^*) follows a discrete distribution, we could partition (or discretize, if observables are continuous) observable characteristics to have the same dimension as the cardinality of the support for (X^*, Z^*), such that equations (3.25) and (3.26) could be written in a matrix form and would provide the nice property of invertibility. Assumption 13(1) implies that we could partition the ability indexes in a way that gives us a good indicator of the true ability level.

Specifically, fixing a certain value of $(\boldsymbol{Y} = \boldsymbol{y}, \tilde{\boldsymbol{X}}, \mu_t, I, J, Q)$, I compute the probability of observing $\boldsymbol{X}^d = \boldsymbol{x}$ as well as various values of $(\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d, \boldsymbol{Z}^d)$

$$\begin{aligned} & M_{\boldsymbol{X}^d=\boldsymbol{x}, \boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d, \boldsymbol{Z}^d | \boldsymbol{Y}=\boldsymbol{y}, \tilde{\boldsymbol{X}}=\boldsymbol{x}, \mu, I, J, Q} \\ & = M_{\boldsymbol{Z}^d | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}=\boldsymbol{x}, \mu, I, J, Q} \cdot D_{\boldsymbol{Z}^*, \boldsymbol{X}^*, \boldsymbol{X}^d=\boldsymbol{x} | \boldsymbol{Y}=\boldsymbol{y}, \tilde{\boldsymbol{X}}=\boldsymbol{x}, \mu, I, J, Q} \cdot M^T_{\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}=\boldsymbol{x}, \mu, I, J, Q} \end{aligned} \tag{2.12}$$

where on the LHS, the element on the i-th row and j-th column corresponds to the probability:

$$Pr(\boldsymbol{X}^d = \boldsymbol{x}, (\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d) = (\boldsymbol{p}, \boldsymbol{z})_j, \boldsymbol{Z}^d = \boldsymbol{z}_i | \boldsymbol{Y} = \mathrm{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q)$$

where $(\boldsymbol{p}, \boldsymbol{z})_j$(resp. $\boldsymbol{z}_i$) is the j-th(resp. i-th) distinct value for the vector $(\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d)$(resp. $\boldsymbol{Z}^d$). The first and third matrix on the RHS is similarly defined. The middle matrix D on the RHS is diagonal whose elements are the probability of observing $(\boldsymbol{Z}^* = \boldsymbol{z}, \boldsymbol{X}^* = \boldsymbol{x}', \boldsymbol{X}^d = \boldsymbol{x})$ for various values of $(\boldsymbol{z}, \boldsymbol{x}')$ conditional on $(\boldsymbol{Y} = \boldsymbol{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q)$. All matrices are of dimension $(m^J \cdot l^I) \times (m^J \cdot l^I)$. In addition,

$$M_{\boldsymbol{P}^d,\tilde{\boldsymbol{Z}}^d,\boldsymbol{Z}^d|\boldsymbol{Y}=\boldsymbol{1}-\boldsymbol{y},\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu_t,I,J,Q} = M_{\boldsymbol{Z}^d|\boldsymbol{Z}^*,\boldsymbol{X}^*,\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \cdot D_{\boldsymbol{Z}^*,\boldsymbol{X}^*|\boldsymbol{Y}=1-\boldsymbol{y},\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \cdot M^T_{\boldsymbol{P}^d,\tilde{\boldsymbol{Z}}^d,\boldsymbol{Z}^*,\boldsymbol{X}^*|\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \tag{2.13}$$

By inverting equation (2.13) and right-multiplying equation (2.12), we get

$$\begin{aligned} & M_{\boldsymbol{X}^d=\boldsymbol{x},\boldsymbol{P}^d,\tilde{\boldsymbol{Z}}^d,\boldsymbol{Z}^d|\boldsymbol{Y}=\boldsymbol{y},\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \cdot M^{-1}_{\boldsymbol{P}^d,\tilde{\boldsymbol{Z}}^d,\boldsymbol{Z}^d|\boldsymbol{Y}=\boldsymbol{1}-\boldsymbol{y},\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu_t,I,J,Q} \\ & = M_{\boldsymbol{Z}^d|\boldsymbol{Z}^*,\boldsymbol{X}^*,\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \cdot D_{\boldsymbol{Z}^*,\boldsymbol{X}^*,\boldsymbol{X}^d=\boldsymbol{x}|\boldsymbol{Y}=\boldsymbol{y},\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \cdot D^{-1}_{\boldsymbol{Z}^*,\boldsymbol{X}^*|\boldsymbol{Y}=1-\boldsymbol{y},\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \cdot \\ & M^{-1}_{\boldsymbol{Z}^d|\boldsymbol{Z}^*,\boldsymbol{X}^*,\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \end{aligned} \tag{2.14}$$

Here, the matrices on the LHS are directly computable from data, and the RHS embeds the distribution of unobservables that we are interested in. I then exploit the independence condition of observable characteristics across jobs:

$$M_{\boldsymbol{Z}^d|\boldsymbol{Z}^*,\boldsymbol{X}^*,\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} = M_{Z^d_1|Z^*_1,\boldsymbol{X}^*_{\mu(1)}} \otimes M_{Z^d_2|Z^*_2,\boldsymbol{X}^*_{\mu(2)}} \otimes \cdots \otimes M_{Z^d_J|Z^*_J,\boldsymbol{X}^*_{\mu(J)}} \tag{2.15}$$

and,

$$
\begin{aligned}
&D_{\boldsymbol{Z}^*,\boldsymbol{X}^*,\boldsymbol{X}^d=\boldsymbol{x}|\boldsymbol{Y}=\boldsymbol{y},\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \cdot D^{-1}_{\boldsymbol{Z}^*,\boldsymbol{X}^*|\boldsymbol{Y}=1-\boldsymbol{y},\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} = \\
&[(D_{Z_1^*|Y_1=y_1} \cdot D^{-1}_{Z_1^*|Y_1=1-y_1}) \otimes D_{\boldsymbol{X}^d_{\mu(1)}|X^*_{\mu(1)},\tilde{\boldsymbol{X}}_{\mu(1)}}] \otimes \cdots \otimes \\
&[(D_{Z_J^*|Y_J=y_J} \cdot D^{-1}_{Z_J^*|Y_J=1-y_J}) \otimes D_{\boldsymbol{X}^d_{\mu(J)}|X^*_{\mu(J)},\tilde{\boldsymbol{X}}_{\mu(J)}}] \qquad (2.16)
\end{aligned}
$$

The RHS of equation (3.29) therefore can be written as

$$
\begin{aligned}
&M_{\boldsymbol{Z}^d|\boldsymbol{Z}^*,\boldsymbol{X}^*,\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \cdot D_{\boldsymbol{Z}^*,\boldsymbol{X}^*,\boldsymbol{X}^d=\boldsymbol{x}|\boldsymbol{Y}=\boldsymbol{y},\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \cdot D^{-1}_{\boldsymbol{Z}^*,\boldsymbol{X}^*|\boldsymbol{Y}=1-\boldsymbol{y},\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \cdot \\
&M^{-1}_{\boldsymbol{Z}^d|\boldsymbol{Z}^*,\boldsymbol{X}^*,\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q} \\
&= (M_{Z_1^d|Z_1^*,\boldsymbol{X}^*_{\mu(1)}} \cdot [(D_{Z_1^*|Y_1=y_1} \cdot D^{-1}_{Z_1^*|Y_1=1-y_1}) \otimes D_{\boldsymbol{X}^d_{\mu(1)}|X^*_{\mu(1)},\tilde{\boldsymbol{X}}_{\mu(1)}}] \cdot M^{-1}_{Z_1^d|Z_1^*,\boldsymbol{X}^*_{\mu(1)}}) \otimes \cdots \\
&\otimes (M_{Z_J^d|Z_J^*,\boldsymbol{X}^*_{\mu(J)}} \cdot [(D_{Z_J^*|Y_J=y_J} \cdot D^{-1}_{Z_J^*|Y_J=1-y_J}) \otimes D_{\boldsymbol{X}^d_{\mu(J)}|X^*_{\mu(J)},\tilde{\boldsymbol{X}}_{\mu(J)}}] \cdot M^{-1}_{Z_J^d|Z_J^*,\boldsymbol{X}^*_{\mu(J)}}) \\
&:= M_1 \otimes M_2 \otimes \cdots \otimes M_J \qquad (2.17)
\end{aligned}
$$

On the RHS, each matrix M_j is a square matrix of dimension $m \cdot l^{|\mu(j)|}$. From the definition of the Kronecker product, we know that matrix M_J could be identified up to a positive scale from the upper-right $(m \cdot l^{|\mu(J)|}) \times (m \cdot l^{|\mu(J)|})$ submatrix of the LHS. Thus, by changing the order of variables when we construct the LHS, we could identify the M_j matrix for any job j up to a positive scale. The next step is to recover the matrix $M_{\boldsymbol{Z}_j^d|(Z_j^*,\boldsymbol{X}^*_{\mu(j)})}$ from the identified matrices M_j.

Suppose we have identified M_j up to a positive scale, $s > 0$. Then, by definition of M_j in equation (2.17), we have

$$s\cdot M_j = M_{Z_j^d|Z_j^*,\boldsymbol{X}^*_{\mu(j)}} \cdot [s\cdot(D_{Z_j^*|Y_j=y_j}\cdot D^{-1}_{Z_j^*|Y_j=1-y_j})\otimes D_{\boldsymbol{X}^d_{\mu(j)}|X^*_{\mu(j)},\tilde{\boldsymbol{X}}_{\mu(j)}}]\cdot M^{-1}_{Z_j^d|Z_j^*,\boldsymbol{X}^*_{\mu(j)}} \tag{2.18}$$

This equation reminds us to use the eigenvalue–eigenvector decomposition method developed in Hu (2008) for nonparametric identification. Specifically, each column in matrix $M_{\boldsymbol{Z}_j^d|(Z_j^*,\boldsymbol{X}^*_{\mu(j)})}$ corresponds to an eigenvector of the LHS matrix (after a normalization). Each corresponding diagonal element of $[s\cdot(D_{Z_j^*|Y_j=y_j}\cdot D^{-1}_{Z_j^*|Y_j=1-y_j})\otimes D_{\boldsymbol{X}^d_{\mu(j)}|X^*_{\mu(j)},\tilde{\boldsymbol{X}}_{\mu(j)}}]$ represents the corresponding eigenvalue of the LHS matrix. Following the argument in Hu (2008), in order to determine the ordering of all the eigenvectors, we exploit the implications from Assumption 4 and 13 to derive the following result:

Lemma 6 *Under Assumption 4 and 13, for each job j, it holds that all the eigenvalues of the matrix $s\cdot M_j$ are distinct.*

This result is derived from the fact that, from Assumption 4, given different complexity levels Z_j^*, we could always rank the probability of observing such complexity level given $Y_j=1$ relative to given $Y_j=0$. Furthermore, from part (1) of Assumption 13, the probability of correctly signaling one's ability by discretized ability indexes is higher than the probability of mismeasurement. Combining these two conditions together, I can pin down the ordering of diagonal elements of the eigenvalue matrix, which further leads to the identification of eigenvectors and eigenvalues in the matrix $s \cdot M_j$ up to the normalization of each column in $M_{Z_{j,1},\boldsymbol{Z}_{j,-1}|(Z_j^*,\boldsymbol{X}^*_{\mu(j)})}$. The next proposition is a major result to identify the joint distribution of $(\boldsymbol{Z}^*, \boldsymbol{X}^*)$ given the market size and observed match, μ.

Proposition 2 *Under Assumptions 3–13, the joint distribution of* $(\boldsymbol{P}, \boldsymbol{Z}^*, \boldsymbol{X}^* | \mu, I, J, Q)$ *is nonparametrically identified from observing the joint distribution of* $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{P}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z}, \tilde{\boldsymbol{X}})$ *conditional on a certain* (μ, I, J, Q).

The proof is in the appendix.

To fully construct the likelihood when estimating the model, the conditional distributions of $(\boldsymbol{Z}_j | Z_j^*, \boldsymbol{X}^*_{\mu(j)})$ and $(\boldsymbol{X}_{\mu(j)} | \boldsymbol{X}^*_{\mu(j)}, \tilde{\boldsymbol{X}}_{\mu(j)})$ must also be known. The following corollary establishes the identification result.

Corollary 1 *For any worker i and job j, the conditional distributions of* $(\boldsymbol{Z}_j | Z_j^*, \boldsymbol{X}^*_{\mu(j)})$ *and* $(\boldsymbol{X}_{\mu(j)} | \boldsymbol{X}^*_{\mu(j)}, \tilde{\boldsymbol{X}}_{\mu(j)})$ *are nonparametrically identified. Moreover, the underlying law of motion* $Pr(X_{it}^* | \widetilde{X}_{it}^*)$ *and initial condition* $Pr(X_{i1} | X_{i1}^*)$ *are nonparametrically identified.*

The proof is in the appendix.

2.3.3 Step 2: Identification of Workers' Utility Primitives

The market-level choice probability is equivalent to the probability of observing the matching allocation μ_t, and is determined solely by workers' preferences once we know their underlying skill levels. As their preferences are affected by $(\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{P})$, the market-level choice probability can be written as $Pr(\mu|\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{P})$.

In order to simplify the notation, the (I, J, Q) is suppressed from now on. One should bear in mind, however, that all of our distributions are conditioning on a certain market size. After identifying the conditional distribution of unobserved types of worker–slot pairs, we could apply the Bayes Theorem:

$$Pr(\mu|\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{P}) = \frac{f(\mathbf{P}, \mathbf{Z}^*, \mathbf{X}^*|\mu) \cdot Pr(\mu)}{\sum_\mu f(\mathbf{P}, \mathbf{Z}^*, \mathbf{X}^*|\mu) \cdot Pr(\mu)},$$

where, $f(\mathbf{P}, \mathbf{Z}^*, \mathbf{X}^*|\mu)$ is identified from the previous section and $Pr(\mu)$ is directly observable from data.

Given the knowledge of $Pr(\mu|\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{P})$, I am able to further decompose it into individual-level choice probability:

$$Pr(\mu|\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{P}) = Pr(u_{1,\mu(1)} \geq \max_{j \cup \emptyset} u_{1,j}) \cdot Pr(u_{2,\mu(2)} \geq \max_{j \neq \mu(1), \emptyset} u_{2,j}) \cdots Pr(u_{I,\mu(I)} \geq \max_{j \neq \mu(i), \forall i < I, \emptyset} u_{I,j})$$

if I order workers such that $X_1^* > X_2^* > \cdots > X_I^*$.

Recall that I parametrize their indirect utilities as:

$$u_{i,j^k} = \beta P_{j^k} + \gamma_0 X_i^* + (\gamma_1 X_i^* + \eta_{it}) \cdot Z_j^* + \nu_{ij}, \tag{2.19}$$

where ν_{ij} is the match-specific idiosyncratic taste shock, following an known i.i.d.

distribution. The coefficient γ_0 is the fixed-effect for a given ability level X_i^*; γ_1 measures each worker's preference over the interaction of her ability and the job's difficulty level. η_i is an unobserved taste determinant that follows a distribution known up to a K-dimensional parameter $\boldsymbol{\sigma}$. The unknown parameters are, therefore, $(\beta, \gamma_0, \gamma_1 \boldsymbol{\sigma})$.

I first write down the conditional probability of the best worker, worker 1, chooses $\mu_t(1)$:

$$
\begin{aligned}
& Pr(\mu(1)|\boldsymbol{P}, \boldsymbol{X}^*, \boldsymbol{Z}^*) \\
& = Pr(\beta P_{\mu(1)} + \gamma_0 X_1^* + (\gamma_1 X_1^* + \eta_1) \cdot Z_{\mu(1)}^* \nu_{1,\mu(1)} \geq \\
& \max\left(\beta P_{j^k} + \gamma_0 X_1^* + (\gamma_1 X_1^* + \eta_1) \cdot Z_j^* + \nu_{1,j}, \forall j^k \in M_i[\mu_t]\right) \\
& := H(\beta P_{j^k} + \gamma_0 X_1^* + (\gamma_1 X_1^* + \eta_1) \cdot Z_j^*, \forall j^k),
\end{aligned}
$$

and the functional form of H is known. Other workers' choice probabilities could also be written similarly, except that their choice sets are constrained by their ability ranking in the market.

First, to identify β, I exploit the situation when all jobs in the market have zero complexity and all workers have zero ability level. [10] Then, the choice probability for

[10] This could be done by normalization of (X^*, Z^*).

the best worker becomes

$$Pr(\mu(1)|\mathbf{Z}^*, \mathbf{X}^*, \mathbf{P})|_{Z_j^*=0, X_i^*=0 \forall i,j} = H(\beta P_{j^k}).$$

The LHS of the equation is identified from the following equation. The RHS has only one unknown parameter β. If the H-function is invertible on β, then it is identified.[11]

$$Pr(\mu(1)|\mathbf{Z}^*, \mathbf{X}^*, \mathbf{P})|_{Z_j^*=0, X_i^*=0 \forall i,j} = \sum_{\mu(2),\cdots,\mu(I)} Pr(\mu|\mathbf{Z}^*, \mathbf{X}^*, \mathbf{P})|_{Z_j^*=0, X_i^*=0 \forall i,j}$$

Next, in order to identify γ_0, I exploit the situation when all jobs in the market have zero complexity level. Firstly we have:

$$\frac{\partial Pr(\mu(1)|\mathbf{Z}^*, \mathbf{X}^*, \mathbf{P})}{\partial P_j}|_{Z_j^*=0, \forall j} = \beta H^{(1)}(\beta P_j + \gamma_0 X_1^*)$$

where $H^{(1)}$ denotes the derivative of H-function over $\beta P_j + \gamma_0 X_1^* + (\gamma_1 X_1^* + \eta_1) Z_j^*$. Similarly, we have:

$$\frac{\partial Pr(\mu(1)|\mathbf{Z}^*, \mathbf{X}^*, \mathbf{P})}{\partial X_1^*}|_{Z_j^*=0, \forall j} = \gamma_0 \sum_j H^{(1)}(\beta P_j + \gamma_0 X_1^*)$$

[11] Typical specifications, such as the multinomial logit model, ensure the invertibility of H.

Therefore, γ_0 could be identified from:

$$\gamma_0 = \beta\{\frac{\partial Pr(\mu(1)|\mathbf{Z}^*,\mathbf{X}^*,\mathbf{P})}{\partial X_1^*} / \sum_j \frac{\partial Pr(\mu(1)|\mathbf{Z}^*,\mathbf{X}^*,\mathbf{P})}{\partial P_j}\}|_{Z_j^*=0,\forall j}$$

Next, identifying γ_1 is done through the following equation:

$$\gamma_1 = \frac{\frac{\partial Pr(\mu(1)|\mathbf{Z}^*,\mathbf{X}^*,\mathbf{P})}{\partial X_1^*} - [\frac{\partial Pr(\mu(1)|\mathbf{Z}^*,\mathbf{X}^*,\mathbf{P})}{\partial X_1^*}|_{Z_j^*=0,\forall j}]}{\sum_j \frac{\partial Pr(\mu(1)|\mathbf{Z}^*,\mathbf{X}^*,\mathbf{P})}{\partial P_j} Z_j^*/\beta}$$

Lastly, to identify the distributional parameters of η_1, we need to make the following assumption to identify $\boldsymbol{\sigma}$.

Assumption 7 *The function $H(\cdot)$ is K times continuously differentiable and there always exists some cash payment such that $H^{(k)}(\beta(P_{\mu(1)} - P_j)) \neq 0, \forall k = 0, 1, \cdots, K$.*

By Assumption 7, we can construct the moment functions of η from taking the derivative of $Pr(\mu(1)|\mathbf{Z}^*,\mathbf{X}^*,\mathbf{P})|_{X_1^*=0}$ when all jobs in the market have the same level of complexity:

$$\frac{\partial Pr(\mu(1)|\mathbf{Z}^*,\mathbf{X}^*,\mathbf{P})}{\partial Z_{\mu(1)}^*}|_{Z_j^*=0,\forall j} = \gamma_1 X_1^* H^{(1)}(\beta P_j + \gamma_0 X_1^*) + H^{(1)}(\beta P_j + \gamma_0 X_1^*) \int_{\eta_1} \eta_1 dF(\eta_1)$$

which indicates:

$$\int_{\eta_1} \eta_1 dF(\eta_1) = \frac{\frac{\partial Pr(\mu(1)|\mathbf{Z}^*,\mathbf{X}^*,\mathbf{P})}{\partial Z_{\mu(1)}^*}|_{Z_j^*=0,\forall j} - \gamma_1 X_1^* H^{(1)}(\beta P_j + \gamma_0 X_1^*)}{H^{(1)}(\beta P_j + \gamma_0 X_1^*)}$$

and similarly,

$$\int_{\eta_1} (\gamma_1 X_1^* + \eta_1)^k dF(\eta_1) = \frac{\frac{\partial^k Pr(\mu(1)|\mathbf{Z}^*,\mathbf{X}^*,\mathbf{P})}{\partial Z^{*}_{\mu(1)}{}^{k}}|_{Z_j^*=0,\forall j}}{H^{(k)}(\beta P_j + \gamma_0 X_1^*)}$$

where the RHS can be perfectly computed now, and the LHS corresponds to the k-th moment of η_1. This suffices for the identification of $\boldsymbol{\sigma}$. This method of identification is a special case of Fox et al. (2012).

Consequently, all the utility primitives are identified in the model.

2.3.4 Step 3: Identification of Firms' Profit Primitives

From Step 1, we have identified the joint distribution of $(\boldsymbol{P}, \boldsymbol{Z}^*, \boldsymbol{X}^*)$. We know that, for each job j, the symmetric equilibrium cash payment depends on (Z_j^*, ω_t), which implies

$$P_j \perp P_{j'} \perp P_{j''} | Z_j^*, Z_{j'}^*, Z_{j''}^*, \omega \tag{2.20}$$

for any j, j', j'' in market t. Therefore, we could recover the distribution of market unobserved heterogeneity, ω_t, from jointly observing at least three jobs in the market. Specifically, suppose we observe $J \geq 3$ jobs per market, then conditional on $J \geq 3$, we could construct the joint probability

$$f(P_1, P_2, P_3|Z_1^*, Z_2^*, Z_3^*) = \sum_{\omega} f(P_1|Z_1^*, \omega) \cdot f(P_2|Z_2^*, \omega) \cdot f(P_3, \omega|Z_3^*) \tag{2.21}$$

and

$$f(P_1, P_3|Z_1^*, Z_2^*, Z_3^*) = \sum_{\omega} f(P_1|Z_1^*, \omega) \cdot f(P_3, \omega|Z_3^*). \tag{2.22}$$

This naturally relates to a measurement-error model; following Hu (2008), we make the following assumption to identify the conditional distribution, $f(P_j|Z_j^*, \omega)$.

Assumption 8 *1. ω is drawn from a finite support $\{w_1, w_2, \cdots, w_n\}$ with $n \leq m$.*

2. There exists a mapping $\rho : \{p_1, p_2, \cdots, p_m\} \to \{1, 2, \cdots, n\}$ such that the following matrix is of full rank $n \times n$.

$$M_{\rho(P_1),\rho(P_3)|Z_1^*,Z_3^*} := [Pr(\rho(P_1) = p, \rho(P_3) = p'|Z_1^*, Z_3^*)]_{p,p' \in \{1,2,\cdots,n\}}$$

The next theorem tells us that we could identify the distribution of a single firm's cash payment, P, conditional on its job complexity, Z^*, and the market unobservable ω. This condition distribution could be viewed as firms' pricing strategies.

Proposition 3 *Given Assumption 11, we can nonparametrically identify $Pr(P_j|Z_j^*, \omega)$ and the marginal distribution of ω.*

As discussed in the previous section, the equilibrium cash prize distribution maximizes the interim payoff of each firm at Stage 1. Specifically, let $\delta^* := Pr(P|Z^*, \omega)$, then

$$\delta^* \in argmax_{\delta \in \Sigma} \sum_{l=1}^{m} [\sum_{Z^*_{-j}, P_{-j}} \pi^{int}(p_l, Z^*_j, Z^*_{-j}, \delta^*_{-j}(\cdot), \omega)] \cdot \delta^l_j$$

where $\pi^{\text{int}}(p_l, Z^*_j, Z^*_{-j}, \delta^*_{-j}(\cdot), \omega)$ is the firm's interim payoff function defined previously. In equilibrium, δ^* is a function of $(c, Z^*, \omega, \beta, \gamma, \sigma, Pr(Z^*), Pr(X^*))$, where only the profit coefficient c is not known. As we already made the assumption with regard to the equilibrium selection rule, we only need to make sure the δ^* function is invertible for c. From Lemma 4, the equilibrium distribution of cash payment δ^* is stochastically increasing with respect to c. Thus, the profit coefficient, c, can be identified from

$$c = (\delta^*)^{-1}(Pr(P|Z^*, \omega), Z^*, \omega, \beta, \gamma, \sigma, Pr(Z^*), Pr(X^*)).$$

So, we have nonparametrically identified the distribution of unobserved heterogeneity on both sides of the market, and more importantly, identified the preference primitives for firms and workers. The next section will discuss the estimation procedure in detail.

2.3.5 Likelihood-Based Estimation

Although the distributions of unobserved types $(\boldsymbol{Z}^*, \boldsymbol{X}^*)$ are entirely nonparametrically identified, empirically it is hard to estimate the distribution fully nonparametrically due to data constraints. Here, we assume a parametric structure on the function of match-specific outcome, Z, and preserve the nonparametric structure for other distributions related to $(\boldsymbol{Z}^*, \boldsymbol{X}^*)$. First, we estimate a simple case, where workers' types, $\boldsymbol{X}^*$, could be entirely inferred from their ability indexes, and we estimate the distribution of job-level unobserved heterogeneity, $\boldsymbol{Z}^*$, along with workers' utility primitives. Then we extend to the general case of latent workers' types and estimate its distribution along with other primitives in the basic model. The dimensionality problem arises in the latter model, as workers participate in different markets. To make the estimation practical, I make further assumptions on the workers' participation rule to reduce the dimensionality of the likelihood. Lastly, I discuss a simulation-based approach to estimate firms' utility primitives.

2.3.5.1 Benchmark Case: When $\boldsymbol{X}^*$ Are Perfectly Observed

In this benchmark model, I use workers' ability indexes (X_i) as the perfect measure of their skill levels, X_i^*. Also, the match-specific outcome from worker i and job j is determined by

$$\begin{aligned} Z_{ij}^t &= \xi_1 X_{it}^* + \xi_2 Z_{jt}^* + b_{jt} \\ &= \xi_1 X_{it} + \xi_2 Z_{jt}^* + b_{jt}, \end{aligned} \tag{2.23}$$

where $\xi_1 > 0$ and $\xi_2 < 0$ are unknown. b follows an atomless distribution known up to a finite-dimensional parameter $\boldsymbol{\kappa}$.

Also for simplicity, I assume cash payment follows the exogenous distribution, $\delta := Pr(P|Z^*, \omega)$. This is because of the complexity of simulating the equilibrium pricing strategies. I will incorporate the equilibrium wage-setting stage in a more general model in section 3.4.3.

The primitives I would like to estimate consist of three sets: (1) $(\beta, \gamma, \boldsymbol{\sigma})$, which are parameters in workers' indirect utility function; (2) $(\xi, \boldsymbol{\kappa})$, which are parameters in the match-specific outcome function in equation (3.11); and (3) the distributions $Pr(Y|Z^*, X^*, \mu_t)$, $Pr(P|Z^*, \omega)$, $Pr(Z^*)$ and $Pr(\omega)$. Including all parameters and distributions into $\boldsymbol{\theta}$, the log-likelihood function is defined as

$$\begin{aligned}
\boldsymbol{LL}(\boldsymbol{\theta}) :=& \sum_{t=1}^{T} \log(Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*)) \\
=& \sum_{t=1}^{T} \log \Bigg(\sum_{\boldsymbol{Z}_t^*} f(\boldsymbol{Z}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \mu_t) \cdot Pr(\boldsymbol{Y}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \mu_t) \cdot Pr(\mu_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \boldsymbol{P}_t) \\
& \qquad \cdot \sum_{\omega_t} Pr(\boldsymbol{P}_t | \boldsymbol{Z}_t^*, \omega_t) \cdot Pr(\omega_t) \cdot Pr(\boldsymbol{Z}_t^*) \Bigg)
\end{aligned} \tag{2.24}$$

where, if we rank workers in market t such that $X_1^* \geq X_2^* \geq \cdots \geq X_I^*$,[12]

$$Pr(\mu_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \boldsymbol{P}_t) = H_1 \cdot H_2 \cdot \cdots \cdot H_I$$

[12] For simplicity, assume when two workers carry the same ability level, all firms strictly prefer the one with a smaller subscript.

and,

$$\begin{aligned} H_i :=& Pr(\mu(i) = j^k | \mu(i'), \forall i' < i; \boldsymbol{X}_t^*, \boldsymbol{Z}_t^*, \boldsymbol{P}_t^*) \\ =& \int_{\eta_i} \frac{1}{\sum_{m \neq \mu(i')} \exp(\beta(P_m - P_{j^k}) + (\gamma X_i^* + \eta_i)(Z_m - Z_j))} dF_\eta(\eta_i) \end{aligned}$$

In words, H_i is the probability of worker i choosing job $\mu(i)$ given the choices by other better workers in the market. Lastly, the probability of observing the match-specific outcomes are:

$$f(\boldsymbol{Z}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \mu_t) = \prod_{j=1}^{J_t} f_{\boldsymbol{Z}_j}(z | Z_j^*, \boldsymbol{X}_{\mu_t(j)}^*) = \prod_{j=1}^{J_t} f_{b_j}(z - X_{\mu(j),1}^* - \xi Z_j^*), \tag{2.25}$$

where $(\mu(j), 1)$ denotes the worker matched with job j and sits in the first place. All other probabilities on the RHS of equation (3.19) are primitives of the model and are discrete.

So far, all the components in the likelihood are fully specified, and the Monte Carlo simulation result is presented in section 2.4.1.

The next lemma establishes the consistency result for the estimator.

Lemma 7 *Assume (i) the product space for estimation primitives are compact;* [13] *(ii) $Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*)$ is continuous in all parameters and probability distributions; (iii) the set of primitives such that $Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*) > 0$ does not depend on the*

[13] The space include the parametric spaces for (1) $(\beta, \gamma, \boldsymbol{\sigma})$, which are parameters in workers' utility function; (2) $(\xi, \boldsymbol{\kappa})$, which are parameters in the score function in equation (3.11), and lastly, the probability space for the distributions $Pr(Y|Z^*)$, $Pr(P|Z^*, \omega)$, $Pr(Z^*)$ and $Pr(\omega)$.

value of primitives; (iv) there exists a function $K(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*)$ *such that* $\log Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*, \boldsymbol{\theta}) - \log Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*, \boldsymbol{\theta_0}) \leq K(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*)$ *and* $EK(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*) < \infty$*; then the likelihood estimator that maximizes the function in equation (3.19) converges in probability to the true values of the primitives.*

Proof can be found in Appendix 2.7.9.

2.3.5.2 General Case: Both $(\boldsymbol{X}^*, \boldsymbol{Z}^*)$ Are Latent

In the general case, where both $(\boldsymbol{X}^*, \boldsymbol{Z}^*)$ are latent, we make one simplification on the generating process for ability indexes. Specifically, we simplify equation (2.8) to be:

$$X_{it} = g(\tilde{X}_{it}, X_{it}^*, a_{it}). \tag{2.26}$$

Thus, the distribution of X_{it} depends only on $(\tilde{X}_{it}, X_{it}^*)$. Compared with the benchmark case, where X^* is completely observable, $Pr(X_{it}|\widetilde{X}_{it}, X_{it}^*)$ and $Pr(X_{it}^*|\widetilde{X}_{it}^*)$ are the additional primitives I would like to estimate. The likelihood function therefore is

$$\begin{aligned}
\boldsymbol{L}(\boldsymbol{\theta}) =& Pr(\boldsymbol{W}_1, \boldsymbol{W}_2, \cdots, \boldsymbol{W}_T) \\
=& \sum_{\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \cdots, \boldsymbol{X}_T^*} Pr(\boldsymbol{W}_1 | \boldsymbol{X}_1^*) \cdot Pr(\boldsymbol{W}_2 | \boldsymbol{X}_2^*, \boldsymbol{W}_{<2}) \cdots Pr(\boldsymbol{W}_T | \boldsymbol{X}_T^*, \boldsymbol{W}_{<T}) \cdot Pr(\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \cdots, \boldsymbol{X}_T^*)
\end{aligned} \tag{2.27}$$

where $\boldsymbol{W}_t$ includes all variables observed in market t, i.e., $(\boldsymbol{Y}_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{X}_t, \mu_t)$. From

Assumption 2, we know that

$$\begin{aligned} Pr(\boldsymbol{W}_t|\boldsymbol{X}_t^*, \boldsymbol{W}_{<t}) &= Pr(\boldsymbol{W}_t|\boldsymbol{X}_t^*, \widetilde{\boldsymbol{X}}_t), \\ &= \sum_{\boldsymbol{Z}_t^*} Pr(\boldsymbol{Z}_t|\boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \mu_t) \cdot Pr(\boldsymbol{Y}_t|\boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \mu_t) \cdot Pr(\boldsymbol{X}_t|\boldsymbol{X}_t^*, \widetilde{\boldsymbol{X}}_t) \\ &\quad \cdot Pr(\mu_t|\boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \boldsymbol{P}_t)) \cdot Pr(\boldsymbol{P}_t, \boldsymbol{Z}_t^*). \end{aligned} \tag{2.28}$$

where $\widetilde{\boldsymbol{X}}_t$ refers to the vector of the most recent rating scores of each worker in market t. And in the first market,

$$Pr(\boldsymbol{W}_1|\boldsymbol{X}_1^*) = \sum_{\boldsymbol{Z}_1^*} Pr(\boldsymbol{Z}_1|\boldsymbol{Z}_1^*, \boldsymbol{X}_1^*, \mu_1) \cdot Pr(\boldsymbol{X}_1|\boldsymbol{X}_1^*) \cdot Pr(\mu_1|\boldsymbol{Z}_1^*, \boldsymbol{X}_1^*, \boldsymbol{P}_1) \cdot Pr(\boldsymbol{P}_1, \boldsymbol{Z}_1^*). \tag{2.29}$$

I have so far constructed the likelihood function for the general case. Ideally, one would estimate the conditional distribution of $Pr(X|\widetilde{X}, X^*)$, the initial condition $Pr(X_1|X_1^*)$ and the law of motion $Pr(X_t^*|\widetilde{X}^*)$ along with other unknown primitives using a Likelihood-Based estimator. Practically, however, due to the high dimensionality of $(\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \cdots, \boldsymbol{X}_T^*)$, this is impossible to do without further modification.

To see why the dimensionality problem arises, suppose in the real data, we have N workers and T markets in total. Given that workers are very likely to appear in multiple markets, they can be viewed as “global players” of the repeated game. On average, if we observe $M > N/T$ workers per market on average, to compute the full likelihood, we need to evaluate the joint probability distribution $Pr(\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \cdots, \boldsymbol{X}_T^*)$ approximately $2^{M \cdot T}$ times, even if each worker's skill level takes only two possible values. This is computationally unrealistic to implement as T grows large. Instead, I make the following assumption to simplify the estimation procedure.

Assumption 9 *Suppose the probability of worker i being present in market t is*

$$e_{it} = \begin{cases} e_0, & \textit{if worker } i \textit{ never participated before,} \\ e_1, & \textit{if worker } i \textit{ participated in market } t-1, \\ 0, & \textit{if worker } i \textit{ appeared both in market } t \textit{ and } t-1. \end{cases}$$

where $0 < e_2 < e_1 < 1$.

Intuitively, workers enter the market and stay for at most two consecutive weeks. Afterwards, they leave the market forever. The entry, stay and exit decisions are nonstrategic though. Furthermore, I assume the populations of both workers and firms consist of countless many candidates. Thus, there always exist positive numbers of workers and firms across all markets. Hence, in each market t, the set of workers, I_t, could be divided into three categories, $\{I_{t,-1}, I_{t,0}, I_{t,+1}\}$, where $I_{t,-1}$ denotes the ones that also appeared in market $t-1$ and stayed in market t, $I_{t,0}$ denotes workers that appeared only in market t and leave forever after t, and $I_{t,+1}$ denotes those who first appear in market t and stay for one more period. The detailed illustration is in in Figure (3.6). Now, I construct a new likelihood function that uses partial information from what we observe:[14]

$$
\begin{aligned}
\widetilde{\boldsymbol{L}}(\theta) =& Pr(\boldsymbol{W}_1, \boldsymbol{W}_3, \cdots, \boldsymbol{W}_{T-1} | [\boldsymbol{Z}_2, \boldsymbol{X}_2, \mu_2], [\boldsymbol{Z}_4, \boldsymbol{X}_4, \mu_4], \cdots, [\boldsymbol{Z}_T, \boldsymbol{X}_T, \mu_T]) \\
=& \frac{Pr(\boldsymbol{W}_1, \boldsymbol{W}_3, \cdots, \boldsymbol{W}_{T-1}, [\boldsymbol{Z}_2, \boldsymbol{X}_2, \mu_2], [\boldsymbol{Z}_4, \boldsymbol{X}_4, \mu_4], \cdots, [\boldsymbol{Z}_T, \boldsymbol{X}_T, \mu_T])}{Pr([\boldsymbol{Z}_2, \boldsymbol{X}_2, \mu_2], [\boldsymbol{Z}_4, \boldsymbol{X}_4, \mu_4], \cdots, [\boldsymbol{Z}_T, \boldsymbol{X}_T, \mu_T])} \\
=& \frac{\sum_{[\boldsymbol{P}_2, \boldsymbol{Y}_2], \ldots, [\boldsymbol{P}_T, \boldsymbol{Y}_T]} \boldsymbol{L}(\boldsymbol{\theta})}{\sum_{\boldsymbol{W}_1, \boldsymbol{W}_3, \cdots, \boldsymbol{W}_{T-1}} \sum_{[\boldsymbol{P}_2, \boldsymbol{Y}_2], \ldots, [\boldsymbol{P}_T, \boldsymbol{Y}_T]} \boldsymbol{L}(\boldsymbol{\theta})}.
\end{aligned} \tag{2.30}
$$

Essentially, I have integrated out some information with regard to even markets, and mainly focus on the odd markets. This enables me to compute market-level likelihood separately without encountering the dimensionality problem. To see this,

$$
\begin{aligned}
\widetilde{\boldsymbol{L}}(\theta) =& Pr(\boldsymbol{W}_1, \boldsymbol{W}_3, \cdots, \boldsymbol{W}_{T-1} | Pr([\boldsymbol{Z}_2, \boldsymbol{X}_2, \mu_2], [\boldsymbol{Z}_4, \boldsymbol{X}_4, \mu_4], \cdots, [\boldsymbol{Z}_T, \boldsymbol{X}_T, \mu_T]) \\
=& \prod_{t=1}^{T/2} Pr(\boldsymbol{W}_{2t-1} | [\boldsymbol{Z}_2, \boldsymbol{X}_2, \mu_2], [\boldsymbol{Z}_4, \boldsymbol{X}_4, \mu_4], \cdots, [\boldsymbol{Z}_T, \boldsymbol{X}_T, \mu_T]) \\
=& \prod_{t=1}^{T/2} Pr(\boldsymbol{W}_{2t-1} | [\boldsymbol{Z}_{2t-2}, \boldsymbol{X}_{2t-2}], [\boldsymbol{Z}_{2t}, \boldsymbol{X}_{2t}]) \\
=& \prod_{t=1}^{T/2} \sum_{\boldsymbol{X}^*_{2t-1}} Pr(\boldsymbol{W}_{2t-1}, \boldsymbol{X}^*_{2t-1} | [\boldsymbol{Z}_{2t-2}, \boldsymbol{X}_{2t-2}, \mu_{2t-2}], [\boldsymbol{Z}_{2t}, \boldsymbol{X}_{2t}, \mu_{2t}]).
\end{aligned} \tag{2.31}
$$

I condition on the information of Z and X from even markets, as they are relevant to workers' underlying types X^*. The second equality follows from the fact that in market $2t-1$ and $2t+1$, no workers is overlapping anymore; thus, the joint

[14] Without loss of generality, assume T is an even number. In addition, there is an abuse of notation – some variables in $\boldsymbol{W}$ are continuously distributed, but I use the summation sign to denote the summation of all possible values for discrete variables, and the integration over the support of continuous variables.

distributions of observables are independent of each other, conditional on neighbor market observables. The third equality follows from the fact that workers stay for at most two periods. Thus the only relevant information from even markets is about the match-specific outcomes from right before and right after market t.

To simplify my notation, I denote variables $(\boldsymbol{Z}_t, \boldsymbol{X}_t)$ as $\boldsymbol{R}_t$. Then I divide all variables into three parts according to whether the worker is from last period (denoted as subscript $\{t, +1\}$), stays only at this period (denoted as subscript $\{t, 0\}$), or stays for one more period (denoted as $\{t, -1\}$). Then, I have,

$$
\begin{aligned}
& Pr(\boldsymbol{W}_t, \boldsymbol{X}_t^* | \boldsymbol{Z}_{t-1}, \boldsymbol{X}_{t-1}, \mu_{t-1}, \boldsymbol{Z}_{t+1}, \boldsymbol{X}_{t+1}, \mu_{t+1}) \\
& = Pr(\boldsymbol{W}_t, \boldsymbol{X}_t^* | \boldsymbol{R}_{t-1}, \boldsymbol{R}_{t+1}) \\
& = Pr(\boldsymbol{W}_t, \boldsymbol{X}_t^* | \boldsymbol{R}_{t-1,-1}, \boldsymbol{R}_{t-1,0}, \boldsymbol{R}_{t-1,+1}, \boldsymbol{R}_{t+1,-1}, \boldsymbol{R}_{t+1,0}, \boldsymbol{R}_{t+1,+1}) \\
& = Pr(\boldsymbol{W}_t, \boldsymbol{X}_t^* | \boldsymbol{R}_{t-1,+1}, \boldsymbol{R}_{t+1,-1}).
\end{aligned} \tag{2.32}
$$

The likelihood function becomes

$$
\begin{aligned}
\widetilde{\boldsymbol{L}}(\theta) &= \prod_{t=1}^{T/2} \sum_{\boldsymbol{X}_{2t-1}^*} Pr(\boldsymbol{W}_{2t-1}, \boldsymbol{X}_{2t-1}^* | [\boldsymbol{Z}_{2t-2}, \boldsymbol{X}_{2t-2}, \mu_{2t-2}], [\boldsymbol{Z}_{2t}, \boldsymbol{X}_{2t}, \mu_{2t}]) \\
&= \prod_{t=1}^{T/2} \sum_{\boldsymbol{X}_{2t-1}^*} Pr(\boldsymbol{W}_{2t-1}, \boldsymbol{X}_{2t-1}^* | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}, \mu_{2t-2,+1}], [\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}, \mu_{2t,-1}]).
\end{aligned} \tag{2.33}
$$

In other words, I have avoided the dimensionality problem by picking up only odd numbers of weeks from the data, and partial information from even numbers of weeks. This makes the estimation much more manageable in practice. I write the log-likelihood function as

$$\widetilde{\boldsymbol{LL}}(\theta) = \sum_{t=1}^{T/2} \log \sum_{\boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1}} Pr(\boldsymbol{W}_{2t-1}, \boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}], [\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}])$$

$$= \sum_{t=1}^{T/2} \log \sum_{\boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1}} Pr(\boldsymbol{Z}_{2t-1} | \boldsymbol{Z}^*_{2t-1}, \boldsymbol{X}^*_{2t-1}, \mu_{2t-1}) \cdot Pr(\boldsymbol{Y}_{2t-1} | \boldsymbol{X}_{2t-2,+1}, \boldsymbol{Z}^*_{2t-1}, \boldsymbol{X}^*_{2t-1}, \mu_{2t-1}) \cdot$$

$$Pr(\mu_{2t-1} | \boldsymbol{Z}^*_{2t-1}, \boldsymbol{X}^*_{2t-1}, \boldsymbol{P}_{2t-1}) \cdot \sum_{\omega_{2t-1}} Pr(\boldsymbol{P}_{2t-1} | \boldsymbol{Z}^*_{2t-1}, \omega_{2t-1}) \cdot Pr(\omega_{2t-1}) \cdot Pr(\boldsymbol{Z}^*_{2t-1}) \cdot$$

$$Pr(\boldsymbol{X}_{2t-1}, \boldsymbol{X}^*_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}], [\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}]). \tag{2.34}$$

The RHS of equation (3.24) corresponds to the primitives we are interested in estimating. The detailed derivation is provided in the appendix.

Assumption 9 may seem to be restrictive at first sight, but empirically it is acceptable to focus on workers' consecutive participation behavior only, and treat reentry behavior separately for the same worker.

The next lemma establishes the consistency result for the estimator.

Lemma 8 *Assume (i) the product space for estimation primitives are compact;* [15] *(ii)* $Pr(\boldsymbol{W}_{2t-1}, \boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}], [\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}])$ *is continuous in all*

[15] The space include the parametric spaces for (1) $(\beta, \gamma, \boldsymbol{\sigma})$, which are parameters in workers' utility function; (2) $(\xi, \boldsymbol{\kappa})$, which are parameters in the score function in equation (3.11), and the probability space for the distributions $Pr(Y|Z^*)$, $Pr(P|Z^*, \omega)$, $Pr(Z^*)$ and $Pr(\omega)$; lastly, it includes the conditional distribution of $Pr(X|\widetilde{X}, X^*)$, the initial condition $Pr(X_1|X_1^*)$ and the law of motion $Pr(X_t^*|\widetilde{X}^*)$

parameters and probability distributions, and (iii) the set of primitives such that $Pr(\boldsymbol{W}_{2t-1}, \boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}], [\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}]) > 0$ *does not depend on the value of primitives; (iv) there exists a function* $K(\boldsymbol{W}_{2t-1})$ *such that* $\log Pr(\boldsymbol{W}_{2t-1} | [\boldsymbol{Z}_{2t-2}, \boldsymbol{X}_{2t-2}], [\boldsymbol{Z}_{2t}, \boldsymbol{X}_{2t}], \boldsymbol{\theta}) - \log Pr(\boldsymbol{W}_{2t-1} | [\boldsymbol{Z}_{2t-2}, \boldsymbol{X}_{2t-2}], [\boldsymbol{Z}_{2t}, \boldsymbol{X}_{2t}], \boldsymbol{\theta_0}) \leq K(\boldsymbol{W}_{2t-1})$ *and* $EK(\boldsymbol{W}_{2t-1}) < \infty$*; then the likelihood estimator that maximizes the function in equation (3.24) converges in probability to the true values of the primitives.*

The proof can be found in Appendix 2.7.10.

2.3.5.3 Including Wage-Setting Stage

As described earlier, the equilibrium decision of the optimal cash payment is set by firms before the matching process. Specifically, it should be related with the underlying complexity level of the firm's job. For simplicity, I assume workers' skill levels are still observable. In the most general case, we can always use the modified likelihood estimator from the previous section.

I conduct a two-step estimation: In the first step, I estimate $\widehat{\delta} := Pr(P_{jt} | Z^*_{jt}, \omega_t)$. In the second step, I match the simulated cash payment distribution with the estimated distribution from Step 1, and use the minimum-distance estimator to estimate the firm's profit parameter c.

2.4 Monte Carlo Evidence

This section presents the Monte Carlo results for the three estimation methods discussed above. I start from the benchmark case with only job-level unobserved heterogeneity, and the cash payments are set by an exogenous rule; next I incorporate worker-level unobserved heterogeneity, and lastly, I include the equilibrium wage-setting process and present the result for the model with only job-level unobserved heterogeneity.

2.4.1 Benchmark Case: When X^* Are Perfectly Observed

I generate a random sample consisting of 1,000 markets. Within each market, each job has a quota of recruiting up to 3 workers, and the number of workers within a market is randomly drawn from $\{8, 9, \cdots, 15\}$. For simplicity, assume each job carries a complexity level that takes two possible values, $\{1, 10\}$. The cash payment and match-specific outcomes are also drawn from the discrete space $\{1, 10\}$.[16] The idiosyncratic utility shock for workers, ν, follows a standard Type I extreme-value distribution. The heterogeneity in taste over complexity, η, follows a mean-zero normal distribution with unknown variance σ_η^2. On the firm side, the unobserved shock

[16] This assumption could be generalized to draw the cash on the positive integer space, and match-specific outcomes drawn from any bounded interval. This setting is simple, but it is sufficient for illustration.

b to match-specific outcomes follows a mean-zero normal distribution with unknown variance σ_b^2. The parameter ξ_1 is normalized to 1, as it could easily be estimated from $(Z_{ij}^t - Z_{lj}^t)/(X_{it}^* - X_{lt}^*)$. Lastly, for simplicity, I assume whether a coder's rating decreases or not only depends on the underlying complexity of the project he is enrolled in. This can be easily extended to the case that the binary indicator is related to both project complexity and the current rating score the coder has. The parameter specifications are given in Table 2.1. The first specification denotes the situation where workers' heterogeneous preferences are mainly driven by the random coefficient η, and they are moderately incentivized by wages. In contrast, the second specification denotes the situation where workers' heterogeneous preferences are mainly driven by the interaction between their skill levels and the job complexity, and they are highly incentivized by cash.

For each estimate, I use 125 Bootstrap iterations. The performance is shown in Table 2.2, with more detailed results provided in the appendix. The bias is quite small for all estimates. The estimates for the unobserved part has larger standard deviations than those for observable characteristics. For instance, the coefficient γ for the interaction term X^*Z^* is slightly more biased and has a larger standard deviation than β, the coefficient for observable cash payment.

2.4.2 General Case: Both $(\boldsymbol{X}^*, \boldsymbol{Z}^*)$ Are Latent

Now I allow workers' skill levels to be latent as well. Specifically, I assume that a worker's skill level on entering the market follows a Bernoulli distribution with parameter $p_X \in (0,1)$. In addition, we assume the observed ability indexes and evolution of underlying abilities are determined by two rules:

$$X_{\tau-1} = X^*_{\tau-1} + u_1 \tag{2.35}$$

and

$$X_\tau = \lambda \cdot X^*_\tau + (1-\lambda) \cdot X_{\tau-1} + u_2. \tag{2.36}$$

Lastly,

$$Pr(X^*_\tau = x_H | X^*_{\tau-1}) = [\delta_{x1}, \delta_{x2}]', \tag{2.37}$$

with (u_1, u_2) following zero-mean joint normal distribution with variance–covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$. The goal of estimation, therefore, is to find $(\sigma_1, \sigma_2, \lambda, \delta_{x1}, \delta_{x2}, p_X)$. Furthermore, I allow the cash payment to be drawn from the support consisting of three elements instead of two: $\{1, 4, 10\}$. The results are shown in Table 2.3. Compared with the benchmark case, we see larger bias and standard deviation, but the overall performance is still good.

2.4.3 Endogenous Cash Prize

In this section, I follow the benchmark case in which workers' skill levels are entirely observable, but add the second step of estimating firms' *ex post* profit primitive c using a minimum-distance estimator. Due to the computational complexity of simulating the equilibrium prize, I assume within each market that there are two firms and four workers who are randomly drawn from the population. Each firm offers two slots, and will award the first slot P, whereas the second slot receives $0.5P$. The cash payment P is drawn from a finite space $\{p_l, p_m, p_h\} = \{1, 4, 10\}$. Results are shown in Table 2.4. The estimate for the firms' profit parameter is less accurate than the other estimates, probably due to the finite-sample bias of the simulated moment.

2.5 Conclusion

This paper develops a two-stage model for two-sided markets where wage contracts are set before the matching process. In the analysis, I establish the formal identification and estimation result when unobserved heterogeneity prevails on both sides. This paper takes the first step to establish a structural model to estimate finite-sized Non-Transferable Utility matching markets with two-sided unobserved heterogeneity. I see at least two directions for future research. First, it would be substantial to account for workers' strategic behavior after matching with the jobs. I.e. there is a "post-matching" stage of the game, in which the match-specific outcomes are generated in a non-cooperative fashion. This requires developing a three-stage game where the last stage is workers' strategic behavior over exerting efforts. Second, researchers are concerned about the learning-by-doing phenomenon in such markets; i.e. when workers decide which job to take, they may hold a clear expectation that they may polish their skills, benefiting future jobs. This dynamic concern is essential if one wants to capture and perhaps improve the learning phenomenon in such markets but may introduce nontrivial theoretical complications to my current model. To sum, the model and econometric discussion in this paper provide a basis for better modeling the real-world matching market.

2.6 Tables and Figures

Table 2.1: Monte Carlo Simulation: Parameter Specification

Specification	β	γ	σ_η	σ_b	ξ_2	$\delta_p(1)$	$\delta_p(2)$	$\delta_p(3)$	$\delta_p(4)$	$\delta_y(1)$	$\delta_y(2)$	δ_{Z^*}	δ_ω
S1	1.2	-1	1	0.8	-1.2	0.1	0.8	0.2	0.9	0.1	0.9	0.3	0.5
S2	2	-2	0.5	0.8	-1.2	0.1	0.8	0.2	0.9	0.1	0.9	0.3	0.5

Table 2.2: Performance of the Likelihood-Based Estimation - Benchmark Case

	S1			S2		
	true value	bias	std.dev	true value	bias	std.dev
β	1.2	-0.00769	0.038696	2	-0.00686	0.06865
γ	-1	0.009296	0.049781	-2	0.010316	0.070334
σ_η	1	-0.011	0.065985	0.5	-0.00418	0.031742
σ_b	0.8	-0.00107	0.007794	0.8	-7.7E-05	0.00859
ξ_2	-1.2	-0.00019	0.002259	-1.2	-7.2E-06	0.002563
$\delta_p(1)$	0.1	0.001751	0.010896	0.1	0.001307	0.011692
$\delta_p(2)$	0.8	0.000186	0.01319	0.8	-0.00033	0.014436
$\delta_p(3)$	0.2	-0.00015	0.01741	0.2	-0.00084	0.018748
$\delta_p(4)$	0.9	0.00111	0.012722	0.9	0.001039	0.013752
$\delta_y(1)$	0.1	0.000492	0.005351	0.1	0.000372	0.005892
$\delta_y(2)$	0.9	0.003515	0.007733	0.9	0.002628	0.008298
δ_{Z^*}	0.3	-0.0003	0.007629	0.3	-0.00015	0.008518
δ_ω	0.5	-0.00048	0.01744	0.5	-3.8E-05	0.018804

Table 2.3: Performance of the Likelihood-Based Estimation - General Case

	true value	bias	std.dev
β	2	-0.01244	0.127455
γ	-2	0.029126	0.170246
σ_η	1	-0.03156	0.129736
σ_b	3	-0.01441	0.072752
ξ_2	-3	-0.00241	0.025119
$\delta_y(1)$	0.1	-0.00019	0.009647
$\delta_y(2)$	0.9	0.00025	0.013505
δ_{Z^*}	0.3	0.00177	0.013729
δ_ω	0.5	-0.01621	0.079816
$Pr(p_L \| Z^* = z_L, \omega = \omega_L)$	0.9	-0.00879	0.034099
$Pr(p_M \| Z^* = z_L, \omega = \omega_L)$	0.05	0.00564	0.029639
$Pr(p_L \| Z^* = z_H, \omega = \omega_L)$	0.7	-0.01372	0.059793
$Pr(p_M \| Z^* = z_H, \omega = \omega_L)$	0.1	0.005325	0.03103
$Pr(p_L \| Z^* = z_L, \omega = \omega_H)$	0.5	-0.0087	0.040982
$Pr(p_M \| Z^* = z_L, \omega = \omega_H)$	0.4	0.008471	0.038291
$Pr(p_L \| Z^* = z_H, \omega = \omega_H)$	0.1	-0.00534	0.048937
$Pr(p_M \| Z^* = z_H, \omega = \omega_H)$	0.2	0.001762	0.035481
$Pr(X^*_{\tau-1} = x_H \| X^*_{\tau-1} = x_L)$	0.1	4.84E-05	0.01417
$Pr(X^*_{\tau-1} = x_H \| X^*_{\tau-1} = x_H)$	0.9	0.000787	0.013756
λ	0.5	0.001182	0.014416
σ_1	2	0.007197	0.053475
σ_2	1.5	-0.00104	0.032806
$Pr(X^* = x_H)$	0.4	-0.00057	0.015655

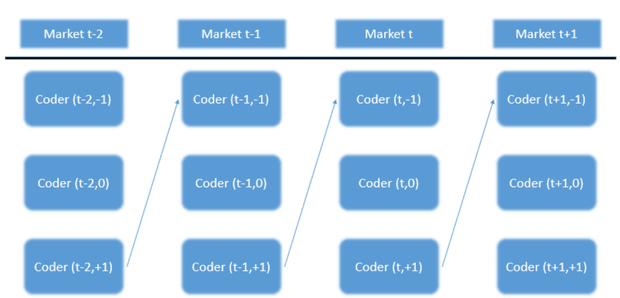


Table 2.4: Performance of the Likelihood-Based Estimation - Endogenous Cash Prize

	true value	bias	std.dev
β	2	0.015883	0.249372
γ	-2	-0.04571	0.324542
σ_η	1	0.023752	0.317615
σ_b	0.8	0.000221	0.016724
ξ_2	-1.2	0.000118	0.004728
$\delta_y(1)$	0.1	-0.00078	0.013076
$\delta_y(2)$	0.9	0.001045	0.016621
δ_{Z^*}	0.3	0.001649	0.014605
δ_ω	0.5	-0.20408	0.021261
$\delta_p(1)$	1	-1E-06	1.33E-06
$\delta_p(2)$	1.16E-06	-5.7E-07	7.71E-07
$\delta_p(3)$	0.1221	0.001095	0.021602
$\delta_p(4)$	0.285709	-0.00055	0.027849
$\delta_p(5)$	2.64E-07	1.45E-06	2.13E-06
$\delta_p(6)$	1	-2.7E-06	3.36E-06
$\delta_p(7)$	3.28E-08	2.77E-06	4.27E-06
$\delta_p(8)$	2.71E-08	2.54E-06	3.44E-06
c	1	-0.08326	0.45173

2.7 Appendix

2.7.1 Proof of Lemma 1

First, I prove that the firms' preferences represented by the profit functions in equation (3.1) are *responsive*, which is defined as follows:

Definition 3 *(Responsiveness) Firms' preference over a set of workers is responsive to the preference over individual workers if, for any set of workers C such that $|C| < Q$, and i, m, (i) firm strictly prefers $C \cup i$ to C if and only if it strictly prefers i to $\emptyset$; (ii) firms strictly prefers $C \cup i$ to $C \cup m$ if and only if it strictly prefers i to m.*

Let C_j denote the set of workers participating in job j. Without loss of generality, assume the first worker has the highest skill level. First, I check part (i) of the definition. The *ex post* profit of job j from hiring C_j is:

$$\pi^{post}_{j,C_j,t} = c \cdot \omega_t \cdot (Z^*_j \cdot \sum_{k \in C_j} X^*_k) - \sum_{k \in C_j} P_{kj}$$

whereas the utility of hiring C_j plus one more worker i is:

$$\pi^{post}_{j,C_j \cup i,t} = c \cdot \omega_t \cdot [Z^*_j \cdot (\sum_{k \in C_j} X^*_k + X^*_i)] - \sum_{k \in C_j \cup i} P_{kj}$$

Therefore,

$$\pi^{post}_{j,C_j\cup i,t} - \pi^{post}_{j,C_j,t} = \begin{cases} c\cdot\omega_t\cdot[Z^*_j\cdot(X^*_i)],\ \text{if } |C_j|\geq 2 \\ \\ c\cdot\omega_t\cdot[Z^*_j\cdot(X^*_i)] - 0.5P_j,\ \text{if } |C_j| = 1 \\ \\ c\cdot\omega_t\cdot[Z^*_j\cdot(X^*_i)] - P_j,\ \text{if } |C_j| = 0 \end{cases}$$

Remember that we have assumed all jobs weakly benefit from having one more worker, regardless of his/her type, which indicates that $\pi^{post}_{j,i,t} := c\cdot\omega_t\cdot[Z^*_j\cdot(X^*_i)] - P_j \geq 0, \forall$. This further implies $\pi^{post}_{j,C_j\cup i,t} - \pi^{post}_{j,C_j,t} \geq 0$; with the strict inequality holds if and only if $\pi^{post}_{j,i,t} > 0$. Part (i) in the definition is checked.

To check part (ii), I calculate the job's profit from having $C_j \cup m$, $m \notin C_j, m \neq i$.

$$\pi^{post}_{j,C_j\cup m,t} = c\cdot\omega_t\cdot[Z^*_j\cdot(\sum_{k\in C_j} X^*_k + X^*_m)] - \sum_{k\in C_j\cup m} P_{kj}$$

Therefore,

$$\pi^{post}_{j,C_j\cup i,t} - \pi^{post}_{j,C_j\cup m,t} = c\cdot\omega_t\cdot[Z^*_j\cdot(X^*_i - X^*_m)]$$

which indicates $\forall i, m, \pi^{post}_{j,C_j\cup i,t} - \pi^{post}_{j,C_j\cup m,t} > 0$ if and only if $X^*_i > X^*_m$. Part (ii) in the definition of responsiveness therefore is checked.

According to Lemma 1 in Roth and Sotomayor (1992), provided preferences are responsive (and strict), the many-to-one matching game share the same stable outcome with a corresponding one-to-one matching problem where each slot within a job is treated as the new "agent" on the firm side. Specifically, I enumerate define the set of slots of all jobs in the market as $K := \{k_1^1, k_1^2, \cdots, k_1^Q, \cdots, k_J^1, k_J^2, \cdots, k_J^Q\}$ for the ease of notation. Slot k_j^q represents the q-the slot in job j, and has an *ex post* profit function of:

$$\pi_{k_j^q}^{post} = \begin{cases} c \cdot \omega_t \cdot (Z_j^* \cdot X_i^*) - P_{i,k_j^q} \text{ if slot is filled by worker } i \\ \\ 0, \text{ otherwise} \end{cases}$$

Within a job, all slots share the same ordinal preference over workers. Consequently, we are describing a market with two sets of disjoint population, set of workers I and set of slots K. Furthermore, this market entails a commonly-known utility profile, $\mathcal{U} := \{\{u\}, \{\pi^{post}\}\}$. Each worker i gains a utility u_{ikt} if he is matched with slot k, whereas this slot gains $\pi_{k_j^q,i,t}^{post}$ from matching with worker i.

This proof for the existence and uniqueness of the stable matching is based on Lemma 1 in Roth and Sotomayor (1992) and Theorem 2 in Clark (2006), via the following deferred acceptance algorithm.

In market t, we can rank workers according to their desirabilities from slots' perspective, such that $X_1^* > X_2^* > \cdots > X_{|I_t|}^*$. First, we execute the slot-optimal Deferred Acceptance Algorithm:

Round 1: all slots propose to worker 1, the most preferred worker, and worker 1 chooses his most preferred slot, denoted as S_1 and hold this offer. Note that worker 1 knows any other offers in later rounds will be worse than S_1, so he would only hold the offer till the end and accept the offer eventually.

...

Round k: slots rejected at Round $k-1$ will propose to their best available choice, worker k, who will, in turn, choose his most favored available firm, denoted as S_k, and will not change her mind until the end of the game, as any later offers will be worse than S_k.

The outcome is pairwise stable in the sense that all workers get their best choices available to them and so are all slots. To see this, consider a worker i and a slot k that are not matched with each other. If worker i prefers slot k to her current slot, $\mu(i)$, then it must be that slot k is chosen before Round i, as otherwise, worker i will certainly choose k instead of $\mu(i)$ in that round. The slot k is chosen by another worker, $\mu(k)$, who is ranked higher than i. Consequently, even if worker i prefers slot k to $\mu(i)$, slot k would not agree to form a coalition with worker i, as k's current match $\mu(k)$ is better than i.

Now consider the case when slot k prefers i to $\mu(k)$, then it must have proposed an offer to worker i in Round i, earlier than Round $\mu(k)$, because i is ranked higher than $\mu(k)$. The only reason why i is not matched with k is that i chose $\mu(i)$ over k, representing that she strictly prefers $\mu(i)$ to k. Therefore, a coalition still cannot be formed by (i, k). As I assume the utility of being matched is better than staying unmatched, the two conditions of pairwise stability are both satisfied.

To prove the matching is unique, we consider a sequence of sets of workers and slots, $\{< I_n, J_n >\}$, $n = 1, 2, \cdots, N$, such that $< I_1, J_1 >:=< I, J >$ is the whole set of workers and slots. If worker i and slot j is the most preferred choice of each other, we define them as a *fixed pair*. In this setup, as both sides have strict preferences over the counterpart, the unique fixed pair is worker 1 and her best choice, S_1. We define $< I_2, J_2 >:=< I_1, J_1 > \setminus < 1, S_1 >$. In the subpopulation, $< I_2, J_2 >$, the unique fixed pair is worker 2 and her best choice, S_2, among all remaining slots. We define for any n, $< I_{n+1}, J_{n+1} >:=< I_n, J_n > \setminus < n, S_n >$. Then, the last element in the sequence, $< I_N, J_N >$ would consist of no workers and all unfavored slots. We, therefore, know that the sequence $\{< I_n, J_n >\}$ is uniquely defined from $< I, J >$.

Let μ be any stable matching of $< I, J >$, and let μ_n be a matching of the subpopulation $< I_n, J_n >$ such that $\mu_n(i) = \mu(i)$ for all $i \in I_n$, and $\mu_n(j) = \mu(j)$ for all $j \in J_n$. By Lemma 2(ii) in Clark (2006), as we know μ is a stable matching of $< I, J >$, then μ_n will be a stable matching of $< I_n, J_n >$. By construction, for any worker i, she will form a fixed pair with S_i in subpopulation $< I_i, J_i >$. Then by Lemma 2(i) in Clark (2006), $\mu_i(i) = S_i$ for the fixed pair $< i, S_i >$. Combining the results, we know that for any $< i, S_i >$ pair, $\mu(i) = S_i$, i.e., the matching μ is unique. Furthermore, by Roth (1986), the set of unmatched slots remains the same across different stable matchings. Therefore, the set $< I_N, J_N >$ will also be the same for different matchings. As a result, the stable matching μ is unique, and the proof completes.

2.7.2 Proof of Lemma 2

To prove Lemma 2, we check the two conditions specified in the definition of pairwise stability. First, the Individual Rationality condition is satisfied on the worker side as the set $M_i[\mu_t]$ always contains the empty set, i.e. staying outside the market. Thus, the worker's choice is always weakly better than staying unmatched. On the firm side, if a worker generates negative incremental profit for the firm, its vacant slots will be excluded from the worker's choice set. Thus, whoever hired by the firm must provide nonnegative incremental profit, satisfying the IR condition.

Second, if we consider a "blocking pair" such that a worker i switch to a job j that still have a vacant slot, and both are better off – this will never happen as job j is included in worker i's choice set when he makes his decision. The fact that he dismisses job j and chooses job $\mu(i)$ proves the nonexistence of such a blocking pair. Similar nonexistence argument follows when job j has no vacancy but is willing to replace a current worker with worker i.

2.7.3 Proof of Lemma 3

First, we define the best response correspondence for firm j given other firms' mixed strategy δ_{-j} to be:

$$B_j(\delta_{-j}) := \{\delta : \{z_1, z_2, \cdots, z_k\} \to \Sigma | \sum_{l=1}^{m} [\sum_{Z^*_{-j}, P_{-j}} \tilde{\pi}(p_l, Z^*_j, Z^*_{-j}, P_{-j}, \omega) \cdot Pr(Z^*_{-j}, P_{-j}|\delta_{-j})] \cdot \delta^l$$

$$\geq \sum_{l=1}^{m} [\sum_{P_{-j}} \tilde{\pi}(p_l, Z^*_j, Z^*_{-j}, P_{-j}, \omega) \cdot Pr(Z^*_{-j}, P_{-j}|\delta_{-j})] \cdot \delta'^l, \forall \delta' \in \Sigma\}$$

And the best response correspondence for all firms is defined as $\boldsymbol{B} := B_1 \times B_2 \cdots \times B_J$. Proof of the existence of a Bayesian Nash Equilibrium is equivalent of showing the existence of a fixed point in $\boldsymbol{B}$. According to Kakutani's theorem, we check the following conditions:

1. The set Σ is compact and convex.
2. $\boldsymbol{B}$ is nonempty for all δ.
3. $\boldsymbol{B}$ is convex-valued.
4. $\boldsymbol{B}$ has a closed graph.

For the first condition, I show that for each $\Sigma_j : \{\delta | \sum_{l=1}^{m} \delta^l = 1\}$ is a simplex of dimension $m - 1$ thus closed and bounded, i.e., compact. It is also easy to show Σ_j is convex. Thus, $\Sigma : \Sigma_1 \times \Sigma_2 \cdots \Sigma_J$ is also compact and convex.

For the second condition, as each Σ_j is nonempty, and the expected profit for firm j is linear in its mixed strategy δ_j, thus continuous in δ_j. Therefore, B_j is nonempty as well.

For the third condition, I show that for any j, B_j is convex valued. This is true as B_j is linear in δ_j. Thus, for any δ_{-j}, pick any pair $\delta, \delta' \in B_j(\delta_{-j})$, then by definition, they both maximize expected profit of firm j. Their linear combination also maximizes expected payoff. From linearity, the linear combination of expected payoffs equals the expected payoff of the linear combination of δ, δ'. Thus, the convexity condition is checked.

For the last condition, I prove by negation. Suppose there exists a sequence $(\delta^n, \hat{\delta}^n) \to (\delta, \hat{\delta})$ where $\hat{\delta}^n \in B(\delta^n)$ but $\hat{\delta} \notin B(\delta)$. That is, there exists $\tilde{\delta}_j$ and $\epsilon > 0$ such that

$$\sum_{l=1}^{m} \pi^{\text{int}}(\delta_{-j}) \cdot \tilde{\delta}^l > \sum_{l=1}^{m} \pi^{\text{int}}(\delta_{-j}) \cdot \delta^l + 3\epsilon$$

By the continuity of the expected profit function, we have

$$\sum_{l=1}^{m} \pi^{\text{int}}(\delta_{-j}^n) \cdot \tilde{\delta}^l > \sum_{l=1}^{m} \pi^{\text{int}}(\delta_{-j}) \cdot \tilde{\delta}^l - \epsilon > \sum_{l=1}^{m} \pi^{\text{int}}(\delta_{-j}) \cdot \delta^l + 2\epsilon$$

But also from the continuity of the expected profit function,

$$\sum_{l=1}^{m} \pi^{\text{int}}(\delta_{-j}) \cdot \delta^{l} > \sum_{l=1}^{m} \pi^{\text{int}}(\delta_{-j}^{n}) \cdot \hat{\delta}^{l,n} - \epsilon$$

This means,

$$\sum_{l=1}^{m} \pi^{\text{int}}(\delta_{-j}^{n}) \cdot \tilde{\delta}^{l} > \sum_{l=1}^{m} pi^{\text{int}}(\delta_{-j}^{n}) \cdot \hat{\delta}^{l,n} + \epsilon$$

which contradicts the fact that $\hat{\delta}^{n} \in B(\delta^{n})$. This completes the proof for B_j to have a closed graph. The product of all B_j's, $\boldsymbol{B}$ therefore has a closed graph, and combining all conditions, it has a fixed point on $\Sigma : \Sigma_1 \times \Sigma_2 \cdots \Sigma_J$.

2.7.4 Proof of Lemma 4

First, I prove that the probability of observing a maximum price increases with the cost coefficient c in a market where all slots are fulfilled. Suppose this is not true, then there exists two coefficient values c_A and c_B, such that $c_A > c_B$ and other things being all equal in the game, but $Pr(P = p_m|c_A) \leq Pr(P = p_m|c_B)$. Without loss of generality, let $Pr(P = p_{m-1}|c_A) \geq Pr(P = p_{m-1}|c_B)$. Then consider firm j in the game with c_B deviate from his current (equilibrium) mixed-strategy δ_B to δ'_B:

Instead of playing $Pr(P = p_m|c_B)$ on the highest price p_m, firm j decreases the probability to $Pr(P = p_m|c_A)$. Correspondingly, it increases the probability of playing p_{m-1} from $Pr(P = p_{m-1}|c_B)$ to $Pr(P = p_{m-1}|c_B) + Pr(P = p_m|c_B) - Pr(P = p_m|c_A)$. Conditional on ω, the difference between the expected payoff given δ' and δ therefore is:

$$
\begin{aligned}
&\pi_j^0(p_m, Z_j^*, c_B, \omega) \cdot Pr(P_j = p_m|\delta_B') - \pi_j^0(p_m, Z_j^*, c_B, \omega) \cdot Pr(P_j = p_m|\delta_B) + \\
&\pi_j^0(p_{m-1}, Z_j^*, c_B, \omega) \cdot Pr(P_j = p_{m-1}|\delta_B') - \pi_j^0(p_{m-1}, Z_j^*, c_B, \omega) \cdot Pr(P_j = p_{m-1}|\delta_B) \\
&= (\pi_j^0(p_m, Z_j^*, c_B, \omega) - \pi_j^0(p_{m-1}, Z_j^*, c_B, \omega)) \cdot \underbrace{(Pr(P = p_m|c_A) - Pr(P = p_m|c_B))}_{\leq 0} \\
&\leq 0
\end{aligned}
$$

and the last inequality follows from the fact that the Bayesian Nash Equilibrium mixed-strategy maximizes the expected payoff. Thus, the following expression has to be nonnegative:

$$(\pi_j^0(p_m, Z_j^*, c_B, \omega) - \pi_j^0(p_{m-1}, Z_j^*, c_B, \omega)) \geq 0$$

Next, I show the following inequality must hold for a bigger c_A:

$$(\pi_j^0(p_m, Z_j^*, c_A, \omega) - \pi_j^0(p_{m-1}, Z_j^*, c_A, \omega)) \geq 0$$

For this inequality to hold, a sufficient condition is that 1) the expected incremental revenue from play p_{m-1} to p_m is increasing in c, and 2) the expected incremental cost is decreasing in c. To see this, remember that when $c = c_A$, the other firms' equilibrium strategy is to put less weight on playing p_m than when $c = c_B$, i.e. $Pr(P = p_m|c_A) \leq Pr(P = p_m|c_B)$. Thus, when c increases from c_B to c_A, the *ex post* revenue $\mathcal{R}$ increases, and given others are now playing a less aggressive strategy, the *ex ante* expected revenue will also increase. This proves the first argument. To see why the second argument also holds, I show the following equality holds:

$$E(P|p_m, c_B) - E(P|p_{m-1}, c_B) = E(P|p_m, c_A) - E(P|p_{m-1}, c_A)$$

where $E(P|p, c)$ denotes firm j's expected payment when it is playing p and the cost coefficient is c, other things being fixed. As all slots are fulfilled in the market, the expected payment from firm j is $\sum_{k=1}^{Q} \phi_k p$ regardless of c, this the equality above holds.

This further indicates the following inequality:

$$\begin{aligned}
&[\pi_j^0(p_m, Z_j^*, c_A, \omega) \cdot Pr(P_j = p_m|\delta_A) - \pi_j^0(p_m, Z_j^*, c_A, \omega) \cdot Pr(P_j = p_m|\delta'_A)+ \\
&\pi_j^0(p_{m-1}, Z_j^*, c_A, \omega) \cdot Pr(P_j = p_{m-1}|\delta_A) - \pi_j^0(p_{m-1}, Z_j^*, c_A, \omega) \cdot Pr(P_j = p_{m-1}|\delta'_A)] \\
&= \underbrace{(\pi_j^0(p_m, Z_j^*, c_A, \omega) - \pi_j^0(p_{m-1}, Z_j^*, c_A, \omega))}_{\geq 0} \cdot \underbrace{(Pr(P = p_m|c_A) - Pr(P = p_m|c_B))}_{\leq 0} \\
&\leq 0
\end{aligned}$$

where the mixed strategy δ_A is the equilibrium mixed-strategy of firm j in the game with c_A, and δ'_A is a deviation such that instead of playing $Pr(P = p_m|c_A)$, the firm increases the probability to $Pr(P = p_m|c_B)$; instead of playing $Pr(P = p_{m-1}|c_A)$, the firm decreases the probability to $Pr(P = p_{m-1}|c_A)+Pr(P = p_m|c_A)-Pr(P = p_m|c_B)$.

Therefore, in the game with c_A, we have found a profitable deviation for firm j, given others' equilibrium mixed-strategy. This contradicts with the fact that δ_A is an equilibrium strategy. The negation indicates the monotonicity of $Pr(P = p_m|c)$ in c holds.

The proof for monotonicity in ω follows the same procedure, by fixing c.

2.7.5 Proof of Lemma 5

I show the following conditions hold:

1. $\boldsymbol{Y}_t \perp (\boldsymbol{P}_t, \widetilde{\boldsymbol{Z}}_t)|(\boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t)$

2. $\boldsymbol{Y}_t \perp \boldsymbol{Z}_t|(\boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t)$

3. $\boldsymbol{X}_t \perp \boldsymbol{Z}_t \perp \widetilde{\boldsymbol{Z}}_t|(\boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t)$

4. $\boldsymbol{X}_t \perp \boldsymbol{P}_t|(\boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t)$

5. $\boldsymbol{Z}_t \perp \boldsymbol{P}_t|(\boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t)$

where condition (2) and (3) indicates the conditional independence between $(\boldsymbol{Y}_t, \boldsymbol{X}_t)$ and $\boldsymbol{Z}_t$; condition (1), (3) and (4) indicates conditional independence between $(\boldsymbol{Y}_t, \boldsymbol{X}_t)$ and $(\boldsymbol{P}_t, \widetilde{\boldsymbol{Z}}_t)$; condition (3) and (5) indicates conditional independence between $\boldsymbol{Z}_t$ and $(\boldsymbol{P}_t, \widetilde{\boldsymbol{Z}}_t)$.

First, I show condition (1) and (2) hold. As the distribution of $\boldsymbol{Y}_t$ is only related with $\boldsymbol{Z}_t^*$, conditional on $\boldsymbol{Z}_t^*$, the variation in $\boldsymbol{Y}_t$ is completely independent of other covariates; the first two condition, therefore, is checked.

Next, I show condition (3) holds:

$$
\begin{aligned}
& f(\boldsymbol{X}_t, \boldsymbol{Z}_t, \widetilde{\boldsymbol{Z}}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t) \\
& = f(\boldsymbol{X}_t | \boldsymbol{Z}_t, \widetilde{\boldsymbol{Z}}_t, \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t) \cdot f(\boldsymbol{Z}_t | \widetilde{\boldsymbol{Z}}_t, \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t) \cdot f(\widetilde{\boldsymbol{Z}}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t) \\
& = f(\boldsymbol{X}_t | \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t) \cdot f(\boldsymbol{Z}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \mu_t) \cdot f(\widetilde{\boldsymbol{Z}}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t)
\end{aligned}
$$

The second equality holds because (i) for any individual worker i, conditional on $(\boldsymbol{X}^*, \mu_t)$, the variation in X_{it} is completely driven by a_{i,μ_t}, which is independent of $(\boldsymbol{Z}_t, \widetilde{\boldsymbol{Z}}_t)$; and (ii) conditional on $(\boldsymbol{Z}^*, \boldsymbol{X}^*, \mu_t)$, the variation in final score Z_{ij}^t is completely driven by b_{jt}, which is also independent of $(\boldsymbol{X}_t, \widetilde{\boldsymbol{Z}}_t)$. Thus, condition (3) is checked.

Similarly, we can check conditions (4) and (5) by showing the following:

$$
\begin{aligned}
f(\boldsymbol{X}_t, \boldsymbol{P}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t) &= f(\boldsymbol{X}_t | \boldsymbol{P}_t, \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t) \cdot f(\boldsymbol{P}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t) \\
&= f(\boldsymbol{X}_t | \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t) \cdot f(\boldsymbol{P}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \mu_t)
\end{aligned}
$$

and

$$
\begin{aligned}
& f(\boldsymbol{Z}_t, \boldsymbol{P}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t) \\
& = f(\boldsymbol{Z}_t | \boldsymbol{P}_t, \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t) \cdot f(\boldsymbol{P}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t) \\
& = f(\boldsymbol{Z}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \mu_t) \cdot f(\boldsymbol{P}_t | \boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \mu_t)
\end{aligned}
$$

As all five conditions are checked, the full conditional independence condition hold and the proof completes.

2.7.6 Proof of Proposition 2

First, I show equation (2.15) and (2.16) hold. That is, the following two equations hold:

$$
f(\boldsymbol{Z}^d | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q) = \prod_{j=1}^{J} f(Z_j^d | Z_j^*, \boldsymbol{X}_{\mu(j)}^*)
$$

and,

$$
f(\boldsymbol{Z}^*, \boldsymbol{X}^d = \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X}^*, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q) = \prod_{j=1}^{J} (f(Z_j^* | Y_j = y) \cdot f(\boldsymbol{X}_{\mu(j)}^d = \boldsymbol{x} | \boldsymbol{X}_{\mu(j)}^*, \tilde{\boldsymbol{X}}_{\mu(j)} = \tilde{\boldsymbol{x}}))
$$

The first equation holds because of part 3(a) in Assumption 4. The second equation holds because of Assumptions 3 and 5 – i.e., Y's are independent amongst different jobs in the same market; the rating scores are conditionally independent amongst workers belonging to different jobs in the market. Thus, by observing the joint distribution of $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{P}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z}, \tilde{\boldsymbol{X}})$ conditional on a certain (μ, I, J, Q), we could rewrite the main equations in the form of equation (2.17), and ultimately, we have the equation (2.18) for eigenvalue-eigenvector decomposition as in Hu (2008).

In order to uniquely determine the ordering of eigenvalues/vectors of the matrix on LHS, we imposed Assumption 7. To see how this works, let us first rewrite the diagonal matrices in its full form:

$$D_{Z_j^*|Y_j=y} \cdot D^{-1}_{Z_j^*|Y_j=1-y} = \begin{bmatrix} q_1, & & & \\ & q_2, & & \\ & & \ddots & \\ & & & q_m \end{bmatrix}$$

and

$$D_{\boldsymbol{X}^d_{\mu(j)}|\boldsymbol{X}^*_{\mu(j)}, \tilde{\boldsymbol{X}}=\tilde{\boldsymbol{x}}_{\mu(j)}} = \begin{bmatrix} c_1, & & & \\ & c_2, & & \\ & & \ddots & \\ & & & c_{l^{|\mu(j)|}} \end{bmatrix}$$

where q's are distinct values and takes up to m values in total, as Z^* could take up to m values. Similarly, as $\boldsymbol{X}^*_{\mu(j)}$ takes up to $l^{|\mu(j)|}$ distinct values, the second diagonal matrix is of dimension $l^{|\mu(j)|}$.

The upper-left element of $D_{Z_j^*|Y_j=y} \cdot D^{-1}_{Z_j^*|Y_j=1-y}$ is $q_1 := Pr(Z_j^* = z_1|Y_j = 1)/Pr(Z_j^* = z_1|Y_j = 0)$. From part (1) in Assumption 5, we know that, $Pr(Z_j^* = z_1|Y_j = 1)/Pr(Z_j^* = z_1|Y_j = 0)$ would be the largest element among all probabilities given different values of Z_j^*, which corresponds to this upper-left element, q_1. Similarly, the upper-left element of $D_{\boldsymbol{X}^d_{\mu(j)}|\boldsymbol{X}^*_{\mu(j)},\tilde{\boldsymbol{X}}=\tilde{\boldsymbol{x}}_{\mu(j)}}$ is $c_1 := Pr(\boldsymbol{X}^d_{\mu(j)} = \boldsymbol{x}|\boldsymbol{X}^*_{\mu(j)} = \{x_1, x_1, \cdots, x_1\}, \tilde{\boldsymbol{X}} = \tilde{\boldsymbol{x}}_{\mu(j)})$. If we choose $\boldsymbol{x} = \{x_1, x_1, \cdots, x_1\}$, then according to part (2) in Assumption 7, the largest number should correspond to this upper-left element, c_1.

Combining the two results together, we know that when we conduct eigenvalue-eigenvector decomposition for the matrix $s \cdot M_j$ where $\boldsymbol{X}^d_{\mu(j)} = \{x_1, x_1, \cdots, x_1\}$, then the largest number in eigenvalue corresponds to $q_1 \cdot c_1$.

Following the same logic, we could identify $q_1 \cdot c_2$ through $q_1 \cdot c_{l^{|\mu(j)|}}$ by conducting the eigenvalue-eigenvector decomposition for matrix $s \cdot M_j$ where $\boldsymbol{X}^d_{\mu(j)} = \{x_1, x_1, \cdots, x_2\}$ through $\boldsymbol{X}^d_{\mu(j)} = \{x_l, x_l, \cdots, x_l\}$ simply by picking the largest number in eigenvalues.

Now that we have identified $q_1 \cdot c_1$ through $q_1 \cdot c_{l^{|\mu(j)|}}$ and their corresponding eigenvectors, we could put them aside and look at the rest of the matrix. For instance, when we decompose $s \cdot M_j$ where $\boldsymbol{X}^d_{\mu(j)} = \{x_1, x_1, \cdots, x_1\}$, the largest number in eigenvalues that does *not* belong to $q_1 \cdot c_1$ through $q_1 \cdot c_{l^{|\mu(j)|}}$ should correspond to $q_2 \cdot c_1$, as q_2 is the largest number except for q_1 in terms of $Pr(Y_j = 1|Z_j^*)$. We therefore could identify $q_2 \cdot c_1$ through $q_2 \cdot c_{l^{|\mu(j)|}}$. After putting aside these identified elements, we are able to identify (in descending sequence) all $q_k \cdot c_1$ through $q_k \cdot c_{l^{|\mu(j)|}}$

for all $k = 3, 4, \cdots, m$.

Given that we identify every single element of the eigenvalue diagonal matrix, the eigenvalue-eigenvector decomposition is unique up to a normalization of each column in the eigenvector matrix. This indicates that, the matrix $M_{\boldsymbol{Z}_j^d|Z_j^*,\boldsymbol{X}^*_{\mu(j)}}$ is nonparametrically identified for each job j. Consequently, the matrix $M_{\boldsymbol{Z}^d|\boldsymbol{Z}^*,\boldsymbol{X}^*,\tilde{\boldsymbol{X}},\mu,I,J,Q}$ is also identified from equation (2.15). Note that $M_{\boldsymbol{Z}^d|\boldsymbol{Z}^*,\boldsymbol{X}^*,\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q}$ is invertible as matrix $M_{(\boldsymbol{P}^d,\widetilde{\boldsymbol{Z}}^d,\boldsymbol{Z}^d|\tilde{\boldsymbol{X}}=\boldsymbol{x},\mu,I,J,Q)}$ is of full rank according to Assumption 6.

For any value $\boldsymbol{P} = \boldsymbol{p}$, the following equation holds:

$$Pr(\boldsymbol{Z}^d, \boldsymbol{P} = \boldsymbol{p}|\mu, I, J, Q) = \sum_{\boldsymbol{Z}^*,\boldsymbol{X}^*} Pr(\boldsymbol{Z}^d|\boldsymbol{Z}^*, \boldsymbol{X}^*, \mu, I, J, Q) \cdot Pr(\boldsymbol{P} = \boldsymbol{p}, \boldsymbol{Z}^*, \boldsymbol{X}^*|\mu, I, J, Q) \tag{2.38}$$

Thus, if we write the above equation into vector and matrix form:

$$V_{\boldsymbol{Z}^d,\boldsymbol{P}=\boldsymbol{p}|\mu,I,J,Q} = M_{\boldsymbol{Z}^d|\boldsymbol{Z}^*,\boldsymbol{X}^*,\mu,I,J,Q} \cdot V_{\boldsymbol{P}=\boldsymbol{p},\boldsymbol{Z}^*,\boldsymbol{X}^*|\mu,I,J,Q} \tag{2.39}$$

where each element of the LHS vector denotes a probability of a distinct value for $\boldsymbol{Z}^d$ and the dimension is $(m^J \cdot l^I) \times 1$. On the RHS, the first matrix is of dimension $(m^J \cdot l^I) \times (m^J \cdot l^I)$ and the second vector is of dimension $(m^J \cdot l^I) \times 1$.

Thus, we could identify the distribution of $(\boldsymbol{P}, \boldsymbol{Z}^*, \boldsymbol{X}^*|\mu, I, J, Q)$ from:

$$V_{\boldsymbol{P}=\boldsymbol{p},\boldsymbol{Z}^*,\boldsymbol{X}^*|\mu,I,J,Q} = M^{-1}_{\boldsymbol{Z}^d|\boldsymbol{Z}^*,\boldsymbol{X}^*,\mu,I,J,Q} \cdot V_{\boldsymbol{Z}^d,\boldsymbol{P}=\boldsymbol{p}|\mu,I,J,Q} \tag{2.40}$$

This completes the proof.

2.7.7 Proof of Corollary 1

As shown in the proof of Theorem 2, the conditional distribution $Pr(\boldsymbol{P}^d, \widetilde{\boldsymbol{Z}}^d, \boldsymbol{Z}^*, \boldsymbol{X}^*|\tilde{\boldsymbol{X}}, \mu, I, J, Q)$ is identified. For each value of $\boldsymbol{Z} = \boldsymbol{z}$, the following condition holds:

$$Pr(\boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{P}^d|\mu, I, J, Q) = \sum_{\boldsymbol{Z}^*,\boldsymbol{X}^*} Pr(\boldsymbol{Z} = \boldsymbol{z}|\boldsymbol{Z}^*, \boldsymbol{X}^*, \mu, I, J, Q) \cdot Pr(\boldsymbol{P}^d, \boldsymbol{Z}^*, \boldsymbol{X}^*|\mu, I, J, Q) \tag{2.41}$$

If we rewrite it into matrix form:

$$V_{\boldsymbol{Z}=\boldsymbol{z},\boldsymbol{P}^d|\mu,I,J,Q} = M_{\boldsymbol{P}^d,\boldsymbol{Z}^*,\boldsymbol{X}^*|\mu,I,J,Q} \cdot V_{\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{Z}^*,\boldsymbol{X}^*,\mu,I,J,Q} \tag{2.42}$$

Again, all the V-vectors are of dimension $(m^J \cdot l^I) \times 1$ and the M-matrix is of dimension $(m^J \cdot l^I) \times (m^J \cdot l^I)$. Therefore, the probability $Pr(\boldsymbol{Z}|\boldsymbol{Z}^*, \boldsymbol{X}^*, \mu)$ could be identified from:

$$V_{\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{Z}^*,\boldsymbol{X}^*,\mu,I,J,Q} = M^{-1}_{\boldsymbol{P}^d|\boldsymbol{Z}^*,\boldsymbol{X}^*,\mu,I,J,Q} \cdot V_{\boldsymbol{P}^d,\boldsymbol{Z}=\boldsymbol{z}|\mu,I,J,Q} \tag{2.43}$$

Once $Pr(\boldsymbol{Z}|\boldsymbol{Z}^*,\boldsymbol{X}^*,\mu,I,J,Q)$ is identified, the job-level conditional probability $Pr(\boldsymbol{Z}_j|Z_j^*,\boldsymbol{X}^*_{\mu(j)})$ could also be identified from the market with only one job.

Next, we could identify $Pr(\boldsymbol{Z},\boldsymbol{X}^*|\mu,I,J,Q)$ by:

$$Pr(\boldsymbol{Z},\boldsymbol{X}^*|\widetilde{\boldsymbol{X}},\mu,I,J,Q) = \sum_{\boldsymbol{Z}^*}\sum_{\boldsymbol{P}^d} Pr(\boldsymbol{Z}|\boldsymbol{Z}^*,\boldsymbol{X}^*,\mu,I,J,Q)\cdot Pr(\boldsymbol{P}^d,\boldsymbol{Z}^*,\boldsymbol{X}^*|\widetilde{\boldsymbol{X}},\mu,I,J,Q) \tag{2.44}$$

Then, for any $\boldsymbol{X}=\boldsymbol{x}$, we have the following equation:

$$Pr(\boldsymbol{Z},\boldsymbol{X}=\boldsymbol{x}|\widetilde{\boldsymbol{X}},\mu,I,J,Q) = \sum_{\boldsymbol{X}^*} Pr(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{X}^*,\widetilde{\boldsymbol{X}},\mu,I,J,Q)\cdot Pr(\boldsymbol{Z},\boldsymbol{X}^*|\widetilde{\boldsymbol{X}},\mu,I,J,Q) \tag{2.45}$$

As $\boldsymbol{Z}$ is continuously distributed, we could discretize its so as to make the matrix $M_{\boldsymbol{Z}^{dd},\boldsymbol{X}^*|\widetilde{\boldsymbol{X}},\mu,I,J,Q}$ invertible and is of dimension $l^I \times l^I$. Therefore, the distribution of

$(\boldsymbol{X} = \boldsymbol{x}|\boldsymbol{X}^*, \mu, I, J, Q)$ could be recovered from:

$$V_{\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{X}^*,\widetilde{\boldsymbol{X}},\mu,I,J,Q} = M^{-1}_{\boldsymbol{Z}^{dd},\boldsymbol{X}^*|\widetilde{\boldsymbol{X}},\mu,I,J,Q} \cdot V_{\boldsymbol{Z}^{dd},\boldsymbol{X}=\boldsymbol{x}|\widetilde{\boldsymbol{X}},\mu,I,J,Q} \tag{2.46}$$

Again, for each job j, the probability of $Pr(\boldsymbol{X}_{\mu(j)}|\boldsymbol{X}^*_{\mu(j)}, \widetilde{\boldsymbol{X}}_{\mu(j)})$ could be recovered from observing markets with only one job.

Lastly, the law of motion $Pr(X^*_i|\widetilde{X}^*_i)$ could be recovered from the following equation:

$$Pr(Z_i, \widetilde{Z}_i) = \sum_{X^*_i, \widetilde{X}^*_i} Pr(Z_i|X^*_i) \cdot Pr(\widetilde{Z}_i|\widetilde{X}^*_i) \cdot Pr(X^*_i, \widetilde{X}^*_i) \tag{2.47}$$

where, $Pr(Z_i|X^*_i) = Pr(Z_i, X^*_i)/\sum_{Z_i} Pr(Z_i, X^*_i)$ and the joint distribution.$Pr(Z_i, X^*_i)$ could be recovered from $Pr(\boldsymbol{Z}, \boldsymbol{X}^*|\widetilde{\boldsymbol{X}}, \mu, I, J, Q)$. Note that due to the data generating process, $Pr(\widetilde{Z}_i|\widetilde{X}^*_i)$ follows the same distribution as $Pr(Z_i|X^*_i)$. Thus, the joint distribution of $(X^*_i, \widetilde{X}^*_i)$ could be recovered from:

$$V_{X^*_i,\widetilde{X}^*_i} = [M_{Z^{dd}_i|X^*_i} \otimes M_{\widetilde{Z}^{dd}_i|\widetilde{X}^*_i}]^{-1} \cdot V_{Z^{dd}_i,\widetilde{Z}^{dd}_i} \tag{2.48}$$

The matrix in the middle is invertible because $M_{Z^{dd}_i|X^*_i}$ is invertible. This is true as $M_{Z^{dd}_i|X^*_i} \cdot D_{X^*_i} = M_{Z^{dd}_i,X^*_i}$, and the matrix on the RHS is invertible from previous proof. Thus, the joint distribution of $(X_i^*, \widetilde{X}_i^*)$ is identified.

Lastly, the initial condition refers to the situation where no worker in the job carries a rating score before participation – i.e., everyone is a newbie. Thus, the distribution of rating score degenerates to $Pr(X_{it}|X_{it}^*)$. For a given value of X_{it}, we have the following equation:

$$Pr(Z_{it}, X_{it}) = \sum_{X_{it}^*} Pr(Z_{it}, X_{it}^*) \cdot Pr(X_{it}|X_{it}^*) \tag{2.49}$$

Therefore, if we discretize Z_{it} such that the matrix $M_{Z_{it}^{dd}, X_{it}^*}$ is of full rank, the distribution of $Pr(X_{it}|X_{it}^*)$ could be recovered from the following:

$$V_{X_{it}|X_{it}^*} = M^{-1}_{Z_{it}^{dd}, X_{it}^*} \cdot V_{Z_{it}^{dd}, X_{it}} \tag{2.50}$$

2.7.8 Proof of Proposition 3

According to Theorem 2, conditional on (μ, I, J, Q), we are able to nonparametrically identify $M_{\boldsymbol{P}, \boldsymbol{Z}^*, \boldsymbol{X}^*|\mu, I, J, Q}$. The following joint distribution can therefore be identified:

$$f(\boldsymbol{P}, \boldsymbol{Z}^*|\mu, I, J, Q) = \sum_{\boldsymbol{X}^*} f(\boldsymbol{P}, \boldsymbol{Z}^*, \boldsymbol{X}^*|\mu, I, J, Q)$$

hence,

$$f(\boldsymbol{P}|\boldsymbol{Z}^*, \mu, I, J, Q) = \frac{f(\boldsymbol{P}, \boldsymbol{Z}^*|\mu, I, J, Q)}{\sum_{\boldsymbol{P}} f(\boldsymbol{P}, \boldsymbol{Z}^*|\mu, I, J, Q)}$$

then,

$$f(\boldsymbol{P}|\boldsymbol{Z}^*, J) = \sum_{\mu, I, Q} f(\boldsymbol{P}|\boldsymbol{Z}^*, \mu, I, J, Q) \cdot f(\mu, I, Q|J)$$

Given $f(\boldsymbol{P}|\boldsymbol{Z}^*, J)$ and the conditional independence condition (3.30), we obtain equation (2.21) and (2.22) conditional on $J \geq 3$:

$$f(P_1, P_2, P_3|Z_1^*, Z_2^*, Z_3^*) = \sum_{\omega} f(P_1|Z_1^*, \omega) \cdot f(P_2|Z_2^*, \omega) \cdot f(P_1, \omega|Z_3^*)$$

as well as,

$$f(P_1, P_3|Z_1^*, Z_2^*, Z_3^*) = \sum_{\omega} f(P_1|Z_1^*, \omega) \cdot f(P_1, \omega|Z_3^*)$$

Next, we write down the following matrix equations:

$$M_{\rho(P_1), P_2=p_m, \rho(P_3)|Z_1^*, Z_2^*, Z_3^*} = M_{\rho(P_1)|Z_1^*, \omega} \cdot D_{P_2=p_m|Z_2^*, \omega} \cdot M^T_{\rho(P_3), \omega|Z_3^*} \tag{2.51}$$

and,

$$M_{\rho(P_1), \rho(P_3)|Z_1^*, Z_2^*, Z_3^*} = M_{\rho(P_1)|Z_1^*, \omega} \cdot M^T_{\rho(P_3), \omega|Z_3^*} \tag{2.52}$$

where,

$$M_{\rho(P_1),P_2=p_m,\rho(P_3)|Z_1^*,Z_3^*} := [\rho(P_1) = p, P_2 = p_m, \rho(P_3) = p'|Z_1^*, Z_3^*]_{p,p'\in\{1,2,\cdots,n\}};$$

$$M_{\rho(P_1),\rho(P_3)|Z_1^*,Z_3^*} := [Pr(\rho(P_1) = p, \rho(P_3) = p'|Z_1^*, Z_3^*)]_{p,p'\in\{1,2,\cdots,n\}};$$

$$M_{\rho(P_j)|Z_j^*,\omega} := [Pr(\rho(P_j) = p|\omega = w, Z_j^*)]_{p=\{1,2,\cdots,n\};w=\{w_1,w_2,\cdots,w_n\}}, j = 1, 3;$$

$$D_{P_2=p_m|Z_2^*,\omega} := diag(Pr(P_2 = p_m|\omega = w, Z_2^*))_{w=\{w_1,w_2,\cdots,w_n\}}.$$

According to part (2) in Assumption 9, $M_{\rho(P_1),\rho(P_3)|Z_1^*,Z_3^*}$ is of full rank. Therefore, we take the inverse of the LHS matrix in equation (2.52) and right-multiply the LHS matrix in equation (2.51), and get:

$$M_{\rho(P_1),P_2=p_m,\rho(P_3)|Z_1^*,Z_2^*,Z_3^*} \cdot M^{-1}_{\rho(P_1),\rho(P_3)|Z_1^*,Z_2^*,Z_3^*} = M_{\rho(P_1)|Z_1^*,\omega} \cdot D_{P_2=p_m|Z_2^*,\omega} \cdot M^{-1}_{\rho(P_1)|Z_1^*,\omega} \tag{2.53}$$

According to Corollary 1, the conditional probability $Pr(P_2 = p_m|Z_2^*, \omega)$ is monotonically increasing with ω. After the eigenvalue–eigenvector decomposition as in Hu (2008), we could rank all numbers in the eigenvalue matrix in an ascending order, and uniquely determine the position of each diagonal element in $D_{P_2=p_m|Z_2^*,\omega}$, thereby identifying the distribution of $Pr(P_2 = p_m|Z_2^*, \omega)$ for each value of Z_2^*. Moreover, we are also able to identify the eigenvector matrix $M_{\rho(P_1)|Z_1^*,\omega}$, from which we could identify $Pr(P_2 = p|Z_2^*, \omega), p = \{p_1, p_2, \cdots, p_{m-1}\}$. To see this,

$$D_{P_2=p|Z_2^*,\omega} = \underbrace{M^{-1}_{\rho(P_1)|Z_1^*,\omega}}_{\text{identified}} \cdot \underbrace{[M_{\rho(P_1),P_2=p,\rho(P_3)|Z_1^*,Z_2^*,Z_3^*} \cdot M^{-1}_{\rho(P_1),\rho(P_3)|Z_1^*,Z_2^*,Z_3^*}]}_{\text{computed from data}} \cdot \underbrace{M_{\rho(P_1)|Z_1^*,\omega}}_{\text{identified}},$$

$$p = \{p_1, p_2, \cdots, p_{m-1}\}$$

To summarize, we are able to identify the distribution of cash prize P_j given different values of (Z_j^*, ω).

In order to identify the marginal distribution of ω, first notice that,

$$Pr(\rho(P_j)|\omega) = \sum_{Z_j^*} Pr(\rho(P_j)|Z_j^*, \omega) \cdot Pr(Z_j^*|\omega)$$

where, $Pr(\rho(P_j)|Z_j^*, \omega)$ is directly identified from the decomposition; $Pr(Z_j^*|\omega) = Pr(Z_j^*)$ due to the independence assumption, and could be identified from $Pr(\boldsymbol{Z}^*|J)$.

Then, I write down the following equation in matrix form:

$$\underbrace{\begin{bmatrix} Pr(\rho(P)=1) \\ Pr(\rho(P)=2) \\ \vdots \\ Pr(\rho(P)=n) \end{bmatrix}}_{\text{data}} = \underbrace{\begin{bmatrix} Pr(\rho(P)=1|\omega=w_1) & Pr(\rho(P)=1|\omega=w_2) & \cdots & Pr(\rho(P)=1|\omega=w_n) \\ Pr(\rho(P)=2|\omega=w_1) & Pr(\rho(P)=2|\omega=w_2) & \cdots & Pr(\rho(P)=2|\omega=w_n) \\ \vdots & \vdots & \ddots & \vdots \\ Pr(\rho(P)=n|\omega=w_1) & Pr(\rho(P)=n|\omega=w_2) & \cdots & Pr(\rho(P)=n|\omega=w_n) \end{bmatrix}}_{\text{Identified; of full rank}} \cdot \begin{bmatrix} Pr(\omega=w_1) \\ Pr(\omega=w_2) \\ \vdots \\ Pr(\omega=w_n) \end{bmatrix}$$

Thus, $Pr(\omega)$ is identified from the following equation:

$$\begin{bmatrix} Pr(\omega=w_1) \\ Pr(\omega=w_2) \\ \vdots \\ Pr(\omega=w_n) \end{bmatrix} = \begin{bmatrix} Pr(\rho(P)=1|\omega=w_1) & Pr(\rho(P)=1|\omega=w_2) & \cdots & Pr(\rho(P)=1|\omega=w_n) \\ Pr(\rho(P)=2|\omega=w_1) & Pr(\rho(P)=2|\omega=w_2) & \cdots & Pr(\rho(P)=2|\omega=w_n) \\ \vdots & \vdots & \ddots & \vdots \\ Pr(\rho(P)=n|\omega=w_1) & Pr(\rho(P)=n|\omega=w_2) & \cdots & Pr(\rho(P)=n|\omega=w_n) \end{bmatrix}^{-1} \cdot \begin{bmatrix} Pr(\rho(P)=1) \\ Pr(\rho(P)=2) \\ \vdots \\ Pr(\rho(P)=n) \end{bmatrix}$$

2.7.9 Proof for Lemma 7

We define the Maximum Likelihood Estimator as:

$$\begin{aligned} \widehat{\boldsymbol{\theta}} &:= argmax_{\boldsymbol{\theta}} \frac{1}{T}\sum_{t=1}^{T} \log Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*, \boldsymbol{\theta}) \\ &= argmax_{\boldsymbol{\theta}} [\frac{1}{T}\sum_{t=1}^{T} \log Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*, \boldsymbol{\theta}) - \frac{1}{T}\sum_{t=1}^{T} \log Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*, \boldsymbol{\theta_0})] \\ &= argmax_{\boldsymbol{\theta}} [\frac{1}{T}\sum_{t=1}^{T} \log \underbrace{(Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*, \boldsymbol{\theta}) / Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*, \boldsymbol{\theta_0}))}_{:=Q(\boldsymbol{\theta})/Q(\boldsymbol{\theta_0})}] \end{aligned} \tag{2.54}$$

By the Strong Law of Large Numbers,

$$\frac{1}{T}\sum_{t=1}^{T} \log Q(\boldsymbol{\theta})/Q(\boldsymbol{\theta_0}) \overset{a.s.}{\to} \int_{\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t} [\log Q(\boldsymbol{\theta})/Q(\boldsymbol{\theta_0})] dF(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*, \boldsymbol{\theta_0}) \leq 0 \tag{2.55}$$

On the RHS, it is known as Kullback-Leibler Divergence, which could be proved to be non-positive by applying Jensen's Inequality. This indicates that for any $\boldsymbol{\theta}$ the following holds:

$$\lim_{T\to\infty} Pr(|\frac{1}{T}\sum_{t=1}^{T} \log Q(\boldsymbol{\theta})/Q(\boldsymbol{\theta_0}) - \int_{\mu_t,\boldsymbol{Z}_t,\boldsymbol{P}_t,\boldsymbol{Y}_t} [\log Q(\boldsymbol{\theta})/Q(\boldsymbol{\theta_0})] dF(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*, \boldsymbol{\theta_0})| = 0) = 1 \tag{2.56}$$

As the space for the primitives are compact, and Q is continuous in $\boldsymbol{\theta}$, we have

$$\lim_{T\to\infty} Pr(\sup_{\boldsymbol{\theta}} |\frac{1}{T}\sum_{t=1}^{T} \log Q(\boldsymbol{\theta})/Q(\boldsymbol{\theta_0}) - \int_{\mu_t,\boldsymbol{Z}_t,\boldsymbol{P}_t,\boldsymbol{Y}_t} [\log Q(\boldsymbol{\theta})/Q(\boldsymbol{\theta_0})] dF(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*, \boldsymbol{\theta_0})| = 0) = 1 \tag{2.57}$$

As the RHS of equation (2.56) equals zero iff $Q(\boldsymbol{\theta}) = Q(\boldsymbol{\theta_0})$, given we have achieved identification, it holds that $Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*, \boldsymbol{\theta}) = Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t | \boldsymbol{X}_t^*, \boldsymbol{\theta_0}))$ iff $\boldsymbol{\theta} = \boldsymbol{\theta_0}$. Thus, $\boldsymbol{\theta_0}$ is the maximizer of the objective function in equation (2.58) when $T \to \infty$, which means it equals $\boldsymbol{\theta}$ when $T \to \infty$. This completes the proof.

2.7.10 Proof for Lemma 8

We define the Maximum Likelihood Estimator as:

$$\begin{aligned}
\widehat{\boldsymbol{\theta}} &:= argmax_{\boldsymbol{\theta}} \frac{1}{T/2}\sum_{t=1}^{T/2} \log Pr(\boldsymbol{W}_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}, \mu_{2t-2,+1}], [\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}, \mu_{2t,-1}], \boldsymbol{\theta}) \\
&= argmax_{\boldsymbol{\theta}} \frac{1}{T/2}\sum_{t=1}^{T/2} \log \frac{Pr(\boldsymbol{W}_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}, \mu_{2t-2,+1}], [\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}, \mu_{2t,-1}], \boldsymbol{\theta})}{Pr(\boldsymbol{W}_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}, \mu_{2t-2,+1}], [\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}, \mu_{2t,-1}], \boldsymbol{\theta_0})}
\end{aligned} \tag{2.58}$$

By the Strong Law of Large Numbers,

$$\frac{1}{T/2}\sum_{t=1}^{T/2}\log\frac{Pr(\boldsymbol{W}_{2t-1}|[\boldsymbol{Z}_{2t-2,+1},\boldsymbol{X}_{2t-2,+1},\mu_{2t-2,+1}],[\boldsymbol{Z}_{2t,-1},\boldsymbol{X}_{2t,-1},\mu_{2t,-1}],\boldsymbol{\theta})}{Pr(\boldsymbol{W}_{2t-1}|[\boldsymbol{Z}_{2t-2,+1},\boldsymbol{X}_{2t-2,+1},\mu_{2t-2,+1}],[\boldsymbol{Z}_{2t,-1},\boldsymbol{X}_{2t,-1},\mu_{2t,-1}],\boldsymbol{\theta_0})}\xrightarrow{a.s.}$$
$$\int_{\boldsymbol{W}_t}[\log\frac{Pr(\boldsymbol{W}_{2t-1}|[\boldsymbol{Z}_{2t-2,+1},\boldsymbol{X}_{2t-2,+1},\mu_{2t-2,+1}],[\boldsymbol{Z}_{2t,-1},\boldsymbol{X}_{2t,-1},\mu_{2t,-1}],\boldsymbol{\theta})}{Pr(\boldsymbol{W}_{2t-1}|[\boldsymbol{Z}_{2t-2,+1},\boldsymbol{X}_{2t-2,+1},\mu_{2t-2,+1}],[\boldsymbol{Z}_{2t,-1},\boldsymbol{X}_{2t,-1},\mu_{2t,-1}],\boldsymbol{\theta_0})}]$$
$$dF(\boldsymbol{W}_t||[\boldsymbol{Z}_{2t-2,+1},\boldsymbol{X}_{2t-2,+1},\mu_{2t-2,+1}],[\boldsymbol{Z}_{2t,-1},\boldsymbol{X}_{2t,-1},\mu_{2t,-1}],\boldsymbol{\theta_0}))\leq 0 \tag{2.59}$$

On the RHS, it is known as Kullback-Leibler Divergence, which could be proved to be non-positive by applying Jensen's Inequality, and it equals zero iff

$$Pr(\boldsymbol{W}_{2t-1}|[\boldsymbol{Z}_{2t-2,+1},\boldsymbol{X}_{2t-2,+1},\mu_{2t-2,+1}],[\boldsymbol{Z}_{2t,-1},\boldsymbol{X}_{2t,-1},\mu_{2t,-1}],\boldsymbol{\theta})=$$
$$Pr(\boldsymbol{W}_{2t-1}|[\boldsymbol{Z}_{2t-2,+1},\boldsymbol{X}_{2t-2,+1},\mu_{2t-2,+1}],[\boldsymbol{Z}_{2t,-1},\boldsymbol{X}_{2t,-1},\mu_{2t,-1}],\boldsymbol{\theta_0}) \tag{2.60}$$

Given we have achieved identification, this equality holds iff $\boldsymbol{\theta}=\boldsymbol{\theta_0}$. Thus, $\boldsymbol{\theta_0}$ is the maximizer of the log-likelihood function when $T\to\infty$, which means it equals $\boldsymbol{\theta}$ when $T\to\infty$. This completes the proof.

2.7.11 Detailed Monte Carlo Results

In this section, I present the Monte Carlo simulation results for the basic model in the next two tables, where workers' types X^* are observable to the econometrician. The starting values are set to be uninformative. We would like to evaluate the performance of our estimation strategy, as this strategy will be used in the empirical part. The results are less accurate as expected but in general performing well.

Table 2.5: Performance of the Likelihood-Based Estimation - Simple Case (bad starting point)

	starting value	true value	bias	std.dev
β	1	2	-0.03143	0.061076
γ	-1	-2	0.040247	0.062758
sigmaE	0.75	0.5	-0.0107	0.02987
sigmaV	1.2	0.8	-0.00018	0.00767
PARAB	-1.08	-1.2	-0.00011	0.00227
PP	0.09	0.1	0.001156	0.0108
PP	0.72	0.8	-0.00069	0.013004
PP	0.18	0.2	-0.0005	0.017138
PP	0.45	0.9	-0.00263	0.014349
PY	0.05	0.1	0.000594	0.005418
PY	0.45	0.9	0.002019	0.007875
PZ0	0.15	0.3	-0.00032	0.007608
POMEGA	0.25	0.5	0.000225	0.017319

Table 2.6: Performance of the Likelihood-Based Estimation - Simple Case (very bad starting point)

	starting value	true value	bias	std.dev
β	2	2	0.011772	0.091076
γ	-2	-2	-0.0128	0.096681
sigmaE	0.5	0.5	0.000208	0.039282
sigmaV	0.8	0.8	0.000212	0.011857
PARAB	-4	-1.2	-0.00052	0.003055
PP	0.1	0.1	-0.00017	0.01525
PP	0.8	0.8	4.2E-05	0.017969
PP	0.2	0.2	-0.00153	0.024809
PP	0.9	0.9	0.002219	0.018691
PY	0.1	0.1	0.000266	0.007567
PY	0.9	0.9	-0.00054	0.010963
PZ0	0.3	0.3	0.000971	0.010876
POMEGA	0.5	0.5	0.000929	0.024391

2.7.12 Detailed Derivation of General Case Log-Likelihood Function

We have equation (3.24) for the log-likelihood function in the general case:

$$\widetilde{\boldsymbol{LL}}(\theta) = \sum_{t=1}^{T/2} \log \sum_{\boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1}} Pr(\boldsymbol{W}_{2t-1}, \boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}, \mu_{2t-2,+1}], [\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}, \mu_{2t,-1}])$$

$$= \sum_{t=1}^{T/2} \log \sum_{\boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1}} Pr(\boldsymbol{Z}_{2t-1} | \boldsymbol{Z}^*_{2t-1}, \boldsymbol{X}^*_{2t-1}, \mu_{2t-1}) \cdot Pr(\boldsymbol{Y}_{2t-1} | \boldsymbol{X}_{2t-2,+1}, \boldsymbol{Z}^*_{2t-1}, \boldsymbol{X}^*_{2t-1}, \mu_{2t-1}) \cdot$$

$$Pr(\mu_{2t-1} | \boldsymbol{Z}^*_{2t-1}, \boldsymbol{X}^*_{2t-1}, \boldsymbol{P}_{2t-1}) \cdot \sum_{\omega_{2t-1}} Pr(\boldsymbol{P}_{2t-1} | \boldsymbol{Z}^*_{2t-1}, \omega_{2t-1}) \cdot Pr(\omega_{2t-1}) \cdot Pr(\boldsymbol{Z}^*_{2t-1}) \cdot$$

$$Pr(\boldsymbol{X}_{2t-1}, \boldsymbol{X}^*_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}, \mu_{2t-2,+1}], [\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}, \mu_{2t,-1}]).$$

If we divide the workers in market $2t-1$ into three parts: $I_{2t-1} := \{I_{2t-1,-1},\ I_{2t-1,0},\ I_{2t-1,+1}\}$, then l

$$\begin{aligned}
&Pr(\boldsymbol{X}_{2t-1}, \boldsymbol{X}^*_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}, \mu_{2t-2,+1}], [\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}, \mu_{2t,-1}]) = \\
&Pr(\boldsymbol{X}_{2t-1,-1}, \boldsymbol{X}^*_{2t-1,-1} | \boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}, \mu_{2t-2,+1}) \cdot \\
&Pr(\boldsymbol{X}_{2t-1,0}, \boldsymbol{X}^*_{2t-1,0}) \cdot \\
&Pr(\boldsymbol{X}_{2t-1,+1}, \boldsymbol{X}^*_{2t-1,+1} | \boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}, \mu_{2t,-1})
\end{aligned} \tag{2.61}$$

Remember the primitives we would like to estimate are $Pr(X_\tau | X^*_\tau, X_{\tau-1})$, $Pr(X^*_\tau, X^*_{\tau-1})$, and $Pr(X_{\tau-1} | X^*_{\tau-1})$. (The initial condition occurs when a worker first appears in the market. For simplicity I will denote the market as $\tau-1$ as they stay at most for two consecutive markets.) Therefore, we have

$$\begin{aligned}&Pr(\boldsymbol{X}_{2t-1,-1},\boldsymbol{X}^*_{2t-1,-1}|\boldsymbol{Z}_{2t-2,+1},\boldsymbol{X}_{2t-2,+1},\mu_{2t-2,+1})\\&=\prod_{i\in I_{2t-1,-1},\mu(i)=j}Pr(\boldsymbol{X}_{\mu(j),2t-1},\boldsymbol{X}^*_{\mu(j),2t-1}|\boldsymbol{Z}_{j,2t-2},\boldsymbol{X}_{\mu(j),2t-2},\mu_{2t-2,+1}),\end{aligned}\tag{2.62}$$

$$Pr(\boldsymbol{X}_{2t-1,0},\boldsymbol{X}^*_{2t-1,0})=\prod_{i\in I_{2t-1,0}}Pr(X_{i,2t-1}|X^*_{i,2t-1})\cdot Pr(X^*_{i,2t-1}),\tag{2.63}$$

and,

$$\begin{aligned}&Pr(\boldsymbol{X}_{2t-1,+1},\boldsymbol{X}^*_{2t-1,+1}|\boldsymbol{Z}_{2t,-1},\boldsymbol{X}_{2t,-1},\mu_{2t,-1})\\&=\prod_{i\in I_{2t-1,+1},\mu(i)=j}Pr(\boldsymbol{X}_{\mu(j),2t-1},\boldsymbol{X}^*_{\mu(j),2t-1}|\boldsymbol{Z}_{j,2t},\boldsymbol{X}_{\mu(j),2t},\mu_{2t,-1}).\end{aligned}\tag{2.64}$$

For equation (2.62), we let $\tau=2t-1$, so

$$Pr(\boldsymbol{X}_{\mu(j),\tau},\boldsymbol{X}^*_{\mu(j),\tau}|\boldsymbol{Z}_{j,\tau-1},\boldsymbol{X}_{\mu(j),\tau-1})=\frac{Pr(\boldsymbol{X}_{\mu(j),\tau},\boldsymbol{X}^*_{\mu(j),\tau},\boldsymbol{Z}_{j,\tau-1},\boldsymbol{X}_{\mu(j),\tau-1}|\mu_{\tau-1,+1})}{Pr(\boldsymbol{Z}_{j,\tau-1},\boldsymbol{X}_{\mu(j),\tau-1}|\mu_{\tau-1,+1})},\tag{2.65}$$

where

$$
\begin{aligned}
&Pr(\boldsymbol{X}_{\mu(j),\tau}, \boldsymbol{X}^*_{\mu(j),\tau}, \boldsymbol{Z}_{j,\tau-1}, \boldsymbol{X}_{\mu(j),\tau-1}|\mu_{2t-2,+1}) \\
&= \sum_{\boldsymbol{X}^*_{\mu(j),\tau-1}} f(\boldsymbol{Z}_{j,\tau-1}|\boldsymbol{X}^*_{\mu(j),\tau-1}, \mu_{\tau-1,+1}) \cdot \prod_{i\in\mu(j)} f(X_{i,\tau}|X^*_{i,\tau}, X_{i,\tau-1}) \cdot \\
&Pr(X^*_{i,\tau}|X^*_{i,\tau-1}) \cdot f(X_{i,\tau-1}|X^*_{i,\tau-1}) \cdot Pr(X^*_{i,\tau-1})
\end{aligned}
$$

and

$$
\begin{aligned}
&Pr(\boldsymbol{Z}_{j,\tau-1}, \boldsymbol{X}_{\mu(j),\tau-1}) \\
&= \sum_{\boldsymbol{X}^*_{\mu(j),\tau-1}} f(\boldsymbol{Z}_{j,\tau-1}|\boldsymbol{X}^*_{\mu(j),\tau-1}, \mu_{\tau-1,+1}) \cdot \prod_{i\in\mu(j)} f(X_{i,\tau-1}|X^*_{i,\tau-1}) \cdot Pr(X_{i,\tau-1}*)
\end{aligned}
$$

and the RHS corresponds to our primitives. Specifically, with respect to the first probability,

$$
f(\boldsymbol{Z}_{j,\tau-1}|\boldsymbol{X}^*_{\mu(j),\tau-1}, \mu_{\tau-1,+1}) = \sum_{Z^*_j} f(\boldsymbol{Z}_{j,\tau-1}|Z^*_j, \boldsymbol{X}^*_{\mu(j),\tau-1}, \mu_{\tau-1,+1}) Pr(Z^*_j).
$$

The RHS of equation (2.63) directly corresponds to our primitives. For equation (2.64), similarly we let $\tau = 2t$, so

$$
Pr(\boldsymbol{X}_{\mu(j),\tau-1}, \boldsymbol{X}^*_{\mu(j),\tau-1}|\boldsymbol{Z}_{j,\tau}, \boldsymbol{X}_{\mu(j),\tau}, \mu_\tau) = \frac{Pr(\boldsymbol{X}_{\mu(j),\tau-1}, \boldsymbol{X}^*_{\mu(j),\tau-1}, \boldsymbol{Z}_{j,\tau}, \boldsymbol{X}_{\mu(j),\tau}|\mu_\tau)}{Pr(\boldsymbol{Z}_{j,\tau}, \boldsymbol{X}_{\mu(j),\tau}|\mu_\tau)}, \tag{2.66}
$$

where

$$
\begin{aligned}
&Pr(\boldsymbol{X}_{\mu(j),\tau-1}, \boldsymbol{X}^*_{\mu(j),\tau-1}, \boldsymbol{Z}_{j,\tau}, \boldsymbol{X}_{\mu(j),\tau}|\mu_\tau) \\
&= \sum_{\boldsymbol{X}^*_{\mu(j),\tau}} f(\boldsymbol{Z}_{j,\tau}|\boldsymbol{X}^*_{\mu(j),\tau}, \mu_\tau) \cdot \prod_{i\in\mu(j)} f(X_{i,\tau}|X^*_{i,\tau}, X_{i,\tau-1}) \cdot \\
&Pr(X^*_{i,\tau}|X^*_{i,\tau-1}) \cdot f(X_{i,\tau-1}|X^*_{i,\tau-1}) \cdot Pr(X^*_{i,\tau-1})
\end{aligned}
$$

and

$$
\begin{aligned}
&Pr(\boldsymbol{Z}_{j,\tau}, \boldsymbol{X}_{\mu(j),\tau}|\mu_\tau) = \sum_{\boldsymbol{X}^*_{\mu(j),\tau}} f(\boldsymbol{Z}_{j,\tau}|\boldsymbol{X}^*_{\mu(j),\tau}, \mu_\tau) \cdot \\
&\prod_{i\in\mu(j)} \int_{X_{i,\tau-1}} f(X_{i,\tau}|X^*_{i,\tau}, X_{i,\tau-1}) \sum_{X^*_{i,\tau-1}} Pr(X^*_{i,\tau}|X^*_{i,\tau-1}) f(X_{i,\tau-1}|X^*_{i,\tau-1}) Pr(X^*_{\tau-1}) dX_{i,\tau-1}
\end{aligned}
$$

and the RHS corresponds to our primitives. Specifically, with respect to the first component,

$$
f(\boldsymbol{Z}_{j,\tau}|\boldsymbol{X}^*_{\mu(j),\tau}, \mu_\tau) = \sum_{Z^*_j} f(\boldsymbol{Z}_{j,\tau}|Z^*_j, \boldsymbol{X}^*_{\mu(j),\tau}, \mu_\tau) Pr(Z^*_j).
$$

Each element on the RHS is a primitive we would like to estimate.

Chapter 3

Matching Games with Unobserved Heterogeneity: A Structural Analysis of Online Labor Markets

3.1 Introduction

In recent decades, the emergence of an entirely novel online labor market has made it possible for previously geographically segmented workers and firms to match with each other at much lower cost. Workers are now hired by the task, therefore enjoying much more flexibility than staying in a long-term labor contract. Firms, on the other hand, enjoy the “crowdsourcing” benefits that enable them to attract multiple job candidates, each providing a unique proposal, and therefore increase the chance of finding a better solution. To date, most of the online labor markets operate in a decentralized fashion: Instead of having a centralized "clearinghouse" system to assign jobs to workers, firms and workers compete with their peers to win over the scarce resource on the opposite side. Probably the most significant difference of such online markets from traditional ones is that workers and jobs are much more idiosyncratic in both their characteristics and their needs (Horton, 2010). As a result, a well-defined price system is usually difficult to establish or maintain in such markets; instead, market participants interact strategically to form monetary transfers contingent with individual transactions. Meanwhile, anecdotal evidence suggests that the labor forces of such online markets work primarily part-time and more importantly, they care other features of a job beyond just cash earned. If this is true, then the incentive problem in such markets becomes a substantial topic for market creators. The multi-dimensionality of online workers' utility profiles, however, has rarely been captured in the literature, which might lead to misspecified model predictions. The current literature tends to assume away further the game-theoretical interaction between firms and workers, which would be inappropriate especially when market size is finite, possibly due to data restrictions and the lack of a suitable structural model.

To fill the gap, in this paper I use a two-sided matching framework to answer two empirical questions regarding such markets: first, how much does a worker in online labor markets care about various dimensions of a job? Second, based on the structural estimates, could the market designer introduce alternative schemes so as to improve two-sided participation and aggregated match-specific outcomes? In particular, I establish a stylized two-stage model to study a contest-based online labor market, in which firms first set their cash prize under asymmetric information, then workers choose which jobs to accept, taking into consideration peer workers' and firms' preferences. Following Chapter 2 on the identification of matching games with two-sided unobserved heterogeneity, I use data on (repeatedly) observed agent- and match-specific characteristics to nonparametrically estimate both sides' unobserved heterogeneity, assuming firms share a vertical preference over workers. The preference primitives are then identified from the equilibrium characterization. The data I use is an individual-level record of a primary contest category on the world's largest online coding community, providing coders around the world access to the software design/development and data-science problems published by mainly U.S.-based client firms. The structural estimates show the significant role of the unobserved firm- and worker-types in determining the observed sorting patterns in the market, as well as the ineffectiveness of the current prize scheme. Both findings shed light on the importance of introducing alternative plans to improve the assortativeness that benefits the platform and market participants in the long run.

From a market-design perspective, this paper takes the first step towards evaluating how the total surplus generated in the market could be improved by introducing alternative schemes. For instance, allowing firms to reward workers based on their ability and/or performance would also make them better off, despite the fact that they incur more cost from engaging in the market. This kind of findings is insightful to the market designer and could empirically boost the development of such markets, as conducting alternative schemes requires much less cost and is much more witnessed compared with traditional offline labor markets.

3.1.1 Related literature:

A growing empirical literature analyzes online marketplaces, a major theme of which is how to efficiently create trade between many buyers and sellers through an efficient price system. For instance, Einav et al. (2015) find that the auction-based pricing scheme is declining as online markets becomes mature and is shifting toward the posted-price scheme. This trend could be partially explained by the significant number of market participants, the time sensitivity of active trades, and intense competition among sellers (auctioneers) in the market. On the other hand, when service delivered in such markets is not standard, and market size is moderate, such as the one I am studying in this paper, it is usually difficult for a well established posted-price scheme to exist. Therefore, it is of great importance to examine how firms strategically determine the monetary transfers and how market outcomes are driven by the strategic behavior. In this paper, I analyze the pricing scheme by characterizing the decentralized wage-setting stage of the model and examine alternative schemes that could potentially improve the market surplus.

Another strand of literature tries to understand the incentives of participants, especially those of workers in such markets. Horton (2010) mentions in his survey paper of online labor markets that, given asymmetric information combined with strategic behavior and job-wise heterogeneity, potential incentive issues impeding workers from delivering satisfactory service can happen in such markets. In this paper, I focus on one particular aspect of workers' incentive issue by trying to answer the question: under what circumstances will they stay away from certain kinds of jobs, due to peer competition and their heterogeneous attitudes toward job characteristics, which determines their equilibrium choices when faced with capacity constraints.

This paper also relates to both the economic and informational system (IT) studies about online "crowdsourcing" platforms, a concept firstly introduced by Howe (2006). See for example the paper by DiPalantino and Vojnovic(2009) and Horton and Chilton (2010) among others. The former models the crowdsourcing markets as a two-stage game, in which the second stage very much resembles all-pay auctions; they found diminishing marginal returns of cash prizes regarding workers' participation level. The latter adopts a different rational model of labor supply and estimates the reservation wage for workers to be willing to join the platform. In this paper, I use a two-sided matching framework to analyze such platforms and focus on both sides' strategic movements leading to the observed market patterns.

Lastly, from a methodological point of view, this paper contributes to the recently growing empirical literature analyzing preferences in two-sided matching markets when unobserved heterogeneity prevails on both sides. See for example Choo and Siow (2006), Chiappori et al. (2006), Fox (2010), Graham (2013), Galichon and Salanie (2012) and Sinha (2014) when utility is modeled as transferrable and Logan et al. (2008), Boyd et al. (2013), Menzel (2015), Agarwal(2016) when utility is modeled as non-transferrable. This is, however, not the focus of this paper, and more detailed relation to this strand of literature is discussed in Chapter 2.

I start the discussion by presenting the market description and relevant institutional backgrounds in Section 2; I then establish the structural model and characterize the equilibrium in Section 3. Also, I briefly discuss how I achieve point identification using this strategy in Section 4.8. Next, in Section 4, I discuss the estimation strategy that nonparametrically recovers the unobserved heterogeneity and parametrically recovers the utility primitives. Following that, Section 5 presents the empirical findings and Section 6 presents the result of the counterfactual experiments. Lastly, I conclude this paper in Section 7.

3.2 Market Description and Data

The market I study in this paper is TopCoder.com, a leading crowdsourcing website. It offers businesses on-demand access to a worldwide community of over 800,000 designers, developers and data science experts.[1] The platform offers three main

[1] Although these participants come from various industries, from now on I uniformly call them coders, as most of the jobs involve primarily coding.

"tracks" for firms and workers to fit themselves in software design, development and data sciences. My data contains the complete participation record for the component development projects, a major subset under the development track, from September 2003 to November 2011, with $1,394$ active coders and $2,846$ projects in total. Each project is observed only once, but coders are observed multiple times throughout the time span. Multiple coders submit to a single project and are awarded according to their ordinal ranks. Specifically, the second place gets half of what the first place gets in cash; and after May 2006, the top five to seven participants get bonus point proportionally as an extra incentive to participate.

The main pipeline could be broadly divided into three stages. At Stage 1, client firms reach out to the website, and with the help of a project manager, decompose the project into smaller, independently manageable parts. They assign each component a set of reward package (including cash payment and accumulative point rewards which could transfer to cash seasonally) according to the rules of the website.[2] At the second stage, known as Competition and Collaboration Stage, coders select from a list of published projects and register themselves without any charge. Coders then work on the project simultaneously and submit by the due date. Communication with project managers is allowed and encouraged during the whole process. At the third stage, the Peer Review Stage, a group of *coding gurus* is hired by the client firms as the review board. The group first screens out any submissions failing to

[2]Specifically, regarding the cash payment, the website requires each client to split their cash rewards proportionally to the first- and second-place coders, on top of a fixed entry payment

meet the baseline requirement, as well as technically trivial solutions. Team members then independently score each submission passed the screening, and submissions are ranked based on the average of their final scores. Finally, at the Results and Rewards Stage, participants are awarded according to their ordinal ranks, and coders' public profiles are updated.

From the data, for every project, I observe the name (which briefly describes the project's content and requirement), the amount of the monetary and non-monetary award such as the "Digital Run Cup Series" introduced in 2006,[3] the duration of the project as measured by the number of days allowed to develop the codes, the programming language requirement and whether the project is for custom or generic use. Furthermore, I observe the number of registrations and submissions per project, the identity of all coders submitting and passing the screening, and the final scores for their submitted code. For every coder, I observe his nationality, active days on the website, and a full history including which projects he/she has participated in, final scores for every submission, monetary and non-monetary rewards collected, and updated skill rating on the website.[4]

Table 3.1 provides descriptive statistics of the project characteristics mentioned

[3] Digital Run (DR) points were introduced in mid-2006 as a complement to award good performers. They allocate proportionally among top five to seven participating coders at the end of each project. The points accumulate until the end of each quarter and those among the top third DR-point holders are rewarded with extra money. Although there have been various changes in the rules for setting the amount of the points, it is, on average, positively related with the winners' cash rewards.

[4] The site provides each registered programmer with a rating score for each track that he/she ever participates in. This rating score is calculated with the well-known Elo rating algorithm widely used to indicate the relative abilities of participants in competitor-versus-competitor games.

above. On average, each project is active for about one week.[5] The levels of the monetary reward provided by different projects vary considerably, reflecting the observed heterogeneity across projects. The dominant programming languages in the data are .NET and Java. It is further decomposed into two categories: generic and custom projects, with approximately 40% being generic. This is to capture whether the code written for the project will be used only once or will be potentially used for multiple purposes. Scores to measure the quality of the submissions are scaled from 0 to 100, averaging about 89, reflecting a satisfactory *ex post* performance of coders. Table 3.2 presents the coders' descriptive statistics. Repeated participation is common: The average number of previous submissions per coder at the time of registration, denoted Experience, is 13. This website is strongly international: over half of registered coders are from China, and only 9.37% are U.S. coders. Moreover, rating scores are relatively diverse: scaled from 0 to 2,500, I observe an average of 1,169 and a standard deviation of almost 400. Lastly, according to the website, contests occur on a roughly weekly basis. I, therefore, divide the whole dataset into 400 weekly markets. As shown in Table 3.3, I observe about seven contests active in a week, with about 22 coders participating. Repeated participation within a market is rare, though. In fact, about 88% of coders only attend once in a weekly market.[6]

By further exploring the data patterns, I have reason to believe there exists a

[5]After a closer look, I find that at the early and middle stages of my data time span, almost all contests are published on each Thursday, but this pattern becomes noisier in later stage.

[6]Initially, the website even restricted new coders to one project per week. This rule was abandoned in February 2008: `http://apps.topcoder.com/forums/?module=Thread&threadID=602646&start=0&mc=115`.

nontrivial degree of unobserved heterogeneity on both sides of this market, beyond observed characteristics. On the coder side, although we can observe the up-to-date rating scores for each person, it is still unclear how well these scores proxy innate skills. For instance, in Table 3.4, I observe that even after controlling for coders' rating score and experience, Chinese coders are preferred over U.S. coders, reflected by their ordinal placement within each contest. Rather than believing firms inherently prefer Chinese coders over U.S. coders, I think it may be true that coders' skill levels are not perfectly captured by rating scores and might be correlated with coders' nationalities.

On the firm side, I find much more extraordinary evidence: As shown in Figure 3.4, we see a clear increasing trend in participants' average rating scores as firms increase the prize money, but there is a decreasing trend in the average final scores as prize money goes up. A reasonable explanation would be the existence of project-wise heterogeneity beyond the observables, which coders may value differently. To take a closer look, I regress final scores coders get from participating in projects on observed project-wise and coder-wise characteristics, and the results are shown in Table 3.5. First, the negative correlation between the cash payment and coders' performance does not vanish after I control for observed characteristics. Second, in the last two column, I consider the potential endogeneity issue concerning projects' winning cash payment and coders' rating scores – the former can be correlated with unobserved project heterogeneity and the latter might be associated with the unobserved coder heterogeneity. To tackle this issue, I use the average winning cash payment within the weekly market and the last-period rating score for the same coder to instrument the two endogenous variables. It can be seen that the coefficient before winning cash payment becomes positive in the last column, where I further incorporate market fixed effect to deal with market-wise unobserved heterogeneity. This serves as stronger evidence showing that after teasing out the potential project-wise and market-wise heterogeneity, higher cash payment indeed motivates better performance measured by the final scores.

The matching allocation is driven by mutual choices of firms and coders. This platform's crowdsourcing feature enables us to observe firms' cardinal preference over coders who submit to the project. The more interesting part, however, is how firms rank all potential participants in the market, which is unknown to the researchers. Nonetheless, from a simple regression analysis, I find that firms' revealed preference over actual submissions indicates a strong vertical pattern. In Table 3.4, I regress the within-contest ranking against coders' rating-score rankings along with other covariates. The coefficient of rating score rank is as high as 0.9 with an R-squared to be as high as 0.81, and barely changes when we add more covariates. Furthermore, as this rating score is uniformly computed and ranked, one has reason to believe that all firms hold a vertical preference over coders, according to their (potentially imperfectly observed) skill levels.

On the other hand, inference on coders' preference over firms is less straightforward because firms can rank multiple workers who submit, but workers can only choose one job at a time. More specifically, for any single coder in a market, we only observe his discrete choice of participation. Nonetheless, it is difficult to use the standard discrete-choice regression model to study coders' preference over the entire project-wise characteristics set, as all the choices (projects) are heterogeneous across different markets. In Table 3.6, I regress coders' rating scores against contests' observable characteristics and see how coders sort themselves into different kinds of projects regarding their observable characteristics. In particular, I divide coders into two subgroups: "top coders" who rank first or second in a contest, and "average coders" who ranked third or worse in a contest. Their regression results are shown in columns 2 and 3, respectively. A high heterogeneity in the sorting pattern is detected here: Top coders have a strong disutility over longer contests, whereas average coders slightly prefer more extended contests, as shown in row 2. The preference for winning cash prizes is also different, as average coders apparently sort away from highly-paid jobs, probably due to their concern with the job complexity. To sum, it is inappropriate to hypothesize coders' preferences to be purely vertical; instead, coders' clearly sort themselves into projects heterogeneously in the data. More interestingly, if for instance short projects with high payment are more likely to be highly complex jobs, we can even detect coders' different preference over the unobserved types(i.e. complexity levels) across jobs.

Lastly, I look at how cash payment is determined in the market, as in labor markets, monetary transfers play a crucial role in determining the final matching allocation. In the empirical literature analyzing matching markets, two completely different ways of modeling the monetary transfers diverge in their methodologies to identify primitives. In reality, however, monetary transfers are very likely to be endogenous, but not arbitrarily negotiable between firms and coders. For instance, in this market, firms first determine the contract terms, publish the information, and then coders participate to win the contract. Setting and negotiating contracts/wages, therefore, is excluded from the matching process; however, firms non-cooperatively set the wage, taking into consideration how the matching game will play afterward. Table 3.7 suggests strong evidence for prize endogeneity: By regressing the winning prize on the degree of competition along with other firm-level covariates, I find a strong causal effect of the highest winning prize of peer contests in the market on cash prizes. When controlling for market-wise fixed effect as shown in the second column, firms are negatively affected by the head-to-head competition. This is very interesting, as firms explicitly take into consideration both the market-specific willingness-to-pay and the competition within a market when setting their cash payment to incentivize coders to participate later.

To summarize, from a series of reduced-form analyses, I have found that: (1) there might exists a non-trivial degree of coder-wise, project-wise and market-wise unobserved heterogeneity in this online market; (2) firms are likely to admit a vertical preference over coders' skill levels, but coders might have heterogeneous preference over the same job; and (3) firms strategically set the cash payment for their projects, reflecting their willingness-to-pay for the talents in the competitive market. Next, I will present a structural model and use a novel method to estimate this game.

3.3 A Structural Model for the Online Labor Matching Markets

I present a stylized structural model that captures how the market works. As the researcher, from market t, I observe three sets of information: the project-wise observable characteristics including the cash prizes set by all J_t firms, $\boldsymbol{P}_t$; the coder-wise observable characteristics including their current rating score, $\boldsymbol{X}_t$; and the matching allocation μ_t that tells us who matches with whom along with a one-dimensional match-specific outcome, $\boldsymbol{Z}_t$. On the other hand, the project's complexity, $\boldsymbol{Z}_t^*$ and coders' abilities, $\boldsymbol{X}_t^*$, and the market condition ω_t, are unobservable heterogeneity, all of which are assumed to be drawn from finite spaces.

I abstract away several aspects of this market. First, to avoid the complexity of coordinating intra-firm pricing strategies, and to focus on the inter-firm pricing competition, I ignore the fact that a firm's business project is divided into multiple components. Instead, I simply assume each firm (or equivalent entity) carries a single project. This is acceptable as in the data; there is no record tracking the identities of firms behind each project; it might, therefore, be inaccurate to decide which projects belong to the same firm. Second, I assume away strategic entry and exit decisions from both sides. I, therefore, do not consider coders' strategic participation and potential learning-by-doing in this model. This is a major simplification, but is consistent with the empirical findings that there is no explicit pattern regarding entries and exits over time, as shown in Figure (3.3). This simplification enables me to focus on the strategic pricing behavior of firms and discrete choices of coders in the market, which I believe are most important to recover for policy implications. Lastly, when coders are making their discrete-choice decisions, I assume they enjoy full information including their peers' utility profiles and idiosyncratic shocks. While it is rare to observe fully informed workers in a real-life labor market, it is much easier and almost costless to achieve full-information scenario for the worker side in online markets. Here I focus on the information asymmetry from the firm side towards the coder side. Next, I establish the timeline for the two-stage market game.

At Stage 1, a finite number J_t of firms are randomly drawn to be present in the market. Firms have common prior about the distribution of peer projects' complexity and potential coders' utility profiles but cannot observe other firms' types nor the actual identity of participating coders. Also, all firms observe the market condition, ω_t, that commonly affects their profits.[7] By the end of this stage, firms simultaneously decide the cash prizes they will pay for the first and second place among all submissions. At Stage 2, a finite number I_t of coders are randomly drawn to enter

[7] An example that determines ω_t is a positive shock to recruiting coders from traditional markets.

the market. Perfectly observing the characteristics and utility profiles of every active project and every peer coder, they simultaneously submit their codes to the project they choose. Without loss of generality, assume all firms provide a slot quota such that $J_t \times Q_t = I_t$. Firms then hire a third-party reviewer board to evaluate all submitted codes. The market game completes with firms publishing the scores for every submission, coders getting cash rewards accordingly.

Next, I define the preference structure of firms and coders. There is an abuse of notation in the subsequent discussion: I_t (resp.J_t) denotes both the number of coders (resp. projects) in market t, and the set of coders (resp. projects) in market t.

3.3.1 Preferences Structure of Firms and Coders

As presented in the reduced-form analysis, from Table 3.4, I detect a strong vertical pattern of firms' preferences over coders. This implies that coders' professional skills might be the primary component affecting firms' preferences. Nonetheless, I need to consider the fact that the willingness to pay for professional skills might vary across firms, especially when they are faced with competitions among other firms of asymmetric types. Therefore, I model Firm j's *ex post* payoff from collecting submissions of the set of coders $\mu_t(j)$ in market t is modeled as:

$$\pi^{post}_{j,\mu_t(j),t} = \mathcal{R}(c, \omega_t, Z^*_j, \boldsymbol{W}_j, \sum_{i \in \mu_t(j)} Z_{ij}) - \sum_{i \in \mu_t(j)} P_{ij} \tag{3.1}$$

where $\mathcal{R}$ is the revenue function that has a known form to the researcher. c is a one-dimensional revenue coefficient that is positive and reflects how profitable the submitted codes are. $\boldsymbol{W}_j$ is a vector of observable firm/project-wise characteristics; here in my data, it is the duration and technology of the project. Lastly,

$$P_{ij} = \begin{cases} P_j, & \text{if } i \text{ is the first place} \\ \frac{1}{2}P_j, & \text{if } i \text{ is the second place} \\ 0, & \text{otherwise.} \end{cases}$$

Note that this is one of the simplest parametric specifications of the payoff function for the ease of illustration, and I can also add nonlinear terms in empirical analyses. In particular, the revenue part mainly depends on three elements: (i) the project-wise complexity Z_j^*; (ii) the sum of Z_{ij}, which denotes the quality of coder i's submission to project j, and (iii) the demand-side shifter that affects the overall willingness to pay in the market, and is denoted by ω_t.

Next, I model the quality of a submission to be determined by the following equation:

$$Z_{ij} = g_2(X_i^*, Z_j^*, \boldsymbol{W}_j, \boldsymbol{V}_i, b_j), \tag{3.2}$$

where g_2-function is increasing in coder i's ability, X_i^*, and decreasing in the complexity of a project, Z_j^*. The final score is also determined by project-wise observed characteristics, $\boldsymbol{W}_j$: the duration and the technology, as well as the coder's experiences, as denoted by $\boldsymbol{V}_i$. The idiosyncratic term b_j is independent of (X_i^*, Z_j^*) and is realized only after the code is submitted. [8]

To sum, this specification of firms' preference captures (i) the vertical preference over coders' skill ability; (ii) the synergy effect between a project's complexity and participating coders' skill levels, and (iii) the trade-off between attracting better coders and saving costs.

Next, I model the preference structure of coders towards projects. Following the discrete choice literature, I assume the indirect utility of coder i choosing project j in market t is determined by the following equations:

$$u_{ijt} = \beta_0 P_{ij} + \tilde{u}_{it}(X_i^*, Z_j^*, \boldsymbol{V}_i) + \boldsymbol{\beta W}_j + \nu_{ijt} \tag{3.3}$$

where β_0 captures coders' vertical preference over cash prizes, and $\beta_0 \geq 0$. The function $\tilde{u}_{it}$ is coders' heterogeneous preference over the complexity of a project, and reflects their benefit from projects with various complexity which may depend on their own skills and experiences. Next, as before, $\boldsymbol{W}_j$ denotes projects' duration and technology requirement, which coders also care about; lastly, ν_{ijt} is the idiosyncratic taste shock that is independent of other variables. In this paper, I assume $\tilde{u}_{it}$ takes

[8] This is to exclude the case where firms know *ex ante* the value of b_j when they decide the cash prizes at Stage 1.

the following form:

$$\tilde{u}_{it}(X_i^*, Z_j^*) = \gamma_0 X_i^* + (\gamma_1 X_i^* + \gamma_2 \boldsymbol{V}_i + \eta_{it}) \cdot Z_j^* \tag{3.4}$$

where η_{it} is a random coefficient following a normal distribution and is independent of X_i^*. Lastly, coders are allowed to choose the outside option. The systematic part of the indirect utility from choosing any outside option is assumed to be zero.

Now I define the matching allocation μ generated in the second stage of the game: if coder i chooses project j, then $\mu(i) = j$ and $i \in \mu(j)$. If coder i remains unmatched, then $\mu(i) = \emptyset$; if no one chooses project j, then $\mu(j) = \emptyset$. This enables me to further define the rational-expectation equilibrium for this extensive-form game.

Definition 4 *The rational-expectation equilibrium (δ_t^*, μ_t^*) is such that: At Stage 2, for any observed $(\boldsymbol{P}_t, \boldsymbol{Z}_t^*)$, the matching allocation $\mu_t^* : I \cup J \to I \cup J \cup \emptyset$ is pairwise stable; At Stage 1, given the rational expectation about the stable matching function and the knowledge of distribution of $(\boldsymbol{X}_t^*, \boldsymbol{Z}_t^*)$, firms play the mixed-strategy pricing function $\delta_t^* : \mathbb{Z} \to \Sigma$ which is Bayesian Nash Equilibrium strategy.*

The equilibrium concept defined in the second-stage market is the *pairwise stability*. An observed μ is said to be *pairwise stable* if it satisfies:

1. (Individual Rationality) $u_{i,\mu(i),t} \geq u_{i,\emptyset,t}, \forall i \in I_t$ and $|\mu_t(j)| \leq Q_t$, $\pi^{post}_{j,\mu_t(j),t} \geq \pi^{post}_{j,\mu_t(j)\backslash i,t}$, $\forall j \in J_t, i \in \mu_t(j)$.

2. (Nonblocking Pairs) For any coder i and project j such that $j \neq \mu_t(i)$, the following situations cannot happen simultaneously: $u_{i,j,t} > u_{i,\mu(i),t}$; and $\pi^{post}_{j,\mu_t(j)\cup i,t} > \pi^{post}_{j,\mu_t(j),t}$ if $|\mu_t(j)| < Q_t$; or $\pi^{post}_{j,\mu_t(j)\setminus i' \cup i,t} > \pi^{post}_{j,\mu_t(j),t}, \exists i' \in \mu_t(j)$ if $|\mu_t(j)| = Q_t$.

The first condition implies that the matching allocation I observe is at least as desirable for all firms and coders as staying unmatched. The second condition implies that, for any coder i in the market, his/her current choice $\mu_t(i)$ is the most desirable project in his/her choice set. This choice set consists of any projects that is willing to swap their current matched coders with coder i, or to fulfill a vacant space with coder i.

At Stage 1, conditional on the prior knowledge of the joint distribution of potential coders' abilities and peer firms' project complexity, along with the distribution of coders' idiosyncratic taste shock and the idiosyncratic shock to the code quality, the mixed-strategy Bayesian Nash Equilibrium is defined as a mapping $\delta^*_t : \mathbb{Z} \rightarrow \Sigma$, where $\Sigma := \{\delta | \sum_{l=1}^m \delta^l = 1\}$, such that for each firm j, given other firms' equilibrium strategy, $\delta^*_{-j,t}$, and the correct belief for the second stage, the following inequality holds:

$$\sum_{l=1}^{m}[\sum_{\boldsymbol{Z}^*_{-j},\boldsymbol{X}^*} Pr(\boldsymbol{Z}^*_{-j}, \boldsymbol{X}^*)\pi^{\text{int}}(p_l, Z^*_j, \boldsymbol{Z}^*_{-j}, \boldsymbol{X}^*, \boldsymbol{\delta}^*_{-j,t}(\cdot), \omega)] \cdot \delta^{l,*}_{j,t} \geq$$
$$\sum_{l=1}^{m} \sum_{\boldsymbol{Z}^*_{-j},\boldsymbol{X}^*} Pr(\boldsymbol{Z}^*_{-j}, \boldsymbol{X}^*)\pi^{\text{int}}(p_l, Z^*_j, \boldsymbol{Z}^*_{-j}, \boldsymbol{X}^*, \boldsymbol{\delta}^*_{-j,t}(\cdot), \omega)] \cdot \delta^{l}_{j,t}, \forall \delta \in \Sigma. \tag{3.5}$$

where π^{int} denotes the interim payoff for firms j that chooses cash prize P_j and believing other firms will play the mixed strategy $\delta^*_{-j,t}$:

$$\pi^{\text{int}}(p_l, Z_j^*, \boldsymbol{Z}^*_{-j}, \boldsymbol{X}^*, \boldsymbol{\delta}^*_{-j,t}(\cdot), \omega) =$$
$$\sum_{\mu_t} Pr(\mu_t | \boldsymbol{Z}^*_{-j,t}, \boldsymbol{X}^*_t, Z^*_j, p_l, \delta_t(\boldsymbol{Z}^*_{-j,t}))\pi^{post}(Z^*_j, \omega_t, \boldsymbol{X}^*_t, \mu_t, p_l) \tag{3.6}$$

3.3.2 Equilibrium Characterization

In this section, I use the result from Section 2.4 of Chapter 2 to establish the existence result of such equilibrium. Specifically, the following lemma hold and proof can be found in the appendix of Chapter 2.

Proposition 4 *There exists a rational-expectation equilibrium (μ_t^*, δ_t^*) for the game defined in the previous section. Moreover, when assuming the symmetric equilibrium mixed-strategy δ^* is uniquely played in the data, the following two results hold:*

1. *The match μ_t is pairwise stable if and only if $u_{i,\mu(i)} \geq \max_{m \in M_i[\mu_t]} u_{i,m}, \forall i \in I_t$, where $M_i[\mu_t]$ is defined as follows:*

$$M_i[\mu_t] := \{j^1 | j \neq \mu(l), \forall X_l^* > X_i^*, \pi^{post}_{j,i,t} \geq 0\} \cup \{j^k | \exists l_m, m = 1, 2, \cdots, k-1,$$
$$\textit{such that } \mu(l_m) = j, X^*_{l_m} > X^*_i, k \leq Q_t, \pi^{post}_{j,l_m \cup i,t} \geq \pi^{post}_{j,l_m,t}\} \cup \emptyset \tag{3.7}$$

2. *In markets where all firms fill up their vacancy, the probability of observing a firm setting the cash prize $p_m = \max \mathbb{P}$ monotonically increases with the market-*

level unobservable ω and the profit coefficient c.

This proposition shows that, at the second stage, as long as we know the skill levels of all coders in market t, the pairwise stable outcome degenerates to a discrete choice problem on the coder side, in which coders make their discrete choices sequentially, even though this is originally a two-sided market. Moving back to the first stage, in equilibrium, the better the market demand condition is (e.g. when it is more costly to recruit coders from alternative platforms), the more probable for me to observe firms posting maximum cash reward to compete for the best coders in the market. This applies also to the profit coefficient that is remained to be identified – the more profitable it is to attract better coders to do the job, the more probable it is for firms to offer the maximum cash reward.

Detailed proof of this proposition can be found in the second section of Chapter 2. In the next section, I will move on to the discussion of how to estimate this game. A stylized identification discussion follows afterward, with minor changes made about the original identification strategy in the third section of Chapter 2.

3.4 Empirical Specification and Estimation Procedure

Suppose, as the econometrician, I observe T weekly markets with I_t coders and J_t projects in market t. Without loss of generality, assume each of the J_t projects offers Q_t slots such that $J_t \times Q_t = I_t$. Once a project is randomly drawn in market t, it leaves the population and will not be selected in future markets. On the other hand, coders enjoy a positive (and exogenous) probability of re-entering the market.

Every coder carries a latent ability level, X_t^*, that is discrete and evolves over time according to some underlying law, $Pr(X_t^*|X_{t'}^*)$ if the coder appears in markets t and t' consecutively. Each project carries a difficulty level Z^*, also discrete. Additionally, each weekly market t carries unobserved heterogeneity, ω_t, which affects firms' willingness to pay. All Q_t slots within a project j share the same characteristics, $\boldsymbol{W}_j$ and the unobserved type Z_j^*, but grant different cash prizes $[P_j, \frac{1}{2}P_j, 0, \cdots, 0]$, respectively.[9]

To summarize, for each market t, I observe a $I_t \times 1$ vector of coder characteristics including their rating scores, $\boldsymbol{X}_t$, and their experiences and nationalities, $\boldsymbol{V}_t$; a $J_t \times 1$ vector of contest characteristics, W_t and cash prizes, P_t; and a matching allocation μ_t such that if coder i chose slot q in contest j, then $\mu(i) = k_j^q$ and $\mu(k_j^q) = i$. The goal of identification, on the other hand, consists of three categories: (1) The joint

[9] In the data there exists an extra percentage bonus for coders who have been reliable in their history; this is abstracted away here but will be taken into account in estimation.

distribution of coders' and projects' unobserved heterogeneity, $(\boldsymbol{X}_t^*, \boldsymbol{Z}_t^*)$; (2) coders' utility primitives in $u_{i,j,t}$; and (3) firms' *ex post* profit primitives in $\pi_{j,i,t}^{\text{post}}$, i.e. the one-dimensional profit coefficient c.

In the following discussion, I suppress the market subscript t for ease of notation. Also, I use bold letters to denote market-wise vectors of characteristics, whereas regular letters denote individual-level characteristics. The letter f is used to indicate the probability mass/density function of any distribution, with a little abuse of notation, and M is used to denote the matrix representation for any discrete distribution.

3.4.1 The Coder Rating System

To keep track of coders' ability ranking in the population, the website provides an up-to-date rating record for every coder. The rating score is calculated using the Elo Algorithm as in chess games and is updated every time a coder participates in a contest.[10]

Under this algorithm, I assume the rating score is calculated (and updated) according to the following function. Specifically, I assume that a coder's skill level when entering the market follows a Bernoulli distribution with parameter $p_X \in [0, 1]$. In addition, I assume the rating scores and evolution of underlying abilities of coders

[10]The particular algorithm for calculating (and updating) a coder's rating score can be found at https://community.topcoder.com/tc?module=Static&d1=help&d2=ratings

are determined by two rules:

$$X_{\tau-1} = X^*_{\tau-1} + u_1 \tag{3.8}$$

when a coder first enters the market $\tau - 1$, and

$$X_\tau = \lambda \cdot X^*_\tau + (1-\lambda) \cdot X_{\tau-1} + u_2. \tag{3.9}$$

when he was present in some previous markets and re-enters the market τ. Rating scores remain unchanged until the next time a coder re-enters the market. The error terms (u_1, u_2) follow a zero-mean joint normal distribution with variance–covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$.

Lastly, a coder's latent ability evolves according to the following rule:

$$Pr(X^*_\tau = x_H | X^*_{\tau-1}, \boldsymbol{V}_i) = [\delta_{x1}(\boldsymbol{V}_i), \delta_{x2}(\boldsymbol{V}_i)]', \tag{3.10}$$

where $\boldsymbol{V}$ is observable characteristics and $X^*_{\tau-1}$ is the coder's most recent latent ability. The goal of estimation, therefore, is to find $(\sigma_1, \sigma_2, \lambda, \delta_{x1}(\boldsymbol{V}_i), \delta_{x2}(\boldsymbol{V}_i)$.

This specification above indicates that a coder's rating in market t is determined by his rating in the most recent market, $\tilde{X}_{it}$, his innate ability in market t, X^*_{it}, and a match-specific shock u independent of X^*_i, independent among different coders and independent across different weekly markets. In reality, this random shock reflects additional information used to calculate and update the rating scores, such as the rating volatility of other peer coders in the same contest. Furthermore, the following conditions hold for any pair of coders i and m such that $\mu(i) \neq \mu(m)$ in market t:

$$X_{kt} \perp X_{it} | \boldsymbol{X}_t^*, \boldsymbol{X}_{t-1}, \mu_t.$$

Next, I specify how the reviewer board evaluates all submitted codes and generate the final score that reflects firms' preferences and determines the rank order within each project.

3.4.2 The Contest Evaluation System

According to the rule specified on the website, the evaluation scores provided by the reviewer board determines the ordinal rank of participating coders. Moreover, both the website and users agree that the score is positively related to each coder's latent ability and the complexity of a project. Specifically, the more difficult a contest is, the more probable that submitted codes cannot meet the universal grading criterion. Thus, the submitted code is more likely to receive a lower grade. I make the following parametric specification about how the final score is related to observed and unobserved characteristics.

Let $Z_{i,j}^t$ represent the final score coder i gets from participating in project j, characterized by

$$
\begin{aligned}
Z_{ij} &= g_2(X_i^*, Z_j^*, \boldsymbol{W}_j, \boldsymbol{V}_i, b_j) \\
&= \xi_0 + \xi_1 X_i^* + \xi_2 Z_j^* + \xi_3 . NET_j + \xi_4 Duration_j + \xi_5 Experience_i + b_j \qquad (3.11)
\end{aligned}
$$

where,

1. the idiosyncratic part b_j reflects anything not captured in the characteristic space, such as the review board's potential bias in grading, and is independent of $(X_i^*, Z_j^*, \boldsymbol{W}_j, \boldsymbol{V}_i)$. In particular, it is assumed to follow a zero-mean normal distribution with variance σ_b^2.

2. g_2 is continuous and monotonically increasing with respect to its first argument ($\xi_1 \geq 0$) and is monotonically decreasing in its second argument ($\xi_2 \leq 0$).

3. Conditional on observed matches μ_t,

   (a) For any two contests j and l in the market, $b_j \perp b_l$.

   (b) For any coder i that is matched with j, $b_j \perp (X_i^*, u_{i,\mu_t})$.

   (c) For any project j in market t, $b_j \perp (Z_j^*, \omega)$.

   (d) For any two markets t and t', $\boldsymbol{b}_t \perp \boldsymbol{b}_{t'}$.

Part (i) above implies that the variation in such scores is driven by the unobserved heterogeneity on both sides. Part (ii) suggests that idiosyncratic taste shocks are unrelated with the market willingness to pay, ω. This is reasonable, as the reviewer board evaluates codes independent of firms' decision-makers and determines scores solely based on functionality and documentation. The parameters to be estimated are therefore $(\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \sigma_b)$.

3.4.3 The Equilibrium Cash Prize

As described in the modeling section, the equilibrium decision of the optimal cash prize is set by firms before the matching process as an equilibrium outcome. This is also supported by reduced-form evidence: The level of the winner's cash prize is significantly affected by the level of winner's cash prize in peer projects in the same market. Therefore, I would like to incorporate the prize-setting stage into the estimation procedure.

I conduct a two-step estimation: In the first step, I estimate $\widehat{\delta} := Pr(P_{jt}|Z^*_{jt}, \omega_t)$. In the second step, I match the simulated cash prize distribution with the estimated distribution from Step 1, and use the minimum-distance estimator to estimate the firm's profit parameter c.

In the data, I observe cash prizes taking integer values and are calculated in US dollars. To make the estimation simpler, in the first step, I discretize the observed cash prize into three intervals: above the median (\$500), equal to the median and below median. Furthermore, the probability of the cash prize lying in each interval is assumed to follow a logit model:[11]

$$Pr(P_{jt} = p^k | Z_j^*, \omega_t) = \frac{\exp[\psi_H + \psi_0^{Z^*} \cdot Z_j^* + \psi_0^{\omega} \cdot \omega_t + \psi_1.\text{NET} + \psi_2\text{Duration} + \psi_3 J]}{\sum_{k=M,H} \exp[\psi_k + \psi_0^{Z^*} \cdot Z_j^* + \psi_0^{\omega} \cdot \omega_t + \psi_1.\text{NET} + \psi_2\text{Duration} + \psi_3 J] + 1}, k = M, H \tag{3.12}$$

The parameters to be estimated therefore are $(\psi_H, \psi_0^{Z^*}, \psi_0^{\omega}, \psi_1, \psi_2, \psi_3)$. All the characteristics in the above equation are observable except for (i) the contest-wise unobserved heterogeneity, i.e. the project complexity level, and (ii) the market-wise unobserved heterogeneity, ω_t. For the former, I assume it is drawn from a binary space, $\{z_L, z_H\}$, and how it relates to observables also follows a logit model:

$$Pr(Z_j^* = z_H) = \frac{\exp[\iota_0(1 + \iota_1.\text{NET} + \iota_2\text{Duration})]}{\exp[\iota_0(1 + \iota_1.\text{NET} + \iota_2\text{Duration})] + 1} \tag{3.13}$$

and the parameters to be estimated are $(\iota_0,\ \iota_1,\ \iota_2)$. The market-level unobservable ω is also drawn from a binary space $\{w_L, w_H\}$, with probability of being w_H depends on the number of projects in the market according to a probit model:

$$Pr(\omega_t = w_H) = Pr(\chi_0 + \chi_1 J + v_t \geq 0) \tag{3.14}$$

where v_t follows a standard normal distribution. The parameters to be estimated

[11] In the actual estimation, I only include the .NET dummy but not the Java dummy, as they compose 98% of the observations, and in order to avoid multicollinearity, I treat all projects that are not .NET as belonging to the Java family.

are (χ_0, χ_1).

3.4.4 Project Generality

As mentioned in the data description section, one feature about software development contests is that, besides the distinction of programming languages (such as Java and C++), applications are also categorized into two classes: custom and generic. The submitted code for generic applications will be delivered to the client firm and simultaneously included in the TopCoder Catalog for potential future use. In contrast, custom code is entirely tailor-made for the client firm's business project: It is not allowed to be used by others.

Anecdotal evidence has suggested that there exists complexity distinction between custom and generic projects. For instance, in a custom project, client firms often expect coders to configure multiple sets of properties to meet their specific functionality requirement. Figure 3.1 shows that custom projects are on average rewarded higher than generic projects, but the number of participation is lower, demonstrated by Figure 3.2. This serves as indirect evidence for the relationship between project generality and its underlying complexity (types). More intuitively, the types of jobs in this market involve certain degree of innovation, and it is reasonable to believe that the customized jobs involve higher degree of innovation than generic jobs, which also inherently adds complexity to the job.

Specifically, let Y_j denotes the binary indicator, equal to one if the contest is generic and zero if it is custom. Then, I make the following specification:

$$Pr(Y_{jt} = 1 | Z_j^*) = \frac{\exp[\phi_0 + \phi_1 \cdot Z_j^* + \phi_2.\text{NET} + \phi_3 \text{Duration}]}{\exp[\phi_0 + \phi_1 \cdot Z_j^* + \phi_2.\text{NET} + \phi_3 \text{Duration}] + 1} \tag{3.15}$$

and the parameters to be estimated therefore are $(\phi_0, \phi_1, \phi_2, \phi_3)$. Furthermore, assume $\log(Pr(Z_j^* = z|Y = 1)) - \log(Pr(Z_j^* = z|Y = 0))$ is decreasing in z and is i.i.d. among all contests with the same Z_j^*. In addition, given Z^*, the distribution of Y is independent of both firms' profit shocks and reviewers' taste shocks . Furthermore, Y_j is excluded from coders' utility functions.

Intuitively, as we compare an easy project with a difficult one, it is relatively more probable to observe the complicated project when $Y = 0$ rather than when $Y = 1$. This assumption mainly restricts the possible correlation between the generality of a project and how reviewers would bias their evaluation conditional on the complexity. This is reasonable, as final scores mainly evaluate the functionality and documentation of codes. Nevertheless, I exclude the situation where reviewers are systematically more favorable to custom (or generic) code.

More implicitly, I restrict coders from having a systematic preference for custom (or generic) contests, beyond the difficulty levels they carry. From the forum discussion, custom and generic contests do not differ much in popularity. If a generic code is sold to other clients, the winning coder will gain a "Royalty Dividend" – i.e., a small portion of the profit; for custom code, however, there will be no potential dividends in the future. Nonetheless, royalty payments are tiny compared with winning prizes. Recently, the website even decided to shut down the Royalty Dividend program due to lack of popularity and motivation. This is a strong exogeneity assumption, as I assume coders do not take the potential Royalty Dividend into account when they make decisions, but we have anecdotal evidence supporting it. In a word, the binary variable of custom vs. generic contests only reflects a variation in difficulty levels, but does not affect how reviewers or coders value a contest, beyond its latent complexity, Z^*.

3.4.5 Coders' Indirect Utility Primitives

Next, I parametrize coders' indirect utilities as:

$$\begin{aligned} u_{i,j^k} = \beta_1 P_{j^k} + \beta_2 \text{Duration}_j + \beta_3 .\text{NET}_j + \beta_4 X_i^* + (\gamma_0 + \gamma_1 X_i^* \\ + \gamma_2 \text{Experience}_i) \cdot Z_j^* + \nu_{ij}, \end{aligned} \tag{3.16}$$

where ν_{ij} is the coder/contest pairwise idiosyncratic taste shock, following an i.i.d. standard Type-I Extreme Value distribution. Specifically, the coefficient β_4 is the fixed-effect for a given ability level X_i^*; γ_0 is the fixed effect given a complexity level of a project; γ_1 measures each coder's preference over the interaction of her ability and the contest's difficulty level, and γ_2 captures the interaction between a coder's experience and the project complexity; η_i is an unobserved taste determinant that is independent of the other covariates. The parameters to be estimated are (β_1, β_2, β_3, β_4, γ_0, γ_1, γ_2).

3.4.6 Firm's Profit Function

As presented in the modeling part, the *ex post* profit for each firm to attract a set of coders is specified as:

$$\pi^{post}_{j,\mu_t(j),t} = \mathcal{R}(c, \omega_t, Z^*_j, \boldsymbol{W}_j, \sum_{i \in \mu_t(j)} Z_{ij}) - \sum_{i \in \mu_t(j)} P_{ij} \tag{3.17}$$

There are two elements that are unknown to the econometrician. First, the market-wise unobserved heterogeneity ω_t has an unknown distribution; second, the one-dimensional profit coefficient c is not known either. To estimate these two, I first exploit the joint distribution of monetary prizes of at least three firms in a market, as they can be viewed as the three measurements to the underlying market demand shifter ω_t and are conditionally independent with each other. After estimating the marginal distribution of the market-wise unobserved heterogeneity, ω_t, I can then estimate the one-dimensional profit coefficient, c, for all firms using the following equation:

$$c = (\delta^*)^{-1}(Pr(P|Z^*, \omega), Z^*, Pr(\omega), Pr(Z^*), Pr(X^*)).$$

where $Pr(P|Z^*, \omega)$ is already estimated, and δ^* is the symmetric equilibrium strategy of each firm, and is assumed to be invertible on c. In the actual estimation, however, due to the complexity of computing δ^*, I did not estimate the profit coefficient, but in the Monte Carlo Simulation section of Chapter 2, I show the performance of my estimation for c.

After presenting the parametric specification of the primitives in the game, I present my likelihood-based estimation procedure in the following section.

3.4.7 A Likelihood-Based Estimation Procedure

To summarize, the primitives I would like to estimate consist of three sets: (1) $(\boldsymbol{\beta}, \boldsymbol{\gamma})$, which are parameters in coders' indirect utility function; (2) $(\boldsymbol{\xi}, \sigma_b)$, which are parameters that determines the match-specific outcome, or the final score for each coder, in equation (3.11); and (3) $(\boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\iota}, \boldsymbol{\xi})$, the parameters that determines the distribution of project complexity, $\boldsymbol{Y}$, equilibrium cash prize $\boldsymbol{P}$, the marginal distribution of project- and market-wise unobserved heterogeneity, $(\boldsymbol{Z}^*, \omega)$, and (4) $(\sigma_1, \sigma_2, \lambda, \delta_{x1}, \delta_{x2})$, the parameters that determines the marginal distribution $Pr(\boldsymbol{X}^*)$. Summarizing all these parameters and distributions into $\boldsymbol{\theta}$, the likelihood function is defined as

$$
\begin{aligned}
&\boldsymbol{L}(\boldsymbol{\theta}) = \\
&Pr([\mu_1, \boldsymbol{P}_1, \boldsymbol{Y}_1, \boldsymbol{Z}_1, \boldsymbol{X}_1, \boldsymbol{W}_1, \boldsymbol{V}_1], [\mu_2, \boldsymbol{P}_2, \boldsymbol{Y}_2, \boldsymbol{Z}_2, \boldsymbol{X}_2, \boldsymbol{W}_2, \boldsymbol{V}_2], \cdots, [\mu_T, \boldsymbol{P}_T, \boldsymbol{Y}_T, \boldsymbol{Z}_T, \boldsymbol{X}_T, \boldsymbol{W}_T, \boldsymbol{V}_T] | \boldsymbol{\theta}) \\
&= \sum_{\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \cdots, \boldsymbol{X}_T^*} Pr([\mu_1, \boldsymbol{P}_1, \boldsymbol{Y}_1, \boldsymbol{Z}_1, \boldsymbol{X}_1, \boldsymbol{W}_1, \boldsymbol{V}_1] | \boldsymbol{X}_1^*) \cdot Pr([\mu_2, \boldsymbol{P}_2, \boldsymbol{Y}_2, \boldsymbol{Z}_2, \boldsymbol{X}_2, \boldsymbol{W}_2, \boldsymbol{V}_2] | \boldsymbol{X}_2^*, [\mu_1, \boldsymbol{P}_1, \boldsymbol{Y}_1, \boldsymbol{Z}_1, \boldsymbol{X}_1, \boldsymbol{W}_1, \boldsymbol{V}_1]) \cdots \\
&Pr([\mu_T, \boldsymbol{P}_T, \boldsymbol{Y}_T, \boldsymbol{Z}_T, \boldsymbol{X}_T, \boldsymbol{W}_T, \boldsymbol{V}_T] | \boldsymbol{X}_T^*, [\mu_{<T}, \boldsymbol{P}_{<T}, \boldsymbol{Y}_{<T}, \boldsymbol{Z}_{<T}, \boldsymbol{X}_{<T}, \boldsymbol{W}_{<T}, \boldsymbol{V}_{<T}]) \cdot Pr(\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \cdots, \boldsymbol{X}_T^*)
\end{aligned}
\tag{3.18}
$$

where the subscript $< T$ denotes all the observed variables prior to market T. We further know that

$$\begin{aligned}&Pr([\mu_t, \boldsymbol{P}_t, \boldsymbol{Y}_t, \boldsymbol{Z}_t, \boldsymbol{X}_t, \boldsymbol{W}_t, \boldsymbol{V}_t] | \boldsymbol{X}^*_t, [\mu_{<t}, \boldsymbol{P}_{<t}, \boldsymbol{Y}_{<t}, \boldsymbol{Z}_{<t}, \boldsymbol{X}_{<t}, \boldsymbol{W}_{<t}, \boldsymbol{V}_{<t}]) = \\ &\sum_{\boldsymbol{Z}^*_t} Pr(\boldsymbol{Z}_t | \boldsymbol{Z}^*_t, \boldsymbol{X}^*_t, \boldsymbol{W}_t, \boldsymbol{V}_t, \mu_t) \cdot Pr(\boldsymbol{X}_t | \boldsymbol{X}^*_t, \widetilde{\boldsymbol{X}}_t, \boldsymbol{V}_t) \\ &\qquad \cdot Pr(\mu_t | \boldsymbol{Z}^*_t, \boldsymbol{X}^*_t, \boldsymbol{P}_t, \boldsymbol{W}_t, \boldsymbol{V}_t)) \cdot Pr(\boldsymbol{Y}_t, \boldsymbol{P}_t, \boldsymbol{Z}^*_t, \boldsymbol{W}_t, \boldsymbol{V}_t). \end{aligned} \tag{3.19}$$

where $\widetilde{\boldsymbol{X}}_t$ refers to the vector of the most recent rating scores of each coder in market t. And in the first market,

$$\begin{aligned}&Pr([\mu_1, \boldsymbol{P}_1, \boldsymbol{Y}_1, \boldsymbol{Z}_1, \boldsymbol{X}_1, \boldsymbol{W}_1, \boldsymbol{V}_1] | \boldsymbol{X}^*_1) = \\ &\sum_{\boldsymbol{Z}^*_1} Pr(\boldsymbol{Z}_1 | \boldsymbol{Z}^*_1, \boldsymbol{X}^*_1, \boldsymbol{W}_1, \boldsymbol{V}_1, \mu_1) \cdot Pr(\boldsymbol{X}_1 | \boldsymbol{X}^*_1, \boldsymbol{V}_1) \cdot Pr(\mu_1 | \boldsymbol{Z}^*_1, \boldsymbol{X}^*_1, \boldsymbol{P}_1, \boldsymbol{W}_1, \boldsymbol{V}_1) \cdot Pr(\boldsymbol{Y}_1, \boldsymbol{P}_1, \boldsymbol{Z}^*_1, \boldsymbol{W}_1, \boldsymbol{V}_1).\end{aligned} \tag{3.20}$$

In equation (3.19), if we rank coders in market t such that $X^*_1 \geq X^*_2 \geq \cdots \geq X^*_I$,[12]

$$Pr(\mu_t | \boldsymbol{Z}^*_t, \boldsymbol{X}^*_t, \boldsymbol{P}_t, \boldsymbol{W}_t, \boldsymbol{V}_t)) = H_1 \cdot H_2 \cdots H_I$$

and,

$$\begin{aligned}&H_i := Pr(\mu(i) = j^k | \mu(i'), \forall i' < i; \boldsymbol{X}^*_t, \boldsymbol{Z}^*_t, \boldsymbol{P}_t, \boldsymbol{W}_t, \boldsymbol{V}_t) \\ &= \int_{\eta_i} \frac{\exp(\beta_1 P_{j^k} + \beta_2 \text{Duration}_j + \beta_3.\text{NET}_j + \beta_4 X^*_i + (\gamma_0 + \gamma_1 X^*_i + \gamma_2 \text{Experience}_i) \cdot Z^*_j)}{\sum_m \exp(\beta_1 P_m + \beta_2 \text{Duration}_m + \beta_3.\text{NET}_m + \beta_4 X^*_i + (\gamma_0 + \gamma_1 X^*_i + \gamma_2 \text{Experience}_i) \cdot Z^*_m)} dF_\eta(\eta_i)\end{aligned}$$

In words, H_i is the probability of coder i choosing contest $\mu(i)$ given the choices by

[12] For simplicity, assume when two coders carry the same ability level, all firms strictly prefer the one with a smaller subscript.

other better coders in the market. Lastly, the probability of observing the scores for submitted codes are:

$$\begin{aligned}
&Pr(\boldsymbol{Z}_t|\boldsymbol{Z}^*_t, \boldsymbol{X}^*_t, \boldsymbol{W}_t, \boldsymbol{V}_t, \mu_t) = \prod_{j=1}^{J_t} f_{\boldsymbol{Z}_j}(z|Z^*_j, \boldsymbol{X}^*_{\mu_t(j)}, W_j, \boldsymbol{V}_{\mu_t(j)}) \\
&= \prod_{j=1}^{J_t} \prod_{i\in\mu(j)} f_{b_j}(z - \xi_1 X^*_i - \xi_2 Z^*_j - \xi_3.\text{NET}_j - \xi_4 \text{Duration}_j - \xi_5 \text{Experience}_i),
\end{aligned} \quad (3.21)$$

Similarly, for the probability of $Pr(\boldsymbol{X}_t|\boldsymbol{X}^*_t, \widetilde{\boldsymbol{X}}_t, \boldsymbol{V}_t)$, we have:

$$\begin{aligned}
Pr(\boldsymbol{X}_t|\boldsymbol{X}^*_t, \widetilde{\boldsymbol{X}}_t, \boldsymbol{V}_t) = &\prod_{i=1}^{I} [f_{u_1}(X_i - X^*_i) \cdot \mathbb{1}(\text{coder } i \text{ first appear in market } t) + \\
&f_{u_2}(X_i - \lambda X^*_i - (1-\lambda)\widetilde{X}_i) \cdot \mathbb{1}(\text{coder } i \text{ not first appear in market } t)]
\end{aligned} \quad (3.22)$$

The last component $Pr(\boldsymbol{Y}_t, \boldsymbol{P}_t, \boldsymbol{Z}^*_t, \boldsymbol{W}_t, \boldsymbol{V}_t)$ can be similarly decomposed using the primitives I am estimating. So far, I have constructed the likelihood function, and everything on the RHS corresponds to the primitives to be estimated. Standard MLE approach is not practically implementable without further modification, however. To see this, recall that coders have a positive probability of re-entering a market. In fact, this panel structure helps me to achieve identification that will be discussed in the next section; coders are therefore viewed as "global players" of this repeated game, leading to the correlation of likelihood across different markets. Taking logarithm is not implementable, as likelihood across markets is not separable. Even if we maximize likelihood function directly, a dimensionality issue will arise here, as we need to sum over all possible values of $Pr(\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \cdots, \boldsymbol{X}_T^*)$, which can be as many as $\prod_{t=1}^{T} M^{I_t}$ times of evaluation if we assume coders' ability levels are drawn from a finite space of order M. For instance, in the real data, we have more than 1,300 coders and 400 weekly markets. On average, there are 22 coders per market, which implies that, to compute the full likelihood, we need to evaluate the joint probability distribution $Pr(\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, \cdots, \boldsymbol{X}_T^*)$ approximately $2^{22\times 400}$ times, even if each coder's skill level takes only two possible values.

3.4.7.1 A Computationally-Feasible Modified Likelihood-Based Estimator

I therefore adopt the modified estimator suggested in section 3.5.2 of Chapter 2. First, I need to make the following assumption:

Assumption 10 *Suppose the probability of coder i being selected in market t is*

$$e_{it} = \begin{cases} e_0, & \textit{if coder } i \textit{ never participated before,} \\ e_1, & \textit{if coder } i \textit{ participated in market } t-1, \\ 0, & \textit{if coder } i \textit{ appeared both in market } t \textit{ and } t-1. \end{cases}$$

where $0 < e_2 < e_1 < 1$.

Intuitively, I assume that coder enters the market and stay for at most two consecutive weeks. Afterward, they leave the market forever. The entry, stay, and exit decisions are nonstrategic, though. Furthermore, I assume the populations of both coders and contests consist of countlessly many candidates. Thus, there always exist a positive number of coders and contests across all weekly markets. Hence, in each market t, the set of coders, I_t, could be divided into three categories, $\{I_{t,-1}, I_{t,0}, I_{t,+1}\}$, where $I_{t,-1}$ denotes the coders that also appeared in market $t-1$ and stayed in market t, $I_{t,0}$ denotes coders that appeared only in market t and leave forever after t, and $I_{t,+1}$ denotes coders that first appear in market t and stay for one more period. This is illustrated in Figure (3.6). This assumption may seem to be restrictive at first sight, but empirically it is acceptable to focus on coders' consecutive participation behavior only, and treat reentry behavior separately for the same coder. Figure (3.7) shows the frequency coders' consecutive participation behavior. Specifically, more than 92% of consecutive attendance is less than or equal to 2 periods, and less than 8% consecutive participation is more than twice. This adds credit to our assumption: We treat any coder who re-enters the market after about of consecutive participation as a separate identity.

Now, I construct a new likelihood function that integrates out partial information from what we observe:[13]

$$\widetilde{\boldsymbol{L}}(\boldsymbol{\theta}) = \frac{\sum_{[\boldsymbol{P}_2,\boldsymbol{Y}_2],\ldots,[\boldsymbol{P}_T,\boldsymbol{Y}_T]} \boldsymbol{L}(\boldsymbol{\theta})}{\sum_{\boldsymbol{A}_1,\boldsymbol{A}_3,\cdots,\boldsymbol{A}_{T-1}} \sum_{[\boldsymbol{P}_2,\boldsymbol{Y}_2],\ldots,[\boldsymbol{P}_T,\boldsymbol{Y}_T]} \boldsymbol{L}(\boldsymbol{\theta})}$$

$$= Pr(\boldsymbol{A}_1, \boldsymbol{A}_3, \cdots, \boldsymbol{A}_{T-1} | [\boldsymbol{Z}_2, \boldsymbol{X}_2, \mu_2, \boldsymbol{W}_2, \boldsymbol{V}_2], [\boldsymbol{Z}_4, \boldsymbol{X}_4, \mu_4, \boldsymbol{W}_4, \boldsymbol{V}_4], \cdots, [\boldsymbol{Z}_T, \boldsymbol{X}_T, \mu_T, \boldsymbol{W}_T, \boldsymbol{V}_T])$$

$$= \prod_{t=1}^{T/2} \sum_{\boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1}} Pr(\boldsymbol{A}_{2t-1}, \boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}, \mu_{2t-2,+1}, \boldsymbol{W}_{2t-2,+1}, \boldsymbol{V}_{2t-2,+1}],$$

$$[\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}, \mu_{2t,-1}, \boldsymbol{W}_{2t,-1}, \boldsymbol{V}_{2t,-1}]). \tag{3.23}$$

where $\boldsymbol{A}_t := [\mu_t, \boldsymbol{P}_t, \boldsymbol{Y}_t, \boldsymbol{Z}_t, \boldsymbol{X}_t, \boldsymbol{W}_t, \boldsymbol{V}_t]$. In the last equation, the subscript $\{2t, -1\}$ denotes the distribution of variables in market $2t$, of coders staying from market $2t-1$ till in market $2t$; similarly, the subscript $\{2t-2, +1\}$ denotes the distribution of variables in market $2t-2$, of coders staying from market $2t-2$ till market $2t-1$. Essentially, I have integrated out some information with regard to even numbers of markets, and mainly focus on the odd numbers of markets. This enables me to compute market-level likelihood separately without encountering the dimensionality problem. We condition on the information of observables such as Z and X from even markets, as they are relevant to coders' underlying types X^*. The last equality follows from the fact that in market $2t-1$ and $2t+1$, there is no coder overlapping anymore; thus, the joint distributions of observables are independent of each other, conditional on neighbor market observables. Moreover, coders stay for at most two periods, thus the only relevant information from even markets is about the codes from right before

[13] Without loss of generality, assume T is an even number. In addition, there is an abuse of notation – some variables in $\boldsymbol{W}$ are continuously distributed, but I use the summation sign to denote the summation of all possible values for discrete variables and the integration over the support of continuous variables.

and right after market t. I then write the log-likelihood function as:

$$\widetilde{\boldsymbol{LL}}(\boldsymbol{\theta}) = \sum_{t=1}^{T/2} \log \sum_{\boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1}} Pr(\boldsymbol{A}_{2t-1}, \boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}, \mu_{2t-2,+1}, \boldsymbol{W}_{2t-2,+1}, \boldsymbol{V}_{2t-2,+1}],$$
$$[\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}, \mu_{2t,-1}, \boldsymbol{W}_{2t,-1}, \boldsymbol{V}_{2t,-1}])$$
$$= \sum_{t=1}^{T/2} \log \sum_{\boldsymbol{X}^*_{2t-1}, \boldsymbol{Z}^*_{2t-1}} Pr(\boldsymbol{Z}_{2t-1} | \boldsymbol{Z}^*_{2t-1}, \boldsymbol{X}^*_{2t-1}, \boldsymbol{W}_{2t-1}, \boldsymbol{V}_{2t-1}, \mu_{2t-1}) \cdot$$
$$Pr(\mu_{2t-1} | \boldsymbol{Z}^*_{2t-1}, \boldsymbol{X}^*_{2t-1}, \boldsymbol{P}_{2t-1}, \boldsymbol{W}_{2t-1}, \boldsymbol{V}_{2t-1})) \cdot$$
$$Pr(\boldsymbol{Y}_{2t-1} | \boldsymbol{Z}^*_{2t-1}, \boldsymbol{W}_{2t-1}) \cdot Pr(\boldsymbol{P}_{2t-1} | \boldsymbol{Z}^*_{2t-1}, \boldsymbol{W}_{2t-1}) \cdot Pr(\boldsymbol{Z}^*_{2t-1}, \boldsymbol{W}_{2t-1}) \cdot$$
$$Pr(\boldsymbol{X}_{2t-1}, \boldsymbol{X}^*_{2t-1}, \boldsymbol{V}_{2t-1} | [\boldsymbol{Z}_{2t-2,+1}, \boldsymbol{X}_{2t-2,+1}, \mu_{2t-2,+1}, \boldsymbol{W}_{2t-2,+1}, \boldsymbol{V}_{2t-2,+1}], [\boldsymbol{Z}_{2t,-1}, \boldsymbol{X}_{2t,-1}, \mu_{2t,-1}, \boldsymbol{W}_{2t,-1}, \boldsymbol{V}_{2t,-1}]). \tag{3.24}$$

The RHS of equation (3.24) corresponds to the primitives we are interested in estimating. The detailed derivation is provided in the appendix.

The next lemma establishes the consistency result for the estimator.

Lemma 9 *Assume (i) the product space for estimation primitives are compact;*[14] *(ii)* $Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t, \boldsymbol{W}_t, \boldsymbol{V}_t | \boldsymbol{X}^*_t)$ *is continuous in all parameters and probability distributions; (iii) the set of primitives such that* $Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t, \boldsymbol{W}_t, \boldsymbol{V}_t | \boldsymbol{X}^*_t) > 0$ *does not depend on the value of primitives; (iv) there exists a function* $K(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t, \boldsymbol{W}_t, \boldsymbol{V}_t | \boldsymbol{X}^*_t)$ *such that*

1. $\log Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t, \boldsymbol{W}_t, \boldsymbol{V}_t | \boldsymbol{X}^*_t, \boldsymbol{\theta}) - \log Pr(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t, \boldsymbol{W}_t, \boldsymbol{V}_t | \boldsymbol{X}^*_t, \boldsymbol{\theta_0}) \leq K(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t, \boldsymbol{W}_t, \boldsymbol{V}_t | \boldsymbol{X}^*_t)$

   and

2. $EK(\mu_t, \boldsymbol{Z}_t, \boldsymbol{P}_t, \boldsymbol{Y}_t, \boldsymbol{W}_t, \boldsymbol{V}_t | \boldsymbol{X}^*_t) < \infty$*;*

[14] The space include the parametric spaces for (1) $(\beta, \gamma, \boldsymbol{\sigma})$, which are parameters in coders' utility function; (2) $(\xi, \boldsymbol{\kappa})$, which are parameters in the score function in equation (3.11), and lastly, the probability space for the distributions $Pr(Y|Z^*)$, $Pr(P|Z^*, \omega)$, $Pr(Z^*)$ and $Pr(\omega)$.

then the likelihood estimator that maximizes the function in equation (3.19) converges in probability to the true values of the primitives.

A detailed proof can be found in the appendix of Chapter 2. In the next section, I briefly discuss how the observed data patterns from many finite-sized markets lead to the identification of the underlying utility and profit parameters through first identifying the unobserved coder and project heterogeneity. The detailed derivations are presented in Chapter 2, and I will focus on how the stylized facts in the market I am studying leads to identification argument.

3.4.8 Identification Discussion

Full identification discussion consists of three steps. First, within each market, conditional on observing stable matches μ, I identify the joint distribution of $(\boldsymbol{X}_t^*, \boldsymbol{Z}_t^*)$ from jointly observing coder-specific, firm-specific and match-specific characteristics. The identification strategy is based on the eigenvalue-eigenvector decomposition technique developed in Hu (2008). Then, the market-level choice probability can be revealed, which leads to the identification of coders' indirect utility parameters following the well-known discrete-choice literature. Lastly, market-wise demand condition can be identified by observing multiple projects within each market. The identification of firms' profit coefficient is identified from the monotonicity result in Proposition 1.

The intuition behind the first step is that, from data, for any individual coder i, we have three conditionally independent pieces of information related with their ability level (current rating X_t and current and previous final scores $(Z_t, \tilde{Z}_t)$). On the contest side, we are unable to observe information for the same project across different markets as each contest appears in only one market. Instead, we must rely on the information observed from the matching outcomes in *one* market to invoke identification. First, let us focus on the final scores, as they not only reflect coders' innate abilities, X^*, but also the complexity of contests, Z^*. We must be cautious here, though, as the observed scores are driven by the matching allocation, through which contests occurring in the same market are implicitly correlated. Another bit of useful information concerns the equilibrium cash prizes decided by firms: They are the major movements firms make during the extensive-form game and are an equilibrium outcome. Again, the distribution of cash prizes is correlated through the market-level unobservable, ω, and further correlated through the matching allocation, μ.[15] Consequently, we cannot separately identify its distribution without looking at other contests in the same market. Instead, the conditional independence could only be built upon observed matching allocations.

The next lemma formally establishes the conditional independence assumption among observed characteristics. To be more accurate, I return to the explicit market subscript t.

Lemma 10 *The following condition holds for the market-level observables, when con-*

[15] To see this, note that the probability of observing an individual match μ depends on the joint distribution of $(\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{P})$. Thus, the distribution of $\boldsymbol{P}$ is variant with different values of μ.

ditioning on $(\boldsymbol{W}_t, \boldsymbol{V}_t)$

$$(\boldsymbol{X}_t|\boldsymbol{Y}_t) \perp (\boldsymbol{P}_t, \tilde{\boldsymbol{Z}}_t) \perp \boldsymbol{Z}_t|(\boldsymbol{Z}_t^*, \boldsymbol{X}_t^*, \tilde{\boldsymbol{X}}_t, \mu_t),$$

where $\boldsymbol{Y}_t$ *denotes the* $J_t \times 1$ *vector of binary variables, each indicating whether a contest is custom or generic;* $\boldsymbol{X}_t$ *represents the* $I_t \times 1$ *vector of coders' contemporary ratings;* $\boldsymbol{P}_t$ *represents the* $J_t \times 1$ *vector of cash prize for all contests;* $\tilde{\boldsymbol{Z}}_t$ *indicates the* $I_t \times 1$ *vector of all coders' final scores up to their most recent participation;* $\boldsymbol{Z}_t$ *is the* $I_t \times 1$ *vector of final scores for all coders;* $(\boldsymbol{Z}_t^*, \boldsymbol{X}_t^*)$ *is the* $I_t \times J_t$ *matrix of all contests' and coders' types; and* $\tilde{\boldsymbol{X}}_t$ *is the* $I_t \times 1$ *vector of all coders' ratings up to their most recent participation. Lastly,* μ_t *is the observed match in market* t.

Therefore, fixing the market size (I, J, Q), other observed characteristics $(\boldsymbol{W}_t, \boldsymbol{V}_t)$ (suppressed here), the matching allocation and coders' previous performance, $(\mu, \tilde{\boldsymbol{X}})$, I can write down the following equations:

$$\begin{aligned} & f(\boldsymbol{X}, \boldsymbol{P}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z}|\boldsymbol{Y}, \tilde{\boldsymbol{X}}, \mu, I, J, Q) \\ &= \sum_{\boldsymbol{Z}^*, \boldsymbol{X}^*} f(\boldsymbol{P}, \tilde{\boldsymbol{Z}}|\boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}, \mu) \cdot f(\boldsymbol{Z}|\boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}, \mu) \cdot f(\boldsymbol{X}, \boldsymbol{Z}^*, \boldsymbol{X}^*|\boldsymbol{Y}, \tilde{\boldsymbol{X}}, \mu) \end{aligned} \tag{3.25}$$

as well as,

$$
\begin{aligned}
&f(\boldsymbol{P}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z} | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \mu_t, I, J, Q) \\
&= \sum_{\boldsymbol{Z}^*, \boldsymbol{X}^*} f(\boldsymbol{P}, \tilde{\boldsymbol{Z}} | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}, \mu) \cdot f(\boldsymbol{Z} | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}, \mu) \cdot f(\boldsymbol{Z}^*, \boldsymbol{X}^* | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \mu)
\end{aligned} \tag{3.26}
$$

Furthermore, when (X^*, Z^*) follows a discrete distribution, we could partition (or discretize, if observables are continuous) observable characteristics to have the same dimension as the cardinality of the support for (X^*, Z^*), such that equations (3.25) and (3.26) could be written in a matrix form and would provide the nice property of invertibility. Detailed assumptions can be found in Appendix 3.9.3.

Specifically, fixing a certain value of $(\boldsymbol{Y} = \boldsymbol{y}, \tilde{\boldsymbol{X}}, \mu_t, I, J, Q)$, I compute the probability of observing $\boldsymbol{X}^d = \boldsymbol{x}$ as well as various values of $(\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d, \boldsymbol{Z}^d)$

$$
\begin{aligned}
&M_{\boldsymbol{X}^d = \boldsymbol{x}, \boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d, \boldsymbol{Z}^d | \boldsymbol{Y} = \boldsymbol{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q} \\
&= M_{\boldsymbol{Z}^d | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q} \cdot D_{\boldsymbol{Z}^*, \boldsymbol{X}^*, \boldsymbol{X}^d = \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q} \cdot M^T_{\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q}
\end{aligned} \tag{3.27}
$$

where on the LHS, the element on the i-th row and j-th column corresponds to the probability:

$$
Pr(\boldsymbol{X}^d = \boldsymbol{x}, (\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d) = (\boldsymbol{p}, \boldsymbol{z})_j, \boldsymbol{Z}^d = \boldsymbol{z}_i | \boldsymbol{Y} = \mathrm{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q)
$$

where $(\boldsymbol{p}, \boldsymbol{z})_j$(resp. $\boldsymbol{z}_i$) is the j-th(resp. i-th) distinct value for the vector $(\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d)$(resp.

$\boldsymbol{Z}^d$). The first and third matrix on the RHS is similarly defined. The middle matrix D on the RHS is diagonal whose elements are the probability of observing ($\boldsymbol{Z}^* = \boldsymbol{z}, \boldsymbol{X}^* = \boldsymbol{x}', \boldsymbol{X}^d = \boldsymbol{x}$) for various values of ($\boldsymbol{z}, \boldsymbol{x}'$) conditional on ($\boldsymbol{Y} = \boldsymbol{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q$). All matrices are of dimension $(m^J \cdot l^I) \times (m^J \cdot l^I)$. In addition,

$$M_{\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d, \boldsymbol{Z}^d | \boldsymbol{Y} = \boldsymbol{1} - \boldsymbol{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu_t, I, J, Q} = M_{\boldsymbol{Z}^d | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q} \cdot D_{\boldsymbol{Z}^*, \boldsymbol{X}^* | \boldsymbol{Y} = 1 - \boldsymbol{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q} \cdot M^T_{\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d, \boldsymbol{Z}^*, \boldsymbol{X}^* | \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q} \tag{3.28}$$

By inverting equation (3.28) and right-multiplying equation (3.27), I get

$$\begin{aligned} & M_{\boldsymbol{X}^d = \boldsymbol{x}, \boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d, \boldsymbol{Z}^d | \boldsymbol{Y} = \boldsymbol{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q} \cdot M^{-1}_{\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d, \boldsymbol{Z}^d | \boldsymbol{Y} = \boldsymbol{1} - \boldsymbol{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu_t, I, J, Q} \\ & = M_{\boldsymbol{Z}^d | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q} \cdot D_{\boldsymbol{Z}^*, \boldsymbol{X}^*, \boldsymbol{X}^d = \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q} \cdot D^{-1}_{\boldsymbol{Z}^*, \boldsymbol{X}^* | \boldsymbol{Y} = 1 - \boldsymbol{y}, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q} \cdot \\ & M^{-1}_{\boldsymbol{Z}^d | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}} = \boldsymbol{x}, \mu, I, J, Q} \end{aligned} \tag{3.29}$$

Here, the matrices on the LHS are directly computable from data, and the RHS embeds the distribution of unobservables that we are interested in. I use the eigenvalue–eigenvector decomposition method developed in Hu (2008) for nonparametric identification. The next proposition is a significant result to identify the joint distribution of $(\boldsymbol{Z}^*, \boldsymbol{X}^*)$ given the market size and observed match, μ.

Proposition 5 *The joint distribution of $(\boldsymbol{P}, \boldsymbol{Z}^*, \boldsymbol{X}^* | \mu, I, J, Q)$, the conditional distributions of $(\boldsymbol{Z}_j | Z_j^*, \boldsymbol{X}^*_{\mu(j)})$ and $(\boldsymbol{X}_{\mu(j)} | \boldsymbol{X}^*_{\mu(j)}, \tilde{\boldsymbol{X}}_{\mu(j)})$ are all nonparametrically identified* *from observing the joint distribution of* $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{P}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z}, \tilde{\boldsymbol{X}})$ *conditional on a certain* (μ, I, J, Q)*. Moreover, the underlying law of motion* $Pr(X^*_{it}|\widetilde{X}^*_{it})$ *and initial condition* $Pr(X_{i1}|X^*_{i1})$ *are nonparametrically identified.*

In the second step, the market-level choice probability is equivalent to the probability of observing the matching allocation μ_t, and is determined solely by coders' preferences once we know their skill levels. In order to simplify our notation, the $(I, J, Q, \boldsymbol{W}_t, \boldsymbol{V}_t)$ is suppressed from now on. As coders' preferences are affected by $(\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{P})$, the market-level choice probability can be written as $Pr(\mu|\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{P})$. After identifying the conditional distribution of unobserved types of coder–slot pairs, we could apply the Bayes Theorem:

$$Pr(\mu|\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{P}) = \frac{f(\mathbf{P}, \mathbf{Z}^*, \mathbf{X}^*|\mu) \cdot Pr(\mu)}{\sum_\mu f(\mathbf{P}, \mathbf{Z}^*, \mathbf{X}^*|\mu) \cdot Pr(\mu)},$$

where, $f(\mathbf{P}, \mathbf{Z}^*, \mathbf{X}^*|\mu)$ is identified from the previous step and $Pr(\mu)$ is directly observable from data.

Given the knowledge of $Pr(\mu|\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{P})$, we are able to recover coders' utility primitives based on

$$Pr(\mu|\boldsymbol{X}^*, \boldsymbol{Z}^*, \boldsymbol{P}) = Pr(u_{1,\mu(1)} \geq \max_{j \cup \emptyset} u_{1,j}) \cdot Pr(u_{2,\mu(2)} \geq \max_{j \neq \mu(1), \emptyset} u_{2,j}) \cdots Pr(u_{I,\mu(I)} \geq \max_{j \neq \mu(i), \forall i < I, \emptyset} u_{I,j})$$

if we order coders such that $X_1^* > X_2^* > \cdots > X_I^*$.

Note that for each coder i, the conditional probability $Pr(u_{i,\mu(i)} \geq \max_{j \cup \emptyset} u_{1,j})$ is very similar to the individual choice probability in single-agent discrete-choice models. The major difference lies in the way I characterize a coder's choice set. Now that the choice set is identified, I can directly use the argument in the discrete-choice literature to identify coders' utility parameters.

From Step 1, we have identified the joint distribution of $(\boldsymbol{P}, \boldsymbol{Z}^*, \boldsymbol{X}^*)$. We know that, for each contest j, the symmetric equilibrium cash prize depends on (Z_j^*, ω_t), which implies

$$P_j \perp P_{j'} \perp P_{j''} | Z_j^*, Z_{j'}^*, Z_{j''}^*, \omega \tag{3.30}$$

for any project j, j', j'' in market t. Therefore, we could recover the distribution of market unobserved heterogeneity, ω_t, from jointly observing at least three contests in the market. This is in line with the measurement-error model, as all prizes can be viewed as noisy measures for ω_t. Following Hu (2008), we make the following assumption to identify the conditional distribution, $f(P_j|Z_j^*, \omega)$.

Assumption 11 *1. ω is drawn from a finite support $\{w_1, w_2, \cdots, w_n\}$ with $n \leq m$.*

2. There exists a mapping $\rho : \{p_1, p_2, \cdots, p_m\} \to \{1, 2, \cdots, n\}$ such that the following matrix is of full rank $n \times n$.

$$M_{\rho(P_1),\rho(P_3)|Z_1^*,Z_3^*} := [Pr(\rho(P_1) = p, \rho(P_3) = p'|Z_1^*, Z_3^*)]_{p,p' \in \{1,2,\cdots,n\}}$$

The next theorem tells us that we could identify the distribution of a single contest's cash prize, P, conditional on its complexity, Z^*, and the market unobservable ω. This condition distribution could be viewed as firms' pricing strategies.

Proposition 6 *Given Assumption 11, we can nonparametrically identify $Pr(P_j|Z_j^*, \omega)$ and the marginal distribution of ω.*

As discussed in the theory section, the equilibrium cash prize distribution maximizes the interim payoff of each firm at Stage 1. Specifically, let $\delta^* := Pr(P|Z^*, \omega)$, which has been identified, then

$$\delta^* \in argmax_{\delta \in \Sigma} \sum_{l=1}^{m} [\sum_{Z^*_{-j}, P_{-j}} \pi^{\text{int}}(p_l, Z_j^*, Z^*_{-j}, \delta^*_{-j}(\cdot), \omega)] \cdot \delta_j^l$$

where $\pi^{\text{int}}(p_l, Z_j^*, Z^*_{-j}, \delta^*_{-j}(\cdot), \omega)$ is the firm's interim payoff function defined previously. In equilibrium, δ^* is a function of $(c, Z^*, \omega, \beta, \gamma, \sigma, Pr(Z^*), Pr(X^*))$, where only the profit coefficient c is not known. As we already made the assumption with regard to the equilibrium selection rule, we only need to make sure the δ^* function is invertible for c. From part 2 of Proposition 4, the equilibrium distribution of cash prize δ^* is stochastically increasing with respect to c. Thus, the profit coefficient, c, can be identified from

$$c = (\delta^*)^{-1}(Pr(P|Z^*, \omega), Z^*, \omega, \beta, \gamma, \sigma, Pr(Z^*), Pr(X^*)).$$

Thus, I have nonparametrically identified the distribution of unobserved heterogeneity on both sides of the market, and more importantly, identified the preference primitives for firms and coders. Detailed derivation and proofs can be found in Section 3 of Chapter 2. In the subsequent section, I will discuss the estimation results and show how the model fits the data.

3.5 Estimation Results and Discussion

In this section, I present the estimation result using the modified likelihood-based estimator I suggested in the previous section. To construct the (log-)likelihood function, I select markets that have no more than 20 coders; This is because when the number of coders increases, the computational burden will increase exponentially. I firstly use odd numbers of markets to generate the point estimate; then I re-estimate using numbers of markets. I include 204 markets in total (over half of the whole sample). Also, I do not observe actual coders choosing the outside option in each market. Here I permute the "outside option" observations by two means: (1) I treat coders who deliver a coder of score lower than 75 to be choosing the outside option. This is acceptable as submissions with a score worse than 75 is treated as trivial and nonfunctional, and can be partially categorized as coder choosing to contribute nothing in the market; (2) I treat coders appearing more than once in the market to choose the outside option in the previous market. This is also reasonable, as an active coder in the current market is more likely to browse the website for a period of time, but decide not to participate until he sees exciting projects. Nonetheless, if given more data, especially data about the actual outside option observations, the estimation should be more accurate. After adding these observations, my sample size increases by around 10%. Other detailed specifications can be found in Appendix 3.9.4.

The results are summarized in Table 3.8. Here, I discuss in detail how the estimation result informs us about the market's underlying patterns and how they are compared with reduced-form evidence.

3.5.1 Coders' Utility Primitives

First, let us look at coders' utility primitives shown in Panel A of Table 3.8. The cash prize plays an active role in motivating coders to participate, but the magnitude of that role is small, as the cash prizes have been scaled down by 1000. On the other hand, duration has a negative impact, showing coders' preference over shorter projects over longer ones. If we compare the reduced-form regression of how coders sort themselves into different projects, this estimates is in line with the sorting pattern: more skilled coders sort into projects with a shorter duration. Regarding programming language, coders slightly prefer .NET projects than Java projects, which is consistent with the fact that the former is less familiar to most coders.

Coders exhibit a positive individual-level fixed effect, reflecting the fact that more skilled coders tend to show more interest to this platform compared with alternative work opportunities outside the platform. In my model, I also allow coders to perceive the complexity of a job differently by adding the interaction terms between project complexity and coder-wise characteristics. On average, the complexity becomes less desirable to coders with higher skills, although the complexity fixed effect *per se* has a positive coefficient in the model. Experience positively affects the probability of choosing complex projects, which indicates that, coders who have more experience tend to sort themselves into more complex jobs.

Compared with results in Table 3.6, this panel at least partially echoes what I have found in the reduced-form sorting pattern. For instance, it is clear that coders with higher rating score sort themselves into Java projects compared with .NET and other projects; from the structural estimation, .NET is less favorable to coders. As mentioned earlier, this pattern also applies to coders' perception over the duration of a job. Lastly, the fact that coders (insignificantly) sort into less paid jobs can be rationalized by the fact that although they love money, they are hesitant to be engaged in more complex jobs, especially when a coder is more skilled, as suggested by the structural estimates.

3.5.1.1 Determinants of the Final Scores

Another interesting pattern is how the average final scores are determined by various factors beyond the participating coder's skill level. This is summarized in

Panel B of Table 3.8. Firstly, it is clear that final score is positively affected by coder ability and negatively affected by project complexity; .NET projects score slightly less than Java projects; coders participating in projects with shorter duration tend to score higher, which is reasonable, as longer projects tend to be easier to complete. More experienced coders tend to get a lower score. Compared with Table 3.5, most results are consistent regarding observable characteristics.

3.5.1.2 Distribution of Cash Prize, Project Generality and the Underlying Complexity

More importantly, let us explore how data informs us about the underlying degree of heterogeneity on the firm side. We see a clear pattern of a project's cash prize being increasingly affected by its underlying complexity, as reflected in the third row of Panel C in Table 3.8. Similarly, higher market-level unobservables are associated with higher average cash prizes. This is all in line with the theoretical prediction. In addition, cash prizes compensate more for .NET projects and/or projects with a shorter duration, as suggested by the coefficient for .NET dummy and Duration covariate. Lastly, as the number of firms in each market increases, the equilibrium cash prize tend to be higher; this might be explained by the fact of competition. This result is also partially consistent with the reduced-form finding in Table 3.7 except for how cash prizes are related to duration and the number of contests per market. Nevertheless, regression analysis in Table 3.7 does not consider the unobserved heterogeneity of a project, thus can result in biased estimates.

The binary variable Y indicates whether the project is for generic or custom use. More complex jobs are less likely to be generic, as reflected by the negative coefficient for the complexity. .NET projects are slightly more likely to be generic. The longer a project is, the more probable it is a generic project.

Regarding the underlying distribution of project complexity, .NET projects are slightly more likely to be complex. The shorter a project is, the more probable it is a complex job. Combining all the estimation regarding the duration, I have found that shorter projects are more challenging for coders to complete, which may be because they require more coding efficiency and bug-freeness.

Moving to Panel F of the table, we can see that the market-level demand shifter is negatively affected by the number of firms in the market. This is reasonable, as the more firms there are in the market, the less profitable it is to buy submitted codes via this platform because of the limited capacity supply of coders.

3.5.1.3 Distribution of Coders' Latent Ability

Lastly, Let me discuss how the structural model tells about the pattern of coders' latent ability and its evolution over time. First, when a coder is of high ability last period, it is more probable for him to remain skilled during the current period than does a coder with low ability, as shown by the first two rows of Panel G. Next, it is demonstrated by the estimate of λ that a coder's rating score depends heavily on his last-period rating. The disturbance to a coder's rating score becomes smaller, as over time, the rating tend to become more stable. This is also consistent with what the website shows us. Regarding the latent marginal probability of being highly skilled, Chinese coders tend to be more skilled than coders of other nationalities. This might arise from the selection as Chinese coders might sort themselves into this platform more often than others. Most of the estimates are reasonable and consistent with what I find in reduced-form analyses.

3.5.2 Model Fit

In this section, I describe the in-sample fit of the estimated moments. Specifically, in the simulated data, I firstly include the same set of coders for all $T = 204$ weekly markets. By using the point estimates from the previous section, I generate the observed projects (resp. coders) within each market, including their latent complexity levels (resp. ability levels), conditional on observed characteristics such as programming language and duration (resp. nationality and experience). Then, I simulate coders' discrete choices and the final scores they would get from participating in the matched project. Lastly, I pool all observations together and re-divide them into four bins according to the experience percentile of each coder. Table 3.9 depicts the predicted and observed first and second moments between data and model against different experience percentiles, with standard deviation from 200 simulated samples provided in the parentheses. Regarding the first moments, it shows that the model well predicts the trend of both rating scores and final scores a coder would get when he accumulates more experience from participation. The estimates for how much cash prize a coder would get is less accurate than the previous two estimates, probably because of the way I discretize the cash prizes. For all first moment estimates, the last column is always less accurate, though, possibly because the experience is capped above by 20 during the estimation. Thus the more experienced coder sub-sample is noisier.

Regarding the second moment estimates, qualitatively my model well predicts the correlation between the (1) final score and cash prize, and (2) rating score and the cash prize. Quantitatively the latter is better predicted than the former. For both estimates, the lower 10% and upper 30% experience subsamples are worse predicted, probably because these samples have noisier observations. To sum, the overall fit is good, indicating that the structural model well approximates the observed data.

3.6 Counterfactual Experiments

Although I am unable to provide closed-form solution for the whole model due to the complexity of computation when the number of agents gets large, it is possible to explore the market patterns via simulation analysis. In particular, I am interested in how to improve the total surplus in such a market. I conduct three counterfactual experiments using the point estimates from the data and an illustrative model where there are two firms in each market, offering three slots each, and four coders are randomly drawn from population to make their participation choices. The whole time window is 12 weeks which is approximately three months after the counterfactual policies go into effect. Besides computing the welfare of all coders and firms in each market, I also define the degree of "assortativeness" as

$$Assort = \frac{Pr(\mu(i) = j|X_i^* = x_h, Z_j^* = z_h) + Pr(\mu(i) = j|X_i^* = x_l, Z_j^* = z_l)}{Pr(\mu(i) = j|X_i^* = x_h, Z_j^* = z_l) + Pr(\mu(i) = j|X_i^* = x_l, Z_j^* = z_h)}$$

and define the participation rate within each market as the percentage of coders not choosing the outside option. In the first experiment, I am interested in how coders' attitude towards complexity would affect firms' welfare.

3.6.1 Coders' Attitudes Towards Complexity

How different coders perceive the complexity of various projects is influenced by the scheme that the website designs. For instance, by introducing extra reputation rewards for complex projects in addition to the cash prize, coders might view those projects as more attractive than before, potentially due to the fact that their reputation will be built up that benefits their future income even in other platforms. This dimension of preference is captured through the γ_1-term in coders' indirect utility functions: to see this, note that currently γ_1 is estimated to be negative, which shows that more talented coders are more interested in easy jobs rather than complex ones.

In this counterfactual experiment, I would like to see what would happen if their attitude toward complexity, γ_1 be shifted to positive (first scenario) and even more negative (second scenario). In particular, the first scenario represents one type of policy change such as delivering extra point rewards for highly skilled coders to enroll in challenging jobs, whereas the second scenario can be implemented by giving them extra points for contributing to easy jobs. Formally, recall the specification about coders' utility functions:

$$\begin{aligned} u_{i,j^k} &= \beta_0 P_{j^k} + \tilde{u}(X_i^*, Z_j^*, \boldsymbol{V}_i) + \boldsymbol{\beta W}_j + \nu_{ij} \\ &= \beta_0 P_{j^k} + \gamma_0 X_i^* + (\gamma_1 X_i^* + \gamma_2 \boldsymbol{V}_i + \eta_{it}) \cdot Z_j^* + \nu_{ij} \end{aligned} \tag{3.31}$$

In particular, the γ_1-term in high-skilled coders' utility function denotes their overall attitudes towards complexity. In this experiment, we shift the values of γ_1 from upward to twice its absolute value, and downward to 2 times of its current value, to see its impact on coders' and firms' total surplus and other market outcomes. The results are shown in Table 3.10. It first indicates that in both scenarios, the coders' welfare is improved, whereas firms' welfare is hurt by the first scenario, largely affecting the total welfare. This is interesting, as what we typically think is that once the assortativeness is increased, as shown in the fourth row of the first column, firms will be as better off as coders do. Instead, they are harmed by the decrease in participation. The reason is that, when high skilled coders are more motivated to take part in difficult jobs, low-skilled ones may potentially be “squeezed out” as their desired jobs are taken, whether it is an easy or a difficult one; this fact on average harms the firm side as a result. From this experiment, we can see there is a subtle trade-off between assortativeness and participation, as what I find in the market is that low-skilled coders are more passionate about complex jobs, and the market will be better off if we further allow them to do so. Therefore, an incentive scheme that encouraging high-skilled coders to be more active in easy jobs might help improve this market. In the long run, however, it remains uncertain which way to go especially if the market creator would like to attract more skilled labor onto this platform.

3.6.2 Allowing Firms to Price-Discriminate

The current market operates in a way that firms are not able to price discriminate coders with different skill levels. This may result in the overly high payment by firms and misalignment of the match between coders and their most suitable jobs. Now I would like to introduce an alternative regime, in which firms have the choice to design discriminative contracts based on coders’ skill levels. Specifically, firms could choose to add a 50% percentage bonus if the coder participated is of high skill level, on top of their basic cash prize levels. The splitting rules remain the same: the second-place always gets half of the prize for the first-place. The results are shown in Table 3.11. First of all, this new regime improves the participation rate, as demonstrated by the last row in the table. Interestingly, both firms and coders benefit from this new scheme: (especially high-skilled) coders are now awarded more money due to price discrimination; meanwhile, firms with complex jobs can attract better coders to participate. As a result, the total surplus in such a market improves. Lastly, the degree of assortativeness is moderately hurt by price discrimination, which is less of a problem, as the price discrimination leads to coders who previously choose the outside option now stay inside, possibly in less complex jobs. To sum, when firms have more freedom of choosing how to award coders in terms of price discrimination, they gain higher utility, which also benefits coders as a whole. This inspires the market designers to redesign the market so as to improve the total surplus in an efficient manner.

3.7 Conclusion

This paper adopts a two-stage structural model to study an online contest-based labor market, where wage contracts are set by the employer before they match with skilled labor. In the analysis, I establish a computationally practical estimation procedure allowing unobserved heterogeneity to prevail on both sides. Estimates of structural parameters using individual level data from a leading contest-based coding community suggest the importance of accounting for workers' multidimensional preferences over short-term jobs on such platforms. To answer the policy question of how to better design such markets, two counterfactual experiments are conducted, which shed light on the importance of incentivizing workers' preference over different task complexities and firms' price discrimination power, which has been less studied in the existing literature.

This paper takes the first step to establishing a structural model to estimate such finite-sized markets with unobserved heterogeneity. One important direction for future research would be to account for workers' strategic behavior after matching with the tasks. This requires a combination of the current model and the well-developed auction/contest literature. By doing so, one could better understand how workers' *ex post* performance is incentivized by different market schemes. It would be equally interesting if one can compare the economic efficiency and potential skill spillovers across different matching platforms, but this requires a new set of data about alternative matching markets.

3.8 Tables and Figures

Table 3.1: Descriptive Statistics: Contest-specific Characteristics

Variable	Mean	Std. Dev.	Min.	Max.	N
Number of Submissions	2846	3.860506	(3.919741)	1	61
Average Final Score	2845	87.20051	(9.24136)	0	99.94
Winning Cash Payment	2778	578.6886	(318.2293)	0	3000
Duration	2835	6.202469	(2.905825)	0	105
Screening Rate	2832	.9330778	(.1411169)	.25	1
Total DR Points	2828	351.6358	(286.9357)	0	2500
Dummy=1 if Custom	2846	.5523542	(.4973389)	0	1
Programming language	**Percentage (%)**				**N**
Java Custom	36.51				1039
Java	26.91				766
.NET Custom	16.62				473
.NET	17.15				488
Others	2.81				80

Note: Each coder gets three independent final scores from the reviewer board, and the average determines coders' ordinal rankings. Cash prize is the amount of money awarded to the first-place participant. Duration is defined as the number of days from the posting date till the submission due date. DR point reward scheme is observed only after May 12, 2006.

Table 3.2: Descriptive Statistics: Coder-specific Characteristics

Variable	Mean	Std. Dev.	Min.	Max.	N
Rating Score	1,169.49	392.126	1	2,488	9,849
Experience	13.169	20.029	0	194	10,052
Tenure	624.815	523.109	0	3,201	10,052
Final Score	87.328	10.366	0	100	10,061
Final Cash Reward	212.641	303.650	0	3,000	10,052
Final DR Reward	122.236	189.141	0	2,500	8,109
# of participation/week	1.156	0.474	1	14	8,718
Nationalities	**Percentage (%)**				**N**
China	53.82				747
United States	9.37				130
India	5.98				83
Ukraine	3.60				50
Others	27.23				378

Note: The tenure of a coder is defined as the number of days from the date he registered on the website to the date he takes part in the current contest.

Table 3.3: Descriptive Statistics: Market-specific characteristics

Variable	Mean	Std. Dev.	Min.	Max.	N
# of contests/week	7.115	4.743	1	24	400
# of active coders/week	21.795	14.885	1	84	400

Table 3.4: Firms' Vertical Preference: Placement Regression on Rating Score

	Placement	Placement	Placement	Placement	Placement
Rank of Rating Score	0.900***	0.891***	0.839***	0.821***	0.782***
	(50.81)	(46.12)	(147.03)	(37.10)	(26.18)
Experience		-0.00172	-0.00384**	-0.000443	-0.00170
		(-1.66)	(-3.17)	(-0.39)	(-1.30)
Winning Cash Payment		-0.000290***	-0.000489***	0.0000491	0.0000412
		(-5.11)	(-4.02)	(0.34)	(0.10)
Duration		-0.0110	-0.0176	-0.0390**	-0.0474
		(-1.14)	(-0.87)	(-2.60)	(-1.68)
Coder Nationality: China		0.0173	-0.0440	0.0377	-0.00247
		(0.37)	(-0.90)	(0.77)	(-0.06)
Coder Nationality: US		0.219*	0.298**	0.250*	0.306**
		(2.47)	(2.66)	(2.44)	(3.18)
Constant	0.416***	0.460***	0.897***	0.404**	0.663***
	(6.64)	(3.99)	(3.68)	(2.95)	(4.27)
Project Generality FE	N	Y	Y	Y	Y
Technology FE	N	Y	Y	Y	Y
Market FE	N	N	Y	N	Y
IV Regression	N	N	N	Y	Y
Obs	10075	9778	9778	8430	8430
R-sq	0.810	0.811	0.728	0.776	

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. *P-value in parentheses.*

Table 3.5: Final Score Regression

	FinalScore	FinalScore	FinalScore	FinalScore	FinalScore
Winning Cash Payment	-0.00414***	-0.00291***	-0.00221***	-0.00229***	0.0200*
	(-9.94)	(-9.27)	(-6.13)	(-3.78)	(2.52)
Rating Score		0.0177***	0.0182***	0.0116***	0.0121***
		(75.48)	(91.80)	(52.55)	(37.92)
Experience		-0.0263***	-0.0471***	0.00794*	-0.0146**
		(-6.86)	(-11.46)	(2.07)	(-3.00)
Duration		-0.398***	-0.148*	-0.374***	-1.111**
		(-9.49)	(-2.45)	(-6.83)	(-2.86)
.NET Custom		-1.400*	-0.378	-1.288*	0.691
		(-2.50)	(-0.60)	(-2.19)	(0.68)
.NET Generic		-3.507***	-0.457	-2.584***	5.966***
		(-6.36)	(-0.72)	(-4.40)	(5.09)
Java Custom		-1.054*	-0.258	-0.901	3.793***
		(-1.99)	(-0.43)	(-1.60)	(5.62)
Java Generic		-3.443***	-1.180	-2.578***	5.489***
		(-6.39)	(-1.89)	(-4.43)	(4.72)
Coder Nationality: China		0.947***	0.460**	1.432***	1.141***
		(5.87)	(3.13)	(8.52)	(5.51)
Coder Nationality: US		-3.348***	-2.033***	-2.301***	-0.894
		(-8.36)	(-6.14)	(-5.03)	(-1.64)
Constant	89.61***	73.00***	69.29***	79.24***	65.73***
	(387.76)	(115.78)	(93.36)	(113.69)	(23.13)
Market FE	N	N	Y	N	Y
IV Regression	N	N	N	Y	Y
Obs	9776	9555	9555	8253	8253
R-sq	0.011	0.530	0.546	0.476	

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. *P-value in parentheses.*

Table 3.6: Coder's Heterogeneous Preferences: Sorting Patterns

	Whole Sample		Top Coders		Average Coders	
	Robust OLS	2SLS	Robust OLS	2SLS	Robust OLS	2SLS
Cash Payment	-0.0258	-0.0300	-0.0747***	0.00192	-0.237***	-0.182***
	(-1.55)	(-1.09)	(-3.84)	(0.06)	(-9.81)	(-3.70)
Duration	-2.984	-2.803	-15.71***	-18.51***	4.468	2.009
	(-1.41)	(-1.21)	(-6.27)	(-6.64)	(1.46)	(0.56)
.NET Custom	149.9***	149.7***	200.0***	201.1***	146.3**	147.7**
	(4.29)	(4.29)	(5.06)	(5.07)	(3.07)	(3.13)
.NET Generic	68.71*	67.82*	199.1***	215.2***	118.4**	123.7**
	(2.01)	(1.97)	(5.08)	(5.41)	(2.60)	(2.74)
Java Custom	107.4**	106.6**	229.6***	242.3***	110.9*	115.3**
	(3.17)	(3.13)	(5.92)	(6.18)	(2.47)	(2.59)
Java Generic	62.66	61.55	230.7***	250.4***	124.2**	132.2**
	(1.85)	(1.79)	(5.91)	(6.32)	(2.77)	(2.95)
Constant	1115.5***	1117.4***	1262.5***	1223.3***	977.1***	960.3***
	(31.45)	(30.41)	(30.77)	(28.55)	(21.02)	(20.10)
Obs	9571	9571	4652	4652	4919	4919
R-sq	0.007	0.007	0.024	0.021	0.019	0.018

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. *P-value in parentheses.*

Table 3.7: Regression Analysis: Endogenous Pricing of Contests

	Cash Payment	Cash Payment
Duration	36.03***	23.80***
	(6.99)	(8.30)
.NET Custom	-26.50	-2.212
	(-0.73)	(-0.08)
.NET Generic	-229.6***	-192.7***
	(-6.30)	(-6.46)
Java Custom	-155.8***	-138.8***
	(-4.64)	(-5.14)
Java Generic	-278.2***	-220.0***
	(-8.05)	(-7.56)
Number of Coders/Market	0.950*	
	(1.96)	
Number of Contests/Market	-4.222*	
	(-2.43)	
Maximum Opponent Payment	0.267***	-1.244***
	(12.10)	(-32.15)
constant	303.7***	1727.0***
	(7.31)	(35.82)
Market FE	N	Y
Obs	2751	2751
R-sq	0.288	0.394

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. *P-value in parentheses.*

Table 3.8: Structural Estimation Result

		Estimate	**Std. Dev.**
Panel A: Coder's Utility Parameters			
β_1	Cash prize	0.0104	2.52e-9
β_2	Duration	-0.0849	1.28e-4
β_3	.NET	-0.1500	1.21e-5
β_4	Ability	1.4309	1.62e-5
γ_0	Complexity	1.6223	1.54e-5
γ_1	Ability × Complexity	-2.3248	6.15e-5
γ_2	Experience × Complexity	0.1436	4.93e-4
Panel B: Final Score Parameters			
ξ_0	Intercept	-0.3700	0.0059
ξ_1	Ability	1.0004	1.49e-9
ξ_2	Complexity	-0.0089	2.08e-5
ξ_3	.NET	-0.0296	0.0016
ξ_4	Duration	-0.0215	0.0355
ξ_5	Experience	-0.0241	0.0726
σ_b	Std.Dev. of b	0.4622	0.0018
Panel C: Cash Payment Parameters			
ψ_M	Mid-level cash FE	-0.2676	3.45e-6
ψ_H	High-level cash FE	-0.2534	5.10e-5
$\psi_0^{Z^*}$	complexity FE	0.2511	3.34e-6
ψ_0^{ω}	demand shifter FE	2.0027	8.93e-5
ψ_1	.NET	0.4985	1.74e-5
ψ_2	Duration	-0.0940	5.38e-4
ψ_3	No. of projects	0.2080	4.57e-4
Panel D: Project Generality Indicator Parameters			
ϕ_0	Intercept	-1.3349	3.77e-5
ϕ_1	complexity FE	-0.0995	2.35e-5
ϕ_2	.NET	0.0202	5.88e-6
ϕ_3	Duration	0.1846	1.04e-4
Panel E: Project Complexity Parameters			
ι_0	intercept	-0.0754	2.00e-5
ι_1	.NET	0.0165	4.47e-6
ι_2	Duration	-0.6023	1.20e-4
Panel F: Market Demand Shifter Parameters			
χ_0	intercept	0.5054	3.40e-5
χ_1	no. of projects	-0.5901	1.74e-4
Panel G: Coder's Latent Ability Parameters			
$Pr(X_t^* = x_H \| X_{t-1}^* = x_L)$	Law of motion	0.4094	4.40e-6
$Pr(X_t^* = x_H \| X_{t-1}^* = x_H)$	Law of motion	0.7262	1.95e-5
λ	coefficient in equation 3.9	0.1252	2.66e-6
σ_1	Std.dev. in equation 3.8	0.8538	1.82e-4
σ_2	Std.dev. in equation 3.9	0.4570	1.38e-4
$Pr(X_{t-1}^* = x_H \| \text{Not Chinese})$	Marginal prob. of high ability if not Chinese	0.7033	1.54e-5
$Pr(X_{t-1}^* = x_H \| \text{Chinese})$	Marginal prob. of high ability if Chinese	0.8049	2.16e-5
T	Number of markets	204	

Note: $*p < 0.05, **p < 0.01, ***p < 0.001$. Standard errors are calculated from 22 Bootstrapped samples due to the length of time taken to do the estimation. In Panel A, cash prize is scaled down by 1000, adjusted for their reliability bonus in Panel A, and further discretized in Panel C into three intervals: p_L for prizes smaller than the median, p_M for equal to median and p_H for larger than median in Panel C. Score is scaled down by 100. Duration is capped from above by 14. Experience is capped up by 20. Rating scores are scaled down by 1000.

Table 3.9: Model Fit

	First Moment						Second Moment			
Experience	Rating Score		Final Score		Cash Prize		Corr(Final Score, Cash)		Corr(Rate,Cash)	
percentile b/w	Data	Model	Data	Model	Data	Model	Data	Model	Data	Model
0 - 10%	1.0552	1.3747	0.8319	0.8987	0.1688	0.1619	0.1410	0.0019	-0.0184	0.0035
		(0.0634)		(0.0126)		(0.0424)		(0.0780)		(0.0738)
10 - 25%	1.0207	1.3565	0.8589	0.9087	0.2576	0.1964	0.0181	0.0102	-0.0618	0.0006
		(0.0434)		(0.0091)		(0.0372)		(0.0632)		(0.0571)
25 - 50%	1.0767	1.3952	0.8778	0.9058	0.4099	0.2437	0.1033	0.0062	0.0157	0.0083
		(0.0381)		(0.0085)		(0.0336)		(0.0541)		(0.0483)
50 - 70%	1.1982	1.4325	0.8982	0.8880	0.4625	0.2734	0.0138	0.0064	-0.0254	-0.0102
		(0.0391)		(0.0097)		(0.0411)		(0.0591)		(0.0534)
70 - 100%	1.4416	1.4283	0.9394	0.8484	0.6919	0.3858	-0.0562	-0.0050	-0.0334	-0.0030
		(0.0334)		(0.0099)		(0.0407)		(0.0568)		(0.0448)

Table 3.10: Counterfactual Experiment 1: Coders' Attitudes Towards Complexity

New values of γ_1	-2γ_1	2γ_1
Coder surplus change (%)	10.5182	57.5541
	(18.4574)	(18.6697)
Firm surplus change (%)	-37.8889	3.2451
	(41.1096)	(10.4817)
Total surplus change (%)	-35.1617	5.3782
	(37.0937)	(10.0079)
Assortativeness change (%)	696.9877	-16.3580
	(1477.3)	(10.9636)
Participation rate change (%)	-54.2458	37.0392
	(22.3489)	(2.0887)

Note: Standard errors are calculated from 100 simulated samples. The scaling and discretization are the same as in the estimation part.

Table 3.11: Counterfactual Experiment 2: Allowing Firms to Price Discriminate

	Adding Bonus for High-type Coders	95% Confidence Interval
Coder surplus change (%)	0.0215	[-0.0264, 0.0566]
	(0.0339)	
Firm surplus change (%)	0.1610	[-0.3304,0.2874]
	(1.8740)	
Total surplus change (%)	0.1723	[-0.2919, 0.2509]
	(1.7628)	
Assortativeness change (%)	-0.3397	[0, 0]
	(11.4086)	
Participation rate change (%)	0.0206	[0, 0]
	(0.2749)	

Note: Standard errors are calculated from 100 simulated samples. The scaling and discretization are the same as in the estimation part.

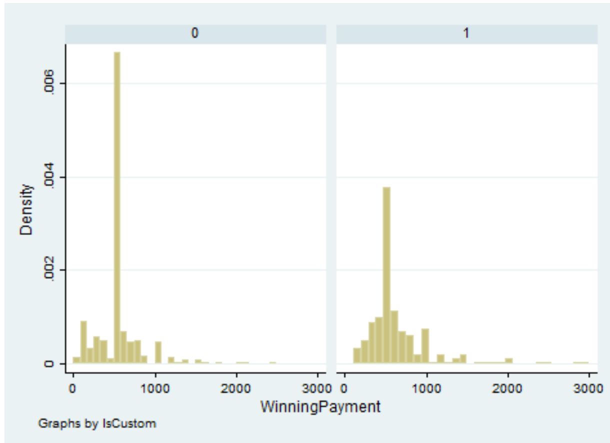


Figure 3.1: Cash Reward and the Generality of the Project

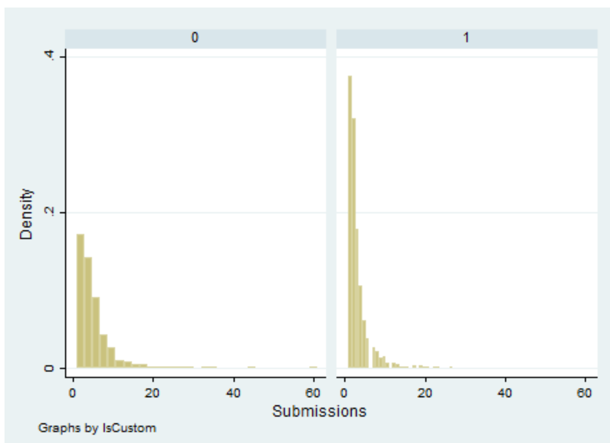


Figure 3.2: Number of Submissions and the Generality of the Project

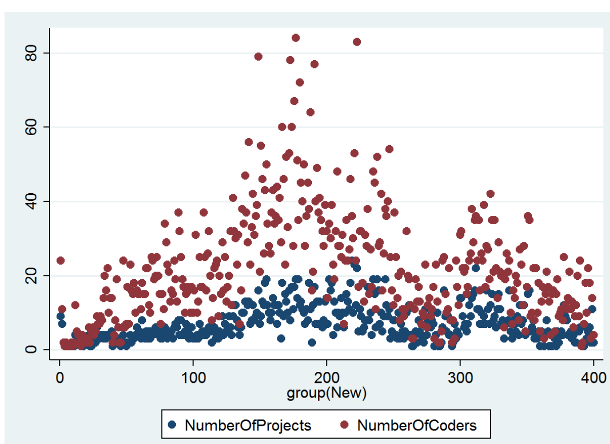


Figure 3.3: The Number of Projects and Coders over Time

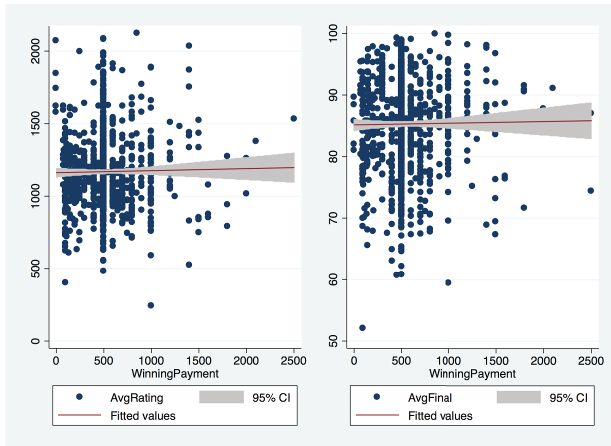


Figure 3.4: The Distribution of Avg. Rating Scores and Avg. Final Scores within a Project

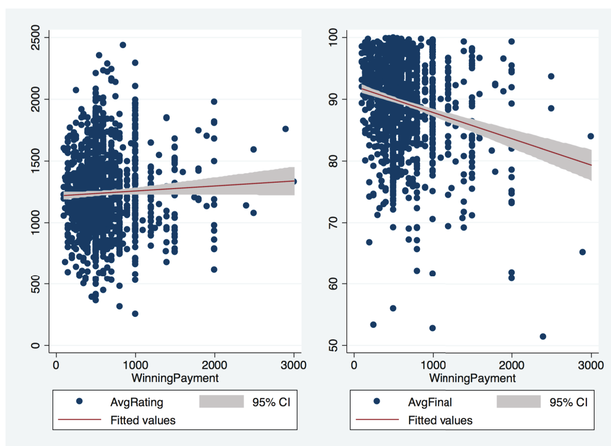


Figure 3.5: The Distribution of Avg. Rating Scores and Avg. Final Scores within a Project

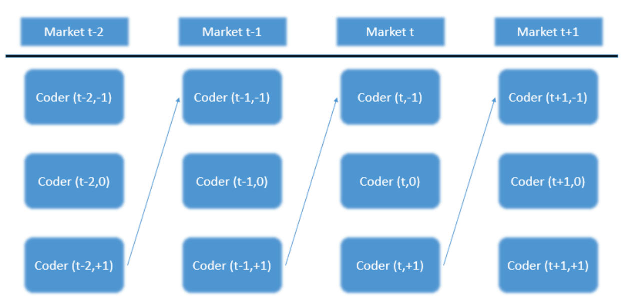


Figure 3.6: Coders' Consecutive Participation Illustration

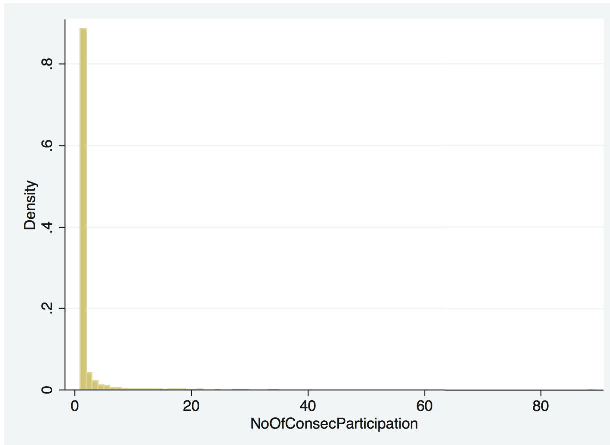


Figure 3.7: Coders' Consecutive Participation Distribution

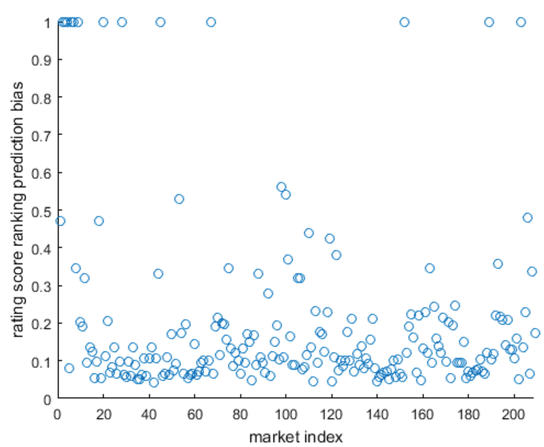


Figure 3.8: Model Fit: Coder's Rating Score Rank Prediction Bias

3.9 Appendix

3.9.1 An Alternative Two-step Estimation Procedure

As an alternative approach to the structural estimation procedure discussed in Section 3.5, I conduct a two-step estimation, wherein the first step, by using the final score and rating score of each coder, I estimate their ability levels from a fixed-effect regression; in the second step, I adopt the benchmark case estimation procedure in Chapter 2 to recover primitives of the model. Compared with the general approach, this approach restricts the data generating the process of coders' rating scores from their observable and unobservable characteristics. Specifically, I assume rating scores are generated from a linear fixed-effect model discussed below. By making this additional restriction, I can incorporate more information in the second step, as I do not encounter the dimensionality problem after recovering coders' latent ability levels in the first step.

3.9.1.1 Step 1: Fixed-Effect Regression to Recover Ability

In market t, for any coder i that submits to contest j, the final score is determined from

$$Z_{ij}^t = \xi_1 X_{it}^* + \xi_2 Z_{jt}^* + \xi_3 W_{ijt} + b_{jt}, \tag{3.32}$$

where X_{it}^* denotes the coders' ability in market t, Z_{jt}^* denotes the contest's unobserved heterogeneity – i.e., its complexity; W_{ijt} are the observed characteristics of coder i and/or contest j, and lastly, b_{jt} is the contest-level idiosyncratic taste shock that follows mean-zero normal distribution with unknown variance. In this equation, both X_{it}^* and Z_{jt}^* are unobserved by the econometrician. However, the coders' ability level X_{it}^* is imperfectly measured by their rating scores:

$$X_{it} = X_{it}^* + \zeta_i + u_{it}, \tag{3.33}$$

where ζ_i is the fixed effect for frequently participating coders, and is independent of the idiosyncratic shock, u. [16] To simplify the estimation procedure, I first separately estimate coders' abilities using a fixed-effects regression model:

$$Z_{ij}^t = \xi_1 X_{it}^* + \xi_3 W_{ijt} + \widetilde{Z}_{jt}^*, \tag{3.34}$$

[16] For ζ_i, I tried three sets of specifications: Fixed effect for coders who participated more than 14 times, more than 20 times, and more than 30 times.

where $\widetilde{Z}^*_{jt}$ is the contest-level fixed effect that summarizes both the unobserved heterogeneity Z^*_{jt} and the contest-level idiosyncratic shock b_{jt}. Thus, by plugging equation (3.43) back to equation (4.14), we have

$$Z^t_{ij} = \xi_1 X_{it} + \xi_{3,1} W_{it} + \xi_{3,2} W_{jt} + \widetilde{Z}^*_{jt} + \widetilde{\zeta}_i + \widetilde{u}_{it}, \tag{3.35}$$

where W_{ijt} is decomposed into $[W_{it}, W_{jt}]$, and $\widetilde{\zeta}_i = -\xi_1 \cdot \zeta_i$ and $\widetilde{u}_{it} = -\xi_1 \cdot u_{it}$. Assuming the independence between $\widetilde{u}$ and other observables and fixed effects, we are able to consistently estimate the coefficient $[\xi_1, \xi_{3,1}]$ and back out the ability level X^*_{it} given observable characteristics. This enables us to construct the residual that relates to the unobserved heterogeneity of contests, Z^*_{jt}.

$$Z^t_{ij} - \widehat{\xi}_1 \widehat{X}^*_{it} - \widehat{\xi}_{3,1} W_{it} \approx \xi_2 Z^*_{jt} + \xi_{3,2} W_{jt} + b_{jt} \tag{3.36}$$

and the parameters $(\xi_2, \xi_{3,2}, \sigma_b)$ are to be estimated in the second stage.

Table 3.12 shows the fixed-effects regression of the average final score. A clear positive causality between the coders' ratings, proxying their skill, and the average final score is reinforced here. This is in line with my theoretical assumption, in which the average final score will monotonically increase with one's skill level. Furthermore, I have tried three specifications of coder-wise fixed effects, which account for the fixed effect of coders who have participated more than (1) 14 times, (2) 20 times and (3) 30 times. The table shows that the point estimates among the three specifications are very similar. In the main estimation part, therefore, I use the estimated result that accounts for fixed effects of coders who have completed more than 30 projects.

3.9.1.2 Step 2: Structural Estimation of Primitives

In this step, I parametrize coders' indirect utility function, equilibrium cash prize distribution and other model primitives as follows. Firstly, I discretize the observed cash prize into two intervals: above or equal to median and below median. Then, the probability of the cash prize lying in each interval follows a logit model:[17]

$$
\begin{aligned}
&Pr(P_{jt} = p^H | Z_j^*, \omega_t) = \\
&\frac{\exp[\psi_H + \psi_0^{Z^*} \cdot Z_j^* + \psi_0^{\omega} \cdot \omega_t + \psi_1.\text{NET} + \psi_2\text{Duration} + \psi_3 J]}{\exp[\psi_H + \psi_0^{Z^*} \cdot Z_j^* + \psi_0^{\omega} \cdot \omega_t + \psi_1.\text{NET} + \psi_2\text{Duration} + \psi_3 J] + 1},
\end{aligned}
\tag{3.37}
$$

The parameters to be estimated therefore are $(\psi_H, \psi_0^{Z^*}, \psi_0^{\omega}, \psi_1, \psi_2, \psi_3)$. The probability of a project being generic or custom also follows a logit model:

$$
Pr(Y_{jt} = 1 | Z_j^*) = \frac{\exp[\phi_0 + \phi_1 \cdot Z_j^* + \phi_2.\text{NET} + \phi_3\text{Duration}]}{\exp[\phi_0 + \phi_1 \cdot Z_j^* + \phi_2.\text{NET} + \phi_3\text{Duration}] + 1}
\tag{3.38}
$$

[17] In the subsequent discussion, I only include the .NET dummy but not the Java dummy, as they compose 98% of the observations, and in order to avoid multicollinearity, I treat all projects that are not .NET as belonging to the Java family.

and the parameters to be estimated therefore are $(\phi_0, \phi_1, \phi_2, \phi_3)$. On the other side of the market, for coder's indirect utility function, we specify

$$\begin{aligned} u_{i,j^k} = {} & \beta_1 P_{j^k} + \beta_2 \text{Duration} + \beta_3.\text{NET} + \beta_4 X_i^* + (\gamma_0 + \gamma_1 X_i^* \\ & + \gamma_2 \text{Experience}) \cdot Z_j^* + \nu_{ij}, \end{aligned} \tag{3.39}$$

and the parameters to be estimated are (β_1, β_2, β_3, β_4, γ_0, γ_1, γ_2). Lastly, I assume the unobserved heterogeneities of contests are drawn from a binary space, $\{z_L, z_H\}$, and how it relates to observables also follows a logit model:

$$Pr(Z_j^* = z_H) = \frac{\exp[\lambda_0(1 + \lambda_1.\text{NET} + \lambda_2 \text{Duration})]}{\exp[\lambda_0(1 + \lambda_1.\text{NET} + \lambda_2 \text{Duration})] + 1} \tag{3.40}$$

and the parameters to be estimated are (λ_0, λ_1, λ_2). The market-level unobservable ω is also drawn from a binary space $\{w_L, w_H\}$, with probability of being w_H depends on the number of projects in the market according to a probit model:

$$Pr(\omega_t = w_H) = Pr(\chi_0 + \chi_1 J + v_t \geq 0) \tag{3.41}$$

where v_t follows a standard normal distribution. The parameters to be estimated are (χ_0, χ_1).

As in the data, I do not observe coders choosing outside option, I impute coders who have appeared in the previous two markets but do not appear in the current market as choosing the outside option. This enlarges my sample size by about 45%. More important, it enables me to give more accurate estimates for the cash prize coefficient.

3.9.2 Result using the Alternative Estimation Procedure

In the empirical specification, I conduct a two-step estimation, wherein the first step, by using the final score and rating score of each coder, I estimate their ability levels from a fixed-effect regression; in the second step, I adopt the benchmark case estimation procedure in Chapter 2 to recover primitives of the model.

3.9.2.1 Step 1: Fixed-Effect Regression to Recover Ability

In market t, for any coder i that submits to contest j, the final score is determined from

$$Z_{ij}^{t} = \xi_1 X_{it}^{*} + \xi_2 Z_{jt}^{*} + \xi_3 W_{ijt} + b_{jt}, \tag{3.42}$$

where X_{it}^{*} denotes the coders' ability in market t, Z_{jt}^{*} denotes the contest's unobserved heterogeneity – i.e., its complexity; W_{ijt} are the observed characteristics of coder i and/or contest j, and lastly, b_{jt} is the contest-level idiosyncratic taste shock that follows mean-zero normal distribution with unknown variance. In this equation, both X^*_{it} and Z^*_{jt} are unobserved by the econometrician. However, the coders' ability level X^*_{it} is imperfectly measured by their rating scores:

$$X_{it} = X^*_{it} + \zeta_i + u_{it}, \tag{3.43}$$

where ζ_i is the fixed effect for frequently participating coders, and is independent of the idiosyncratic shock, u. [18] To simplify the estimation procedure, I first separately estimate coders' abilities using a fixed-effects regression model:

$$Z^t_{ij} = \xi_1 X^*_{it} + \xi_3 W_{ijt} + \widetilde{Z}^*_{jt}, \tag{3.44}$$

where $\widetilde{Z}^*_{jt}$ is the contest-level fixed effect that summarizes both the unobserved heterogeneity Z^*_{jt} and the contest-level idiosyncratic shock b_{jt}. Thus, by plugging equation (3.43) back to equation (4.14), we have

$$Z^t_{ij} = \xi_1 X_{it} + \xi_{3,1} W_{it} + \xi_{3,2} W_{jt} + \widetilde{Z}^*_{jt} + \widetilde{\zeta}_i + \widetilde{u}_{it}, \tag{3.45}$$

where W_{ijt} is decomposed into $[W_{it}, W_{jt}]$, and $\widetilde{\zeta}_i = -\xi_1 \cdot \zeta_i$ and $\widetilde{u}_{it} = -\xi_1 \cdot u_{it}$. Assuming the independence between $\widetilde{u}$ and other observables and fixed effects, we are able to consistently estimate the coefficient $[\xi_1, \xi_{3,1}]$ and back out the ability level X^*_{it} given

[18] For ζ_i, I tried three sets of specifications: Fixed effect for coders who participated more than 14 times, more than 20 times, and more than 30 times.

observable characteristics. This enables us to construct the residual that relates to the unobserved heterogeneity of contests, Z^*_{jt}.

$$Z^t_{ij} - \widehat{\xi}_1 \widehat{X}^*_{it} - \widehat{\xi}_{3,1} W_{it} \approx \xi_2 Z^*_{jt} + \xi_{3,2} W_{jt} + b_{jt} \tag{3.46}$$

and the parameters $(\xi_2, \xi_{3,2}, \sigma_b)$ are to be estimated in the second stage.

Table 3.12 shows the fixed-effects regression of the average final score. A clear positive causality between the coders' ratings, proxying their skill, and the average final score is reinforced here. This is in line with my theoretical assumption, in which the average final score will monotonically increase with one's skill level. Furthermore, I have tried three specifications of coder-wise fixed effects, which account for the fixed effect of coders who have participated more than (1) 14 times, (2) 20 times and (3) 30 times. The table shows that the point estimates among the three specifications are very similar. In the main estimation part, therefore, I use the estimated result that accounts for fixed effects of coders who have completed more than 30 projects.

3.9.3 Detailed Assumptions and Derivations in Identification Discussion

Recall the two equations I establish in the first step of identification:

Table 3.12: Fixed Effect Regression on Final Score

Final Score	(1) Model 1	(2) Model 2	(3) Model 3
RatingScore	0.0196***	0.0195***	0.0192***
	(0.000350)	(0.000330)	(0.000310)
Experience	-0.0679***	-0.0585***	-0.0437***
	(0.00686)	(0.00656)	(0.00627)
Intercept	64.07***	64.77***	65.42***
	(0.358)	(0.339)	(0.324)
$\geq$ 14 experience Coder FE	Y		
$\geq$ 20 experience Coder FE		Y	
$\geq$ 30 experience Coder FE			Y
Contest FE	Y	Y	Y
N	9,823	9,823	9,823
R^2	0.622	0.606	0.595

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

$$f(\boldsymbol{X}, \boldsymbol{P}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z} | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \mu, I, J, Q)$$
$$= \sum_{\boldsymbol{Z}^*, \boldsymbol{X}^*} f(\boldsymbol{P}, \tilde{\boldsymbol{Z}} | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}, \mu) \cdot f(\boldsymbol{Z} | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}, \mu) \cdot f(\boldsymbol{X}, \boldsymbol{Z}^*, \boldsymbol{X}^* | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \mu) \tag{3.47}$$

as well as,

$$f(\boldsymbol{P}, \tilde{\boldsymbol{Z}}, \boldsymbol{Z} | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \mu_t, I, J, Q)$$
$$= \sum_{\boldsymbol{Z}^*, \boldsymbol{X}^*} f(\boldsymbol{P}, \tilde{\boldsymbol{Z}} | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}, \mu) \cdot f(\boldsymbol{Z} | \boldsymbol{Z}^*, \boldsymbol{X}^*, \tilde{\boldsymbol{X}}, \mu) \cdot f(\boldsymbol{Z}^*, \boldsymbol{X}^* | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \mu) \tag{3.48}$$

Then, by making the following assumptions, a formal identification result is es-

Table 3.13: Empirical Result – Alternative Approach

Parameter	Description	point estimate	std.dev.
Panel A: Coders' utility parameters:			
β_1	Cash prize	0.2729***	0.0553
β_2	Duration	-1.2234***	0.1447
β_3	.NET projects	-0.3109***	0.0479
β_4	Ability	-0.8383**	0.3930
γ_0	Complexity	1.1980***	0.4502
γ_1	Ability$\times$ Complexity	0.5238	0.5745
γ_2	Experience $\times$ Complexity	-0.2481***	0.0099
Panel B: Average final score parameters:			
ξ_2	Complexity	-0.0205***	0.0044
ξ_3	.NET projects	0.0019	0.0038
ξ_4	Duration	0.0580***	0.0031
σ_b	Std.Dev. of b	0.0648***	0.0019
Panel C: Cash prize parameters:			
ψ_H	High-level prize FE	2.9052***	0.2339
$\psi_0^{Z^*}$	Complexity FE	1.8209***	0.1920
ψ_0^{ω}	Demand shifter FE	0.6492	1.3534
ψ_1	.NET projects	-0.1552	0.1677
ψ_2	Duration	-0.6726	0.6580
ψ_3	# of Projects	-0.3048***	0.0500
Panel D: Generic project indicator parameters:			
ϕ_0	Intercept	-0.6115	1.4515
ϕ_1	Complexity FE	-0.2437	0.2631
ϕ_2	.NET projects	0.6007***	0.1128
ϕ_3	Duration	0.4463	0.5730
Panel E: Project-level unobservable parameters:			
λ_0	Intercept	-0.3506	0.4735
λ_1	.NET projects	-0.1111	0.1418
λ_2	Duration	-0.4448***	0.1994
Panel F: Distribution of market-level unobservable:			
χ_0	Intercept	0.7835***	0.2997
χ_1	# of projects	-0.1936**	0.0726
T	No. of Markets	344	

Note: $*p < 0.1, **p < 0.05, ***p < 0.01$. Standard errors are calculated from 100 Bootstrapped samples. In Panel A, cash prize is scaled down by 1000, adjusted for their reliability bonus in Panel A, and further discretized in Panel C into two intervals: p_L for prizes smaller than the median, and p_H for higher than or equal to median in Panel C. Duration is scaled down by taking log(1+Duration). Experience is scaled down by taking log(1+Experience).The final scores are scaled down by 100.

tablished in Proposition 2.

Assumption 12 *1. Each contest's, j, complexity is drawn from the finite space $\{z_1, z_2, \cdots, z_m\}$.*

2. Each coder's, i, ability is drawn from the finite space $\{x_1, x_2, \cdots, x_l\}$.

3. The final score, Z_{ij}, is drawn from a bounded atomless support $[0, 100]$.

4. The cash prize is drawn from the finite space $\{p_1, p_2, \cdots, p_M\}$ with $M \geq m$.

5. The rating score, X_{it}, is drawn from an arbitrarily large but discrete support $\{\widetilde{x_1}, \widetilde{x_2}, \cdots, \widetilde{x_L}\}$.

As all latent types are discrete, we would like to write equations (3.25) and (3.26) in matrix forms. The following assumption is crucial.

Assumption 13 *For any observed $(\mu, \widetilde{\boldsymbol{X}} = \boldsymbol{x}, I, J, Q)$,*

1. There exists a mapping $\phi_X : \{\widetilde{x_1}, \widetilde{x_2}, \cdots, \widetilde{x_L}\} \to \{x_1, x_2, \cdots, x_l\}$ such that for any contest j: $\forall \boldsymbol{x}, \boldsymbol{x'} \in \{\widetilde{x_1}, \widetilde{x_2}, \cdots, \widetilde{x_L}\}^{|\mu(j)|}, \boldsymbol{x} \neq \boldsymbol{x'}$, and for any observed $\tilde{\boldsymbol{X}}_{\mu(j)}$,

$$Pr(\boldsymbol{X}^d_{\mu(j)} = \boldsymbol{x} | \boldsymbol{X}^*_{\mu(j)} = \boldsymbol{x}, \tilde{\boldsymbol{X}}_{\mu(j)}) > Pr(\boldsymbol{X}^d_{\mu(j)} = \boldsymbol{x} | \boldsymbol{X}^*_{\mu(j)} = \boldsymbol{x'}, \tilde{\boldsymbol{X}}_{\mu(j)}),$$

where, for any coder i, $X^d_i = \phi_X(X_i)$.

2. There exists a mapping $\phi_P : \{p_1, p_2, \cdots, p_M\} \to \{1, 2, \cdots, m\}$.

3. *There exists a mapping $\phi_{z1} : [0, 100] \to \{1, 2, \cdots, l \cdot m\}$ for each final score of the winning coders $Z_{j,1}$.*

4. *There exists a mapping $\phi_{z2} : [0, 100] \to \{1, 2, \cdots, l\}$ for each final score of the non-winning coders $Z_{j,-1}$ and a vector of values $\boldsymbol{y}$, such that the following matrix is of full rank $m^J \cdot l^I$:*

$$M_{\boldsymbol{P}^d, \tilde{\boldsymbol{Z}}^d, \boldsymbol{Z}^d | \boldsymbol{Y}=\boldsymbol{y}, \tilde{\boldsymbol{X}}=\boldsymbol{x}, \mu_t, I, J, Q}. \tag{3.49}$$

where for each nonwinning coder i, $Z^d = \phi_{z2}(Z_i)$ and $\tilde{Z}^d = \phi_{z2}(\tilde{Z}_i)$; for each winning coder i, $Z^d = \phi_{z1}(Z_i)$ and $\tilde{Z}^d = \phi_{z1}(\tilde{Z}_i)$; and for each contest j, $P^d = \phi_p(P_j)$.

3.9.4 Detailed Specifications in Structural Estimation

In the structural estimation, I assume coders' latent ability levels are drawn from a binary space $\{0, 2\}$. The project-wise latent complexity level is drawn from $\{-1, 1\}$. The market-wise demand condition is similarly drawn from $\{-1, 1\}$. The cash prizes are discretized into three values: $\{0, 500, 3000\}$. Final scores are divided by 100, thus ranging from 0 to 1. Rating scores are divided by 1000, thus ranging from 0 to 2.5. Experiences are capped above by 20. When calculating the log-likelihood function, I selected markets with less than or equal to 20 coders, and in total there are 204 such markets, a bit over than half of the whole sample.

Chapter 4

Identification of Simultaneous Contests with Two-sided Unobserved Heterogeneity

4.1 Introduction

In recent years, a growing amount of markets begins to adopt contests and/or auctions to allocate scarce economic goods and services to multiple players. Examples can be found in rent-seeking activities by politicians such as lobbying a desirable political prize, research and development races on different new drugs among pharmaceutical companies and procuring temporary jobs to part-time talents via the Internet. There exists an abundant amount of economic questions to be answered while seeing this trend in real life. From a market-design point of view, researchers are interested in how different regimes of allocating the economic good to participants can achieve a better market welfare. Also, it remains uncertain whether it would be better off to introduce more competition on the other side of the market, i.e. auctioneers/contest holders, to boost the performance of bidders/contestants. To answer questions like these, one should firstly understand what the current market patterns tell us about players' preferences over the economic good to be allocated via an auction or a contest. This is not an easy task, as unobserved characteristics may mostly affect players' behavior in equilibrium. In this paper, I develop a two-stage structural model where players initially choose among several contests held simultaneously, and then participate, getting paid according to the rank order of their submitted bids/efforts. In the model, it is also allowed that non-winners get part of the economic good according to a commonly-observed splitting rule. The major theme of this paper is to suggest a nonparametric identification and estimation strategy to recover both player-wise and contest-wise unobserved heterogeneity, from observing (1) all players' first-stage equilibrium choices over different contests, and (2) the second-stage equilibrium effort they exert to win.

In particular, my model assumes that for each economic good to be auctioned [1], there is one extra layer of heterogeneity beyond observable characteristics such

[1] I will use terms in auctions and contests literature interchangeably, though primarily I am studying a multi-prize contest problem. For instance, I call what the contestants contribute in a

as the monetary prize. This heterogeneity could originate from the complexity of meeting some minimum requirement for the job, and/or a virtual reward benefiting contestants' future activities, such as a good reputation in the market. This is usually difficult to observe or measure from data but is also as important in determining market players' behavior. To see this, contestants' will bid according to their preference over the unobserved part of a job; on the other hand, contest holders will determine how much prize they would like to propose given all the relevant features of their jobs. The econometric model is further complicated by the fact that contestants may also carry unobserved characteristics determining the market outcomes. Examples can be their valuation over the prize awarded by a contest, or their latent abilities to complete certain jobs. Using a simple additively separable parametric setting, I can exploit the variations in contestants' observable choice probabilities over multiple contests when facing different numbers of peer contestants to uncover the underlying joint distribution of their unobserved heterogeneity by solving the two-stage model described in the previous paragraph. Furthermore, contestants' unobserved types can be nonparametrically recovered from the distribution of equilibrium bids. Identification through the one-to-one mapping between equilibrium bids and unobserved bidders' valuation has been established by Guerre, Perrigne and Voung (2000), and the major departure here is that I am solving an all-pay "auction" with multiple prizes awarded according to the rank order. By combining the two steps, I can nonparametrically identify the

contest as "bids".

full model, which can be well estimated when having enough contestant variations across markets in the data.

4.2 Related Literature

This paper contributes to several strands of the literature. First, it borrows theoretical results from all-pay auctions and multi-prize contests with private information of bidder- (contestant-) types. When a contest is static in the sense that no multiple rounds exist, it is isomorphic to private-valued all-pay auctions with multiple prizes. (Moldovanu and Sela, 2006) The existence and characterization results of all-pay auctions date back to Milgrom and Weber (1982) but only one prize is granted in a standard all-pay auction. On the other hand, in the strand of contest design literature, Moldovanu and Sela (2001) proves the existence of a (unique) symmetric equilibrium bidding strategy when multiple prizes are granted within a contest. Here, I extend Moldovanu and Sela's model to study multiple parallel contests held at the same time faced by multiple contestants. Kvasov (2006) characterizes mixed-strategy equilibria in a two-player complete-information simultaneous contests with budget constraint and identical valuation, whereas in this paper I study the simultaneous contests among multiple bidders with private information and heterogeneous valuation over auctioned items. Compared with the optimal contest design literature such as Modovanu and Sela (2001), I do not consider contest designer's optimization problem; instead, I assume contest prizes and splitting rules are exogenously given, but allow the contest prize to be correlated with contest types in an arbitrary way.

The competition among bidders over multiple objects also naturally relates to the growing study of multi-unit and combinatorial auctions. Multi-unit Auction literature examines the case where the auctioneer holds multiple homogeneous and divisible goods, allowing bidders to submit a price-quantity pair of bids. In this paper, however, the items to be auctioned are assumed to be heterogeneous in various aspects, and therefore is more closely related to the (single-round) combinatorial auctions over heterogeneous goods studied in Cantillon and Pesendorfer (2006) and Kim, Olivares and Weintraub (2014). One feature is that only one auctioneer exists in their models, trying to solve a complex "Winner Determination Problem" (Rothkopf et.al., 1998 and Sandholm, 1999) that chooses the best bidder-item allocation to maximize the total revenue from all auctions. In this paper, I study the case that multiple auctioneers are present in the market, each holding a single-item multi-prize contest and implicitly competing to win over the best bidders. Therefore the winner determination problem is much more straightforward than in the standard combinatorial auctions. To my best knowledge, the most closely related auction literature to this paper is Gentry, Komarova, and Schiraldi (2016), in which they study the simultaneous first-price single-item sealed-bid auctions, allowing bidders to hold non-additive preference over combinations of auctioned items. Bidders submit a vector of bids to all items being auctioned in the market at the same time. In this paper, I split the game into two stages: at the first stage, bidders need to make an entry decision of being involved in at most one auction in the market; afterward, they engage in a multi-prize single-unit contest as described previously. To restrict bidders to involve no more than one auction/contest a time seems to be less general than in Gentry, Komarova, and Schiraldi (2016), but can instead provide very useful information on bidders' choice probability which leads to the full identification of auction/contest-wise and bidder-wise unobserved heterogeneity. Furthermore, my model provides a clean-cut way to explicitly model capacity constraint which is sometimes present in real life applications.

Lastly, the theoretical framework is closely related to the two-stage model established in DiPalantino and Vojnovic (2009), except for the multi-prize aspect of the contests, extending their all-pay auction framework in the second stage. This two-stage model provides a way to depict the capacity constraint faced with every bidder and most importantly serves to identify the underlying heterogeneity across auctions/contests in empirical studies, which provides a novel insight into the structural literature that mainly discusses bidder-wise unobserved heterogeneity. The extension to multiple prizes within a contest has its empirical virtue of better fitting into real-life markets such as sports tournaments, crowd-souring coding platform, and lobbying in political activities.

To summarize, complementing existing literature, this paper studies simultaneous contests that (1) extends standard single-item auctions to provide multiple prizes; (2) endogenizes the entry decision of bidders toward multiple parallel contests before the competition, and (3) nonparametrically identifies both contest-wise and contestant-wise unobserved heterogeneity.

In the subsequent section, I set up a two-stage model, defining the preference structure of the players on both sides. The equilibrium characterization is then established in Section 4. Afterward, I discuss in Section 5 the nonparametric identification results when players' unobserved types are continuous and bounded, and I allow auction-wise unobserved types to be either discrete or continuous. I then suggest an estimation procedure in Section 6. Monte Carlo simulation is then provided in Section 7. Lastly, I conclude in Section 8.

4.3 Model Setup

From this section, I denote one side of the market as "contests" and the other side as "bidders," with a bit abuse of terminology. An outline of the model is as follows. Consider a two-stage game where there are N bidders and M contests randomly selected to be present in market t. At Stage 1, each of the bidders chooses one contest to enter, after perfectly observing his valuation towards every contest award, but not the valuation of others. Choosing the outside option generates a payoff normalized to zero. At Stage 2, every bidder exerts a certain level of effort in the contest he chooses without knowing the first-stage choices of other bidders, and the payoff depends on (1) which contest he steps into, (2) his valuation towards that contest, and (3) the rank order of his effort level within the contest. The rule of splitting the prize according to the rank order is common knowledge.

For each contest, the prize consists of two parts: the actual monetary reward, R, and a virtual reward (or cost if negative), ϵ, that is correlated with money and could be interpreted as the reputation for performing Ill in the contest, and/or the difficulty of meeting the minimum requirement of the contest. In particular, the payoff of a bidder with valuation v exerting effort level b is $v(R+\epsilon)-b$ if he ranks the highest, $\phi_{j,2}v(R+\epsilon)-b$ if he is the second-place, $\phi_{j,3}v(R+\epsilon)-b$ if third-place and so on. Without loss of generality, I assume $\phi_{j,k}=0, \forall k\geq 3$, and $\phi := \phi_{j,2} \in [0,1)$ does not differ across contests. I further assume that the total reward is larger than zero all the time, i.e. $R+\epsilon > 0, \forall R, \epsilon$, as a contest would never attract any participation if the total reward is negative. Here the all-pay feature is clear – no matter which rank a bidder ends up with, his effort is already exerted and cannot be contingent on his rank order or be reverted.

The valuation of each bidder is independently drawn from a commonly known distribution, $F(\cdot)$, on a bounded interval, $[0, \bar{V}]$. Both the monetary prize R and the virtual prize ϵ can be drawn from either a discrete or a continuous space. Suppose in market t, there are in total J_t different values for $R + \epsilon$, and for each value, there are M_j contests in the market. In other words, $\sum_{j=1}^{J_t} M_j = M$. This total value of prizes denoted as the "class" of contests and all other information regarding the contests and bidders is commonly known to every market participants at the initial stage. From now on, I suppress the market subscript t for the ease of illustration.

4.4 Equilibrium Notion

A mixed-strategy for bidder i with valuation v consists of two parts: $[\boldsymbol{\pi}(v), \boldsymbol{\beta}(v)]$, where $\boldsymbol{\pi}(v)$ is a $(J-1) \times 1$ vector that tells us the probability of bidder i choosing one contest in each class of the contests[2] , and within the class, he uniformly randomizes in choosing which contest to join; $\boldsymbol{\beta}(v)$ is a $J \times 1$ vector of bidding functions at the second stage for each class of the contests. Here, bidders (resp. contests) are interchangeable as long as they have the same "type", i.e. valuation of a bidder (resp. the 2-dimensional rewards, $[R, \epsilon]$, of a contest). I restrict myself in discussing symmetric Bayes-Nash Equilibrium only, which specifies that each bidder yields the highest expected payoff by playing the equilibrium strategy, believing that others are

[2]And of course, if a bidder plays a pure strategy of entering one contest for sure, $\boldsymbol{\pi}(v)$ will be all zeros except for one element.

following the symmetric Bayes-Nash Equilibrium strategy. The following proposition establishes the existence result of such an equilibrium:

Proposition 1 There exists a symmetric mixed-strategy equilibrium to this two-stage game.

Proof: I first prove there exists a (unique) monotone symmetric pure-strategy Bayes Nash equilibrium for the second-stage sub-game. At Stage 2, assuming other bidders all play a monotone symmetric bidding strategy $\beta(\cdot)$, the expected payoff for bidder i with value v_i to bid b in contest class j is:

$$\begin{aligned} E\Pi_j(v_i, b) &= v_i(R_j + \epsilon_j)(1 - p_j(1 - \hat{F}_j(\beta^{-1}(b)))^{N-1} \\ &+ \phi_j v_i(R_j + \epsilon_j)(N-1)p_j(1 - \hat{F}_j(\beta^{-1}(b)))(1 - p_j(1 - \hat{F}_j(\beta^{-1}(b)))^{N-2} - b, \end{aligned} \tag{4.1}$$

where $p_j := Pr(\text{bidder } k \text{ chooses contest class } j)$ and
$\hat{F}_j := Pr(\beta(v_k) \leq b|\text{bidder } k \text{ chooses contest class } j)$. Further, I denote $\hat{f}$ as the derivative of $\hat{F}$ agains v. Note that p_j is the same for all other bidders as bidder i has no prior knowledge regarding other bidders' types at the time of bidding; this equivalently means other bidders' choice probability at Stage 1 is *ex ante* equal to bidder i's. Deriving the First Order Condition with respect to b gives us the following

equation:

$$\begin{aligned}\beta'(v_i) =& v_i(R_j+\epsilon_j)(N-1)(1-p_j(1-\hat{F}_j(v_i)))^{N-2}p_j\hat{f}_j(v_j) \\ &- v_i\phi(R_j+\epsilon_j)(N-1)(1-p_j(1-\hat{F}_j(v_i)))^{N-2}p_j\hat{f}_j(v_i) \\ &+ v_i\phi(R_j+\epsilon_j)(N-1)(N-2)(1-p_j(1-\hat{F}_j(v_i)))^{N-3}p_j(1-\hat{F}_j(v_i))p_j\hat{f}_j(v_i)\end{aligned} \tag{4.2}$$

Moreover, when $v_i = 0$ it is obvious that the optimal strategy is to bid zero, therefore:

$$\begin{aligned}\beta(v_i) = \int_0^{v_i} x(1-\phi)(R_j+\epsilon_j)(N-1)(1-p_j(1-\hat{F}_j(x)))^{N-2}p_j\hat{f}_j(x)dx \\ + \int_0^{v_i} x\phi(R_j+\epsilon_j)(N-1)(N-2)(1-p_j(1-\hat{F}_j(x)))^{N-3}p_j(1-\hat{F}_j(x))p_j\hat{f}_j(x)dx\end{aligned} \tag{4.3}$$

It is easy to check that the second derivative function $\beta'(v) \geq 0$, thus satisfying the monotonic assumption of the bidding strategy. Furthermore, the equilibrium payoff for bidder i is:

$$\begin{aligned} E\Pi^*(v_i) = & v_i\phi(R_j+\epsilon_j)(1-p_j(1-\hat{F}_j(v_i)))^{N-1} \\ & + (1-\phi)(R_j+\epsilon_j)\int_0^{v_i}(1-p_j(1-\hat{F}_j(x)))^{N-1}dx \\ & + \phi(R_j+\epsilon_j)(N-1)\int_0^{v_i}(1-p_j(1-\hat{F}_j(x)))^{N-2}p_j(1-\hat{F}_j(x))dx \\ & - \phi(R_j+\epsilon_j)(N-1)\int_0^{v_i}(1-p_j(1-\hat{F}_j(x)))^{N-2}p_jx\hat{f}(x)dx \end{aligned} \tag{4.4}$$

Next, conditional on knowing what bidders will behave in the second stage, I move back to the first stage – This is a simultaneous-move game with finite action space and independent type space regarding bidders, then according to Milgrom and Weber(1985), I know that a mixed-strategy Bayes-Nash Equilibrium always exists. As the game is symmetric, I may select one symmetric equilibrium. In fact, when the bidders' types are invariant to which contest they participate in, the symmetric equilibrium is unique according to Proposition 4.1 of DiPanlatino and Vojnovic (2009).

I now characterize the equilibrium entry strategy $\boldsymbol{\pi}(\cdot)$ for any bidder i at Stage 1. Following a similar argument as in DiPalantino and Vojnovic (2009), I firstly calculate the marginal benefit for bidder i to choose one contest in class j:

$$\Pi'_j(v_i) := \frac{\partial E\Pi^*(v_i)}{\partial v_i} \tag{4.5}$$

$$= (R_j + \epsilon_j)(1 - p_j(1 - \hat{F}_j(v_i)))^{N-1} + \phi(R_j + \epsilon_j)(N-1)(1 - p_j(1 - \hat{F}_j(v_i)))^{N-2} p_j(1 - \hat{F}_j(v_i)$$

By checking the second order derivative of $E\Pi^*$, I know that the marginal benefit stated above is positive and increasing in v_i, i.e. $E\Pi^*(\cdot)$ is a non-decreasing continuous convex function of v_i, and $\Pi'(\cdot)$ is continuous in v_i. Applying Corollary B.1 of DiPalantino and Vojnovic (2009), I have:

$$\Pi'_j(0) = \Pi'_k(0), \forall j, k \text{ such that } P_j > 0, P_k > 0; \tag{4.6}$$

$$\Pi'_j(0) \geq \Pi'_k(0), \forall j, k \text{ such that } P_j > 0, P_k = 0; \tag{4.7}$$

where,

$$\Pi'_j(0) = (R_j + \epsilon_j)(1 - p_j)^{N-1} + \phi(R_j + \epsilon_j)(N-1)(1 - p_j)^{N-2} p_j$$

Intuitively, it means that for any contest that has zero entry probability, it must be the case that the marginal benefit of entering it cannot exceed that of any contest with a positive entry probability for the lowest-type bidder; and if this is true, it must be true for all higher-type bidders. Otherwise, in equilibrium, there is always a profitable deviation for the lowest-type bidder to choose the contest that all bidders currently have zero probability to enter.

Using this result, I could derive the equilibrium entry probability for each contest. I do this progressively. Rank the contest classes such that $R_1 + \epsilon_1 > R_2 + \epsilon_2 > \cdots > R_J + \epsilon_J > 0$. First, consider the case that all classes of contests have positive entry probabilities. I therefore solve for the following system of equations:

$$\begin{cases} (R_1 + \epsilon_1)(1 - p_1)^{N-1} + \phi(R_1 + \epsilon_1)(N-1)(1-p_1)^{N-2}p_1 = C \\ (R_2 + \epsilon_2)(1 - p_2)^{N-1} + \phi(R_2 + \epsilon_2)(N-1)(1-p_2)^{N-2}p_2 = C \\ \cdots \\ (R_J + \epsilon_J)(1 - p_J)^{N-1} + \phi(R_J + \epsilon_J)(N-1)(1-p_J)^{N-2}p_J = C \\ \sum_{j=1}^{J} p_j M_j = 1 \end{cases} \tag{4.8}$$

where only $[p_1, p_2, \cdots, p_J, C]$ are unknowns and there are $J+1$ polynomial equations. Note that by fixing other coefficients, the LHS of all the equations but the last one is decreasing in p on its interval $[0, 1]$. I consider the case that $p_1 = p_2 = \cdots = p_J = 1$. Thus, all LHS of the first J equations are equal to zero, but clearly the last equation does not hold. I now let p_J decrease from 1 towards 0. Meanwhile, for each value of p_J, I solve for $p_1, \cdots, p_{J-1}$ so as to equate the LHS of all the first J equations. If N

is an even number, the real-valued solutions for $p_j, j < J$ always exist; if N is odd, I compute that the Hermite form of the following polynomial:

$$(1-\phi)(R_j+\epsilon_j)(1-p_j)^{N-1}+(N-1)\phi(R_j+\epsilon_j)(1-p_j)^{N-2}-C \tag{4.9}$$

As long as the signature of the Hermite form is positive, I could ensure the existence of a real solution for p_j. As p_J decreases and other p's decreases accordingly, their weighted sum $\sum_{j=1}^{J} M_j p_j$ gets closer to 1. If it gets to 1 before p_J approaches zero, I have found the solution $[p_1, p_2, \cdots, p_J]$, as other p's must not exceed p_J. If when $p_J = 0$, the weighted sum is still larger than 1, I let $p_J = 0$ and decrease p_{J-1} from 1 towards 0. Following similar argument, I find a solution or move to the next p_j. Now consider the case where all $p_j, j \geq 2$ are set to be zero. Then it must be the case that $p_1 > \frac{1}{M_1}$, where p_1 is such that $(R_1+\epsilon_1)(1-p_1)^{N-1}+\phi_1(R_1+\epsilon_1)(N-1)(1-p_1)^{N-2}p_1 = R_2+\epsilon_2$. Given the monotonicity of the LHS as a function of p_1, I further know that

$$\Pi_1'(0) := (R_1+\epsilon_1)(1-\frac{1}{M_1})^{N-1}+\phi(R_1+\epsilon_1)(N-1)(1-\frac{1}{M_1})^{N-2}\frac{1}{M_1} > R_2+\epsilon_2$$

In this case, the equilibrium would be that all bidders randomize in class 1 of the contests, with an entry probability being $\frac{1}{M_1}$. To sum, I could always find a solution to this problem.

Next, I solve for the individual entry probability, $\boldsymbol{\pi}(v)$, given a bidder's type. First, I claim that the following lemma holds:

Lemma 1: Suppose in equilibrium, class $1, 2, \cdots, K$ contests have positive entry probabilities, $K \leq J$. Then there exists $0 \leq v_K \leq v_{K-1} \leq \cdots v_2 \leq v_1 \leq \bar{V}$ such that:

1. $\Pi'_j(v) = \Pi'_1(v) \forall j = 1, 2, \cdots, l, v \in [v_{l+1}, v_l)$
2. $\Pi'_j(v) < \Pi'_1(v) \forall j = l+1, \cdots, K, v \in [v_{l+1}, v_l)$

Proof: Following a similar argument in DiPalantino and Vojnovic (2009), I define the $[v_1, v_2 \cdots, v_K]$ as:

$$p_j(1 - \hat{F}_j(v)) > 0, \forall v \in [0, v_j)$$

$$p_j(1 - \hat{F}_j(v)) = 0, \forall v \in [v_j, \bar{V}]$$

I firstly show that $0 \leq v_K \leq v_{K-1} \leq \cdots v_2 \leq v_1 \leq \bar{V}$. It is easily shown that $v_K \geq 0$ as otherwise, I have $p_K(1 - \hat{F}_K(0)) = 0$, which corresponds to $p_K = 0$. This, however, contradicts our assumption that $p_K > 0$. Similarly, to show $v_K \leq v_{K-1}$, I consider the case that $v_{K-1} < v_K$, then there exists $v \in [v_{K-1}, v_K)$ such that

$$p_{K-1}(1 - \hat{F}_{K-1}(v)) > 0$$

$$p_K(1 - \hat{F}_K(v)) = 0$$

This, however, cannot be true, as I have $\Pi'_{K-1}(v) = R_{K-1} + \epsilon_{K-1} \geq R_K + \epsilon_K > \Pi_K(v)$. This means it cannot happen that a bidder of type v only participates in contest class K but not class $K-1$ as the marginal benefit of moving to the latter is more than staying with the former. Following the same logic, I could prove $0 \leq v_K \leq v_{K-1} \leq \cdots v_2 \leq v_1 \leq \bar{V}$. Also, for $v \in [v_{l+1}, v_l)$, I know that he will have a positive probability in joining contest class $1, 2, \cdots, l$ but not class $l+1, \cdots, K$.

To show $\Pi'_j(v) = \Pi'_1(v) \forall j = 1, 2, \cdots, l, v \in [v_{l+1}, v_l)$, I consider a case that $\exists j \leq l$ such that $\Pi'_j(v) > \Pi'_1(v)$ for some $v \in [v_{l+1}, v_l)$. This cannot happen as I already have $E\Pi^*_j(v) = E\Pi^*_1(v), \forall j \leq l$, and once Π'_j exceeds Π'_1, it would be profitable for the bidder of type v to choose contest class j instead of class 1. This contradicts the statement that when $v \in [v_{l+1}, v_l)$, the probability of entering contest class 1 through l are all positive. It follows similar arguments to show that $\Pi'_j(v) < \Pi'_1(v)$ cannot happen either.

To show $\Pi'_j(v) < \Pi'_l(v) \forall j = l+1, \cdots, K, v \in [v_{l+1}, v_l)$, I consider that $\Pi'_j(v) \geq \Pi'_1(v)$ for some $v \in [v_{l+1}, v_l)$ and $j \geq l+1$. I then choose a positive number ν such that, $R_j + \epsilon_j - \Pi'_j(v_{l+1} - \nu) < \Pi'_j(v) - \Pi'_j(v_{l+1})$. This is ensured by the fact that Π' is a continuous function, and as v approaches v_{l+1} from left side, Π' approaches $R_j + \epsilon_j$ arbitrarily. Thus, I have $\Pi'_j(v) - \Pi'_j(v_{l+1} - \nu) \geq \Pi'_j(v) - \Pi'_j(v_{l+1}) > R_j + \epsilon_j - \Pi'_j(v_{l+1} - \nu)$, which indicates $\Pi'_j(v) > R_j + \epsilon_j$ which can never be true. This completes the proof.

To solve for the individual entry probability, $\boldsymbol{\pi}(v)$, given a bidder's type, I firstly state the following equation holds:

$$p_j(1-\hat{F}_j(v)) = \int_v^{\bar{V}} \pi_j(x)dF(x) \tag{4.10}$$

Thus, I have: $\pi_j(v) = -\frac{1}{f(v)}\frac{\partial p_j(1-\hat{F}_j)}{\partial v}$. Suppose $v \in [v_{l+1}, v_l)$, then $\pi_j(v) = 0, \forall j \geq l+1$. Then it can be shown that, $\forall j \leq l$, $\pi_j(v) = \frac{x_j(v)}{M_j}$, where $x_j(v) := Prob($ choose class j $|V \in [v, v_{l+1}))$. This probability could be solved iteratively. First solve for $x_1(v_1)$ which is the probability of choosing class 1 when $v \geq v_1$. This is trivial as I have $x_1(v_1) = \frac{1}{M_1}$. Next, I solve for $x_1(v_2), x_2(v_2)$ via the following system of equations:

$$\begin{cases} (R_1+\epsilon_1)(1-\frac{1}{M_1}(1-F(v_1)) - \frac{x_1(v_2)}{M_1}(F(v_1)-F(v_2)))^{N-1} \\ +\phi(R_1+\epsilon_1)(N-1)(1-\frac{1}{M_1}(1-F(v_1)) - \frac{x_1(v_2)}{M_1}(F(v_1)-F(v_2)))^{N-2}(\frac{1}{M_1}(1-F(v_1)) \\ +\frac{x_1(v_2)}{M_1}(F(v_1)-F(v_2))) = C \\ (R_2+\epsilon_2)(1-\frac{x_2(v_2)}{M_2}(F(v_1)-F(v_2)))^{N-1} \\ +\phi(R_2+\epsilon_2)(N-1)(1-\frac{x_2(v_2)}{M_2}(F(v_1)-F(v_2)))^{N-2}(\frac{x_2(v_2)}{M_2}(F(v_1)-F(v_2))) = C \\ x_1(v_2)+x_2(v_2) = 1 \end{cases}$$

Complicated as this system of nonlinear equations seems, I can use a similar argument to track down its solution as what I did in deriving the unconditional entry probabilities, p_j. I do this progressively, and could derive $x_j(v_1), x_j(v_2), \cdots, x_j(v_l), x_j(v)$, $j = 1, 2, \cdots, l$. After solving $x_j(v)$, I could derive the individual entry probabilities. The most intuitive implication of Lemma 1 is that when I divide bidders into $K+1$ ordered groups according to their skill levels, in equilibrium, a bidder of skill level $v \in [v_{k+1}, v_k)$ randomizes his entry in the first k classes of contests. Thus, I proved the existence and characterized the symmetric Bayes-Nash Equilibrium for this two-stage game.

4.5 Nonparametric Identification

Recall that the goal of identification in this game consists of two parts: (i) The distribution of the contest complexity, ϵ; (ii)The distribution of the valuation, v_i, for bidders that have positive entry probabilities to at least one of the contests in the market. I now discuss the strategies for two cases respectively: when contests' unobserved types are discrete and when they are continuous.

4.5.1 When Contest Types are Discrete

In this section, I assume that the contest types are drawn from a finite space with a cardinality of Q. Here, I do not restrict the contest type to be independent of its monetary prize, i.e. $Pr(\epsilon|R) \neq Pr(\epsilon)$ for some (R, ϵ). Furthermore, I allow contest prizes in the same market to be correlated with each other, i.e. $Pr(\boldsymbol{R}|\boldsymbol{\epsilon}) \neq \Pi_j Pr(R_j|\epsilon_j)$. This is to capture the possibility of firms strategically setting their monetary rewards to compete with their peer firms in the market, and the market-level unobserved factors affecting contest holders' pricing behavior. The only restriction on the contest side is that the contest types are independent of the number of bidders in the market when conditioning on their monetary prizes. This assumption naturally holds when the contest designers have no knowledge about how many bidders will appear in the market when they design their rewarding schemes. The type space on the bidder side is assumed to be i.i.d. drawn from $[0, \bar{V}]$. More importantly, it is assumed that fixing the number of contests and their monetary prizes; the researcher can observe at least Q distinct values of the number of bidders, N, from data. This is to ensure enough variation in equilibrium choice probabilities in the first stage to recover the discrete distribution of contests' unobserved types.

The identification strategy consists of two steps. First, I exploit the variation in entry probabilities in response to the different amount of bidders in a market to recover the joint distribution of contest types, $\boldsymbol{\epsilon} := [\epsilon_1, \epsilon_2, \cdots, \epsilon_Q]$. Suppose in the data, there are infinitely many repeated markets where we can observe many markets with the same monetary prize vector, $\boldsymbol{R} = \boldsymbol{r}$. I therefore am able to compute the empirical probability toward each contest in the market. I write down the following equation:

$$Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, N) =$$

$$\sum_{\boldsymbol{\epsilon}} \underbrace{Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, \boldsymbol{\epsilon}, N)}_{:=\hat{p}_j} \cdot Pr(\boldsymbol{\epsilon}|\boldsymbol{R}, N) \tag{4.11}$$

Note that once I knew the values of $(\boldsymbol{r}, \boldsymbol{\epsilon}, N)$, $\hat{p}_j$ could be computed by the system of equations (4.8) and the algorithm provided in the modeling part, as all other coefficients in the system of equations are observable to the econometrician. More important, I have the following independence result:

$$Pr(\boldsymbol{\epsilon}|\boldsymbol{R}, N) = Pr(\boldsymbol{\epsilon}|\boldsymbol{R}) \tag{4.12}$$

Therefore, fixing the value of $\boldsymbol{R} = \boldsymbol{r}$, equation (4.11) could be written using every distinct value of N. Suppose I can observe $Q^j \geq Q$ distinct values of N, I will have the following equation:

$$\begin{bmatrix} Pr(R=r_j \text{ is chosen } |\boldsymbol{r}, n_1) \\ Pr(R=r_j \text{ is chosen } |\boldsymbol{r}, n_2) \\ \cdots \\ Pr(R=r_j \text{ is chosen } |\boldsymbol{r}, n_{Q^J}) \end{bmatrix}_{Q^J \times 1} =$$

$$\begin{bmatrix} Pr(R=r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_1}, n_1), Pr(R=r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_2}, n_1), \cdots, Pr(R=r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_{Q^J}}, n_1) \\ Pr(R=r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_1}, n_2), Pr(R=r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_2}, n_2), \cdots, Pr(R=r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_{Q^J}}, n_2) \\ \cdots \\ Pr(R=r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_1}, n_{Q^J}), Pr(R=r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_2}, n_{Q^J}), \cdots, Pr(R=r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_{Q^J}}, n_{Q^J}) \end{bmatrix}_{Q^J \times Q^J} \cdot$$

$$\begin{bmatrix} Pr(\boldsymbol{e_1}|\boldsymbol{r}) \\ Pr(\boldsymbol{e_2}|\boldsymbol{r}) \\ \cdots \\ Pr(\boldsymbol{e_{Q^J}}|\boldsymbol{r}) \end{bmatrix}_{Q^J \times 1}$$

The LHS is an $Q^J \times 1$ vector that is observed from data. The first matrix on the RHS could be computed from equilibrium argument. The last vector on the RHS is the goal of identification. A just-identified case would be when $Q_J = Q$, and assuming the first two matrics on the RHS are both invertible, I could identify the joint distribution of $\boldsymbol{\epsilon}$ given $\boldsymbol{R} = \boldsymbol{r}$ by the following:

$$\begin{bmatrix} Pr(\boldsymbol{e_1}|\boldsymbol{r}) \\ Pr(\boldsymbol{e_2}|\boldsymbol{r}) \\ \cdots \\ Pr(\boldsymbol{e_{Q^J}}|\boldsymbol{r}) \end{bmatrix} =$$

$$\begin{bmatrix} Pr(R = r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_1}, n_1), Pr(R = r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_2}, n_1), \cdots, Pr(R = r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_{Q^J}}, n_1) \\ Pr(R = r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_1}, n_2), Pr(R = r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_2}, n_2), \cdots, Pr(R = r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_{Q^J}}, n_2) \\ \cdots \\ Pr(R = r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_1}, n_{Q^J}), Pr(R = r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_2}, n_{Q^J}), \cdots, Pr(R = r_j \text{ is chosen } |\boldsymbol{r}, \boldsymbol{e_{Q^J}}, n_{Q^J}) \end{bmatrix}^{-1} \cdot$$

$$\begin{bmatrix} Pr(R = r_j \text{ is chosen } |\boldsymbol{r}, n_1) \\ Pr(R = r_j \text{ is chosen } |\boldsymbol{r}, n_2) \\ \cdots \\ Pr(R = r_j \text{ is chosen } |\boldsymbol{r}, n_{Q^J}) \end{bmatrix} \tag{4.13}$$

The over-identified case, on the other hand, would be when $Q_J > Q$. In this case, I could always come up a way to combine some distinct values of the number of bidders to reconstruct a similar equation above, but with Q vector on the LHS. The conditional distribution of $\boldsymbol{\epsilon}$ given $\boldsymbol{R} = \boldsymbol{r}$ can still be identified.

Next, I move to the bidder side. First of all, I use the observed distribution of bids within each contest to identify the conditional distribution, $\hat{F}_j(v)$. To do this, I follow the seminal paper by Guerre, Perrigne and Voung (2000), GPV henceforth, and derive an equation linking the observed bids to the unobserved bidder types. Fixing the total number of bidders N and the vector of monetary rewards $\boldsymbol{R}$, and conditional on observing bidder i choosing contest class j, I have:

$$\begin{aligned} v_i^{-1} =&(1-\phi)(R_j+\epsilon_j)(N-1)(1-p_j(1-\hat{G}_j(b))^{N-2}p_j\hat{g}_j(b) \\ &+\phi(R_j+\epsilon_j)(N-1)(N-2)(1-p_j(1-\hat{G}_j(b))^{N-3}p_j(1-\hat{G}_j(b))p_j\hat{g}_j(b) \\ &:=\xi_j(b,R_j,\epsilon_j,N,\phi,p_j) \end{aligned} \tag{4.14}$$

where, $\hat{G}_j(b)$ is the observed cdf of bids in contest class j, and $\hat{g}_j$ is the corresponding pdf function. The choice probability p_j is a known function of $(\boldsymbol{R},\boldsymbol{\epsilon},N)$. I can use this equation to recover the conditional cdf, $\hat{F}_j(v)$, for any contest class j,

$$\begin{aligned} F_j(v|\boldsymbol{R},\boldsymbol{\epsilon},N) &= Pr(V\le v|\text{ choose class j};\boldsymbol{\epsilon},\boldsymbol{R},N) \\ &= Pr(\xi_j(b,R_j,\epsilon_j,N,\phi,p_j(\boldsymbol{R},\boldsymbol{\epsilon},N))\ge\frac{1}{v}) \\ &= Pr(b\le\xi_j^{-1}(\frac{1}{v},R_j,\epsilon_j,N,\phi,p_j(\boldsymbol{R},\boldsymbol{\epsilon},N))) \\ &= \hat{G}_j(\xi_j^{-1}(\frac{1}{v},R_j,\epsilon_j,N,\phi,p_j(\boldsymbol{R},\boldsymbol{\epsilon},N))) \end{aligned} \tag{4.15}$$

In order to recover the unconditional cdf, $F(v)$, I need the following equation which is derived from equation (4.14):

$$f(v|\boldsymbol{R},N)=\sum_{\boldsymbol{\epsilon}}\frac{\partial[p_j(\boldsymbol{R},\boldsymbol{\epsilon},N)\cdot(1-\hat{F}_j(v|\boldsymbol{R},\boldsymbol{\epsilon},N))]}{\partial v}\cdot(-\frac{1}{\pi_j(v|\boldsymbol{R},\boldsymbol{\epsilon},N)})\cdot Pr(\boldsymbol{\epsilon}|\boldsymbol{R}) \tag{4.16}$$

Note that the first part is identified from the previous argument, and $\pi_j(v|\boldsymbol{R},\boldsymbol{\epsilon},N)$

could be computed using the algorithm in the modeling part, and I already identified the distribution of contest types, $Pr(\boldsymbol{\epsilon}|\boldsymbol{R})$.

4.5.2 When Contest Types are Continuous

When ϵ is drawn from a continuous space such as the real line, I can still achieve nonparametric identification under suitable conditions. To see this, suppose in the data, there are many repeated markets where contests offer J distinct values of monetary rewards, $\boldsymbol{R}$, and N bidders. I therefore am able to compute the empirical probability of a contest with monetary rewards $R \in \{r_1, r_2, \cdots, r_{Q_R}\}$ chosen by bidders. I then write down the following equation:

$$\begin{aligned}
&Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, N) \\
=& \int_{\boldsymbol{\epsilon}} Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, \boldsymbol{\epsilon}, N) \cdot f(\boldsymbol{\epsilon}|\boldsymbol{R}, N) d\boldsymbol{\epsilon} \\
=& \int_{\boldsymbol{\epsilon}} Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, \boldsymbol{\epsilon}, N) \cdot f(\boldsymbol{\epsilon}|\boldsymbol{R}) d\boldsymbol{\epsilon} \\
:=& \int_{\boldsymbol{\epsilon}} g(-(\boldsymbol{R} + \boldsymbol{\epsilon}), N) \cdot f(\boldsymbol{\epsilon}|\boldsymbol{R}) d\boldsymbol{\epsilon}
\end{aligned} \tag{4.17}$$

Here, again both $\boldsymbol{R}$ and $\boldsymbol{\epsilon}$ are $1 \times J$ vectors. Let $\boldsymbol{t} := -\boldsymbol{R}$, then the above equation can be written as a convolution:

$$Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, N) := f * g(\boldsymbol{t}) = \int_{\boldsymbol{\epsilon}} g(\boldsymbol{t} - \boldsymbol{\epsilon}) \cdot f(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon} \tag{4.18}$$

In the above equation, I supress $(\boldsymbol{R}, N)$ for the ease of illustration. Assume both f and g are continuous and absolutely integrable with absolute integrable Fourier transform [3], then according to the Convolution Theorem, I can write down the following equation:

$$\mathcal{F}(f * g(\boldsymbol{\xi})) = \mathcal{F}(f(\boldsymbol{\xi})) \times \mathcal{F}(g(\boldsymbol{\xi})) \tag{4.19}$$

where $\mathcal{F}$ denotes the Fourier Transform operation. I could then recover the density function of $\boldsymbol{\epsilon}$ given $\boldsymbol{R}$:

$$\begin{aligned} f(\boldsymbol{\epsilon}|\boldsymbol{R}) =& \mathcal{F}^{-1}\left(\frac{\mathcal{F}(f * g(-\boldsymbol{R}))}{\mathcal{F}(g(-(\boldsymbol{R} + \boldsymbol{\epsilon})))}\right) \\ =& \mathcal{F}^{-1}\left(\frac{\mathcal{F}(Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, N))}{\mathcal{F}(Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, \boldsymbol{\epsilon}, N))}\right) \end{aligned} \tag{4.20}$$

In particular, the numerator can be computed from data, and the denominator can be computed from equilibrium argument in the modeling section. Most importantly, given a fixed number of bidders N, I need to observe the choice probability at any value

[3] Alternatively, I assume the Fourier Transform of f_ϵ is everywhere non-vanishing.

of $\boldsymbol{R}$ to construct the Fourier Transform in the numerator; in other words, variations in monetary prizes provide the primary identification power of contest-wise unobserved heterogeneity. In reality, this full-support condition is difficult to satisfy; instead, I can exploit the variations in the number of bidders like in the discrete case to ensure full identification. To see this, suppose I construct a new function, $\boldsymbol{H}(\boldsymbol{R}, N_1, N_2, \cdots, N_K)$, such that it equals $Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, N_k)$ when we actually observe contests' types to be $\boldsymbol{R}$ and the number of bidders to be N_k in the data; The number K is defined as the number of distinct N's that covers the full support of $\boldsymbol{R}$. Another function $\boldsymbol{G}(-(\boldsymbol{R}+\boldsymbol{\epsilon}), N_1, N_2, \cdots, N_K) := Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, \boldsymbol{\epsilon}, N_k),$ when N_k and $\mathbf{R}$ are jointly observed in the data. can be similarly defined and computed in equilibrium, except that I condition on a certain value of $\boldsymbol{\epsilon}$ now. We then have the following equation:

$$\boldsymbol{H}(\boldsymbol{R}, N_1, N_2, \cdots, N_K) = \int_{\boldsymbol{\epsilon}} \boldsymbol{G}(-(\boldsymbol{R}+\boldsymbol{\epsilon}), N_1, N_2, \cdots, N_K) f(\boldsymbol{\epsilon}|\boldsymbol{R}) d\boldsymbol{\epsilon} \tag{4.21}$$

Then the Convolution Theorem still applies here, except that we substitute the functions to be Fourier Transformed on the RHS of equation (4.20) with the $\boldsymbol{H}$ and $\boldsymbol{G}$ functions respectively:

$$f(\boldsymbol{\epsilon}|\boldsymbol{R}) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\boldsymbol{H}(\boldsymbol{R}, N_1, N_2, \cdots, N_K))}{\mathcal{F}(\boldsymbol{G}(-(\boldsymbol{R}+\boldsymbol{\epsilon}), N_1, N_2, \cdots, N_K))}\right) \tag{4.22}$$

These two functions ensure that for any values of contests' monetary prizes $\boldsymbol{R}$, so long as we observe some markets in the data with the number of bidders to be N_k, we can still achieve identification of contests' unobserved types. Essentially it still requires the full support of $\boldsymbol{R}$ to be observed, but allows the values to be observed with different numbers of bidders in the market; therefore the conditional independence condition in equation (4.12) is also assumed. So far, the distribution of contest-wise unobserved heterogeneity is nonparametrically identified. Identifying bidder-wise unobserved heterogeneity adopts the same strategy as in the discrete case, except that I need to take integration over $\boldsymbol{\epsilon}$ according to their joint distribution.

4.6 Nonparametric Estimation

The estimation procedure consists of two parts: uncovering the bidders' and contests' unobserved type distribution. I start by estimating the distribution of contest-wise unobserved heterogeneity by using the variation in the amount of participating bidders to recover the distribution of contests' unobserved types. This can be directly done using equation 4.13 for the discrete case. For the continuous case, following equation 4.20, I use Monte Carlo method to numerically integrate when estimating the Fourier transform of observed density functions on the RHS. Specifically,

$$\widehat{f(\boldsymbol{\epsilon}|\boldsymbol{R})} = \widehat{\mathcal{F}^{-1}}\left(\frac{\widehat{\mathcal{F}}(Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, N))}{\widehat{\mathcal{F}}(Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, \boldsymbol{\epsilon}, N))}\right) \tag{4.23}$$

where, $\widehat{\mathcal{F}}(Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, N)) := \frac{1}{M}\sum_{m=1}^{M} Pr(\text{a contest with } R = x_{mj} \text{ is chosen}|\boldsymbol{x}_m, N))e^{2\pi i \boldsymbol{x}_m \boldsymbol{R}}$ and $\boldsymbol{x}_m$ is (pseudo-) randomly generated from a multivariate uniform distribution. Similarly, $\widehat{\mathcal{F}}(Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, \boldsymbol{\epsilon}, N)) := \frac{1}{M}\sum_{m=1}^{M} Pr(\text{a contest with } R = x_{mj} \text{ is chosen}|\boldsymbol{x}_m, \boldsymbol{\epsilon}, N)e^{2\pi i \boldsymbol{x}_m(\boldsymbol{R}+\boldsymbol{\epsilon})}$. Lastly, I generate M (pseudo-) random vectors from multivariate uniform distribution, $\boldsymbol{y}_m, m = 1, 2, \cdots, M$, and estimate the inverse Fourier transform as
$\widehat{\mathcal{F}^{-1}} := \frac{1}{M}\sum_{m=1}^{M} \frac{\widehat{\mathcal{F}}(Pr(\text{a contest with } R=r_j \text{ is chosen}|\boldsymbol{R},N))}{\widehat{\mathcal{F}}(Pr(\text{a contest with } R=r_j \text{ is chosen}|\boldsymbol{R},\boldsymbol{y}_m,N))} e^{2\pi i \boldsymbol{\epsilon} \boldsymbol{y}_m}$.

Here, I also show a special case, where ϵ is drawn from a continuous function known up to a finite-dimensional parameter vector $\boldsymbol{\theta}$. In this case, I use a moment-based estimator that relies on various values of N, the number of bidders in the market, to estimate $\boldsymbol{\theta}$. In other words,

$$\begin{aligned}\widehat{\boldsymbol{\theta}} =& argmin_{\boldsymbol{\theta}}(||Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, N) - \\ &\sum_{m=1}^{M} Pr(\text{a contest with } R = r_j \text{ is chosen}|\boldsymbol{R}, \boldsymbol{\epsilon}_{m,\boldsymbol{\theta}}, N)||)\end{aligned} \tag{4.24}$$

where $\boldsymbol{\epsilon}_{m,\boldsymbol{\theta}}$ is a (pseudo-) randomly generated vector with a density function $f(\cdot|\boldsymbol{R}, \boldsymbol{\theta})$. So long as there are as many different values of N as the dimensionality of $\boldsymbol{\theta}$, the distribution of $\boldsymbol{\epsilon}$ can be estimated.

After recovering the conditional distribution of contests' unobserved types, I show the estimation of the distribution of bidders' value function. Following GPV, a well-known two-step estimation strategy is used: in the first step, I estimate the distribution of observed bids for each class of contests using kernel methods; after that, I recover the pseudo-private value for each bidder using the following equation, then use the kernel method to estimate the conditional distribution of bidders' value functions within each class of contests. Formally, in the first step, suppose a set of bidders I_{jt} choose contests of class j in market t, and I observe their bids $\{B_{it}, i \in I_{jt}, t = 1, 2, \cdots, T\}$, then the empirical distribution function $\hat{G}$ and the kernel density estimator $\hat{g}$ for contest class j are estimated as:

$$\hat{G}_j(b) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_{jt}} \sum_{i \in I_{jt}} \mathbb{1}(B_{it} \leq b) \tag{4.25}$$

$$\hat{g}_j(b) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N_{jt} h_{jg}} \sum_{i \in I_{jt}} K_{jg}\left(\frac{b - B_{jt}}{h_{jg}}\right) \tag{4.26}$$

where h_{jg} is a bandwidth that may vary across contest classes and $K_{jg}(\cdot)$ is a kernel function on a compact support. One thing to mention here is that, by solving the functional form of the equilibrium bidding strategy within each auction, I know that the underlying density of equilibrium bids is infinite at its loIr bound. Furthermore, it is also Ill known that on its boundaries, the kernel density estimator is biased. Therefore, kernel methods could be quite inaccurate near the loIr bound. I hence trim the pseudo-private value using the following criteria. In particular, bidder i's pseudo-private value is estimated through

$$\widehat{v_i} = \begin{cases} \sum_{\boldsymbol{\epsilon}}\{(1-\phi)(R_j+\epsilon_j)(N-1)(1-\widehat{p_j}(\boldsymbol{R},N)(1-\hat{G}_j(b))^{N-2}\widehat{p_j}(\boldsymbol{R},N)\hat{g}_j(b) \\ +\phi(R_j+\epsilon_j)(N-1)(N-2)(1-\widehat{p_j}(\boldsymbol{R},N)(1-\hat{G}_j(b))^{N-3}\widehat{p_j}(\boldsymbol{R},N)\cdot \\ (1-\hat{G}_j(b))\widehat{p_j}(\boldsymbol{R},N)\hat{g}_j(b)\}^{-1}\cdot Pr(\boldsymbol{\epsilon}|\boldsymbol{R}) \text{ , if } B_{min}+\rho_{jg}h_{jg}/2 \leq B_{it} \leq B_max-\rho_{jg}h_{jg}/2 \\ +\infty \text{ , otherwise} \end{cases}$$

when contest-wise types, $\boldsymbol{\epsilon}$'s, are discrete and ρ_{jg} denotes the length of support of the kernel function used for contest class j. In the equation above, the estimated function $\widehat{p_j}(\boldsymbol{R},N) := \sum_{k=1}^{Q} p_j(\boldsymbol{R},\boldsymbol{\epsilon}_k,N)Pr(\boldsymbol{\epsilon}=\boldsymbol{\epsilon}_k)$. When $\boldsymbol{\epsilon}$'s are continuous, it is specified as:

$$\widehat{v_i} = \begin{cases} \int_{\boldsymbol{\epsilon}}\{(1-\phi)(R_j+\epsilon_j)(N-1)(1-\widehat{p_j}(\boldsymbol{R},N)(1-\hat{G}_j(b))^{N-2}p_j(\boldsymbol{R},N)\hat{g}_j(b) \\ +\phi(R_j+\epsilon_j)(N-1)(N-2)(1-\widehat{p_j}(\boldsymbol{R},N)(1-\hat{G}_j(b))^{N-3}\widehat{p_j}(\boldsymbol{R},N)\cdot \\ (1-\hat{G}_j(b))\widehat{p_j}(\boldsymbol{R},N)\hat{g}_j(b)\}^{-1}\cdot dF(\boldsymbol{\epsilon}|\boldsymbol{R}) \text{ , if } B_{min}+\rho_{jg}h_{jg}/2 \leq B_{it} \leq B_max-\rho_{jg}h_{jg}/2 \\ +\infty \text{ , otherwise} \end{cases}$$

and the estimated function $\widehat{p_j}(\boldsymbol{R},N) := \int_{\boldsymbol{\epsilon}} p_j(\boldsymbol{R},\boldsymbol{\epsilon},N)d\widehat{F}(\boldsymbol{\epsilon}|\boldsymbol{R})$. In practice, this integration can be done numerically if not having a closed-form solution. In the following section, I show how this estimation procedure performs using Monte Carlo simulated data.

4.7 Monte Carlo Evidence

4.7.1 When Contest Types are Discrete

In this Monte Carlo simulation analysis, I generate 100 markets with i.i.d. distributed 2, 3, 4 or 5 bidders and i.i.d. distributed two contests in each market. As discussed in the identification argument, this is to create variation in number of contests within each class for the first-stage estimation. Bidders' types are drawn from a truncated standard lognormal distribution on $[0.1, 4]$. Further, I let the monetary reward take only two distinct values, $\{5, 3\}$. For simplicity, I further assume that there is only one project having monetary reward equal to 5 and one project having monetary reward equal to 3 in each market. The contest complexity is drawn from a binary support of $\{-0.1, 0.1\}$, and the conditional probability of ϵ given monetary reward R is given by:

$$Pr(\epsilon = 0.1|R) = \begin{cases} 0.9, & \text{if } R = 5 \\ 0.1, & \text{if } R = 3 \end{cases} \tag{4.27}$$

This also suggests that the ϵ's across different contests are independent of each other. This assumption is only for computational simplicity. I use triangular Kernel function, and the bandwidth is set same as in GPV. Figure 4.1 and Table 4.1 are the results from two steps of the estimation with 100 repetitions. The starting values are set to be zeros. It can be seen that, for the estimation of contest-wise complexity distribution, the overall bias is small, despite the relatively high variation as shown by the standard deviations. For the estimation of bidders' valuation function, it can be shown that on the interval of $[0.8, 3.3]$, the estimation is relatively good. On the lower and higher bounds, however, it is less accurate due to the inaccurate kernel estimation of bid density near boundaries.

4.7.2 When Contest Types are Continuous

Now I present the performance when the contest unobserved types are continuous. I only show results estimating parametric distribution for contest-wise complexity levels, and nonparametric distribution for bidders' valuation, as fully nonparametrically estimating the whole model may be computationally burdensome and require a lot more variations in the simulated data. I generate 100 i.i.d. markets with 2, 3, 4, 5 or 6 bidders in each market. Bidders' valuation is again drawn from a truncated standard lognormal distribution on $[0.1, 4]$. Further, I let the monetary reward take only two distinct values, $\{5, 3\}$. For simplicity, I further assume that there is only one project having a monetary reward equal to 5 and one project having a monetary reward equal to 3 in each market. The contest complexity is drawn from a normal distribution, which parameters are given below. The conditional mean is:

$$\mu_{\epsilon|R} = \begin{cases} 1, & \text{if } R = 5 \\ \\ -1, & \text{if } R = 3 \end{cases} \tag{4.28}$$

and the conditional variance is:

$$\sigma_{\epsilon|R} = \begin{cases} 1, & \text{if } R = 5 \\ \\ 2, & \text{if } R = 3 \end{cases} \tag{4.29}$$

I estimated the parameters using two specifications. First I fix the variances of ϵ's and only estimated the conditional means, $\mu_{\epsilon|R=5}$ and $\mu_{\epsilon|R=3}$. The results are shown in the first panel of Table 4.2. Then I estimated all four parameters of ϵ, including $\sigma_{\epsilon|R=5}$ and $\sigma_{\epsilon|R=3}$. The starting values are set to be 0.9 times the true values in the first specification and the true values in the second specification. To estimate the parameters for contest complexity levels, I adopt the Monte Carlo numerical integration method and generate 200 random samples to approximate the integrated value. It can be seen that there is more noises and inaccuracy regarding estimating the contest-wise complexity levels, compared with the discrete case, as I see higher biases and larger variation in Table 4.2. This inaccuracy increases when I try to estimate more parameters in the model. One conjecture is that, when trying to estimate the distribution of contest-wise complexity nonparametrically, I must need a much richer variation in the number of bidders appeared in each market.

To sum, it can be seen that compared with the continuous case, it is more accurate to estimate the contest-wise unobserved heterogeneity in the discrete space. This is mainly due to the data availability and the bias arisen from numerical integration.

4.8 Conclusion

This paper develops a new method to identify and estimate primitives in simultaneous contests with multiple prizes. In theoretical modeling part, I establish a two-stage game where contestants first choose one among multiple contests, then in the second stage, they compete within each contest by submitting their bids or contributing their efforts to win over the pre-determined prize based on the rank order. Non-winners may get part of the prize due to a pre-determined rule. I show that by jointly observing their first-stage choice probability and the second-stage bidding strategy, I can nonparametrically identify the joint distribution of unobserved heterogeneity on both sides of the market. I then present a corresponding estimation strategy and show the performance of Monte Carlo experiments.

While this novel strategy can be potentially applied to many real-life scenarios from political lobbying to online labor markets, I see at least three directions for future work to suit more complicated markets better. First, it is widely observed that bidders/contestants incur nonrefundable cost when bidding in the second stage of the game. Hence, how to incorporate this bidding cost as another layer of contest-wise and/or bidder-wise unobserved heterogeneity is important to study. The main complication here is the lack of a closed-form solution to characterize players' equilibrium strategies in the presence of the bidding cost, except that I know there will be a minimum level of bidders' skills to enter a particular contest class. Another direction is to introduce supermodularity or submodularity between bidders' skills and contests' types. For instance, some bidders may encounter synergy effect by participating in certain classes of contests. It would be fascinating to see how market outcomes are affected by this synergy effect, and most importantly, how the market designer could stimulate a better participation pattern using alternative policy intervention. Lastly, as contest holders stand for the other side of the market, it would be substantial to explicitly study their equilibrium behavior by incorporating a pricing stage before the contestants' movements. This is also related to the other two chapters of this thesis but requires more work regarding proving the existence of a reasonable equilibrium notion.

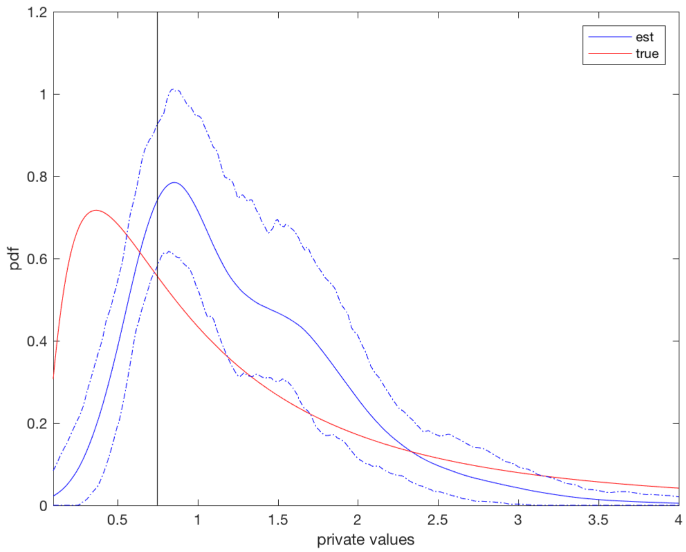


Figure 4.1: Kernel Density Fitting of Bidder's Value Function – Discrete Case

Table 4.1: Estimation of the Distribution of Contest-wise Unobserved Heterogeneity – Discrete Case

	starting value	true value	bias	std.dev
$Pr(\epsilon = 1 \mid R = 1.1)$	0.9	0	-0.0303	0.3072
$Pr(\epsilon = 1 \mid R = 1)$	0.1	0	-0.0150	0.3144

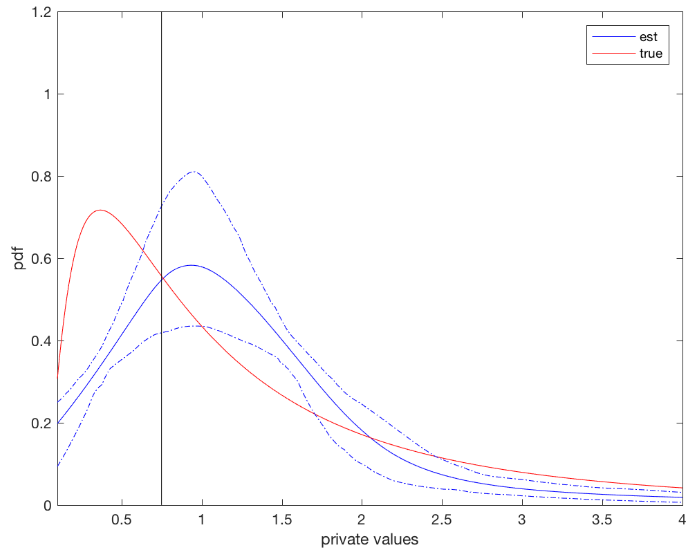


Figure 4.2: Kernel Density Fitting of Bidder's Value Function – Continuous Case (first specification)

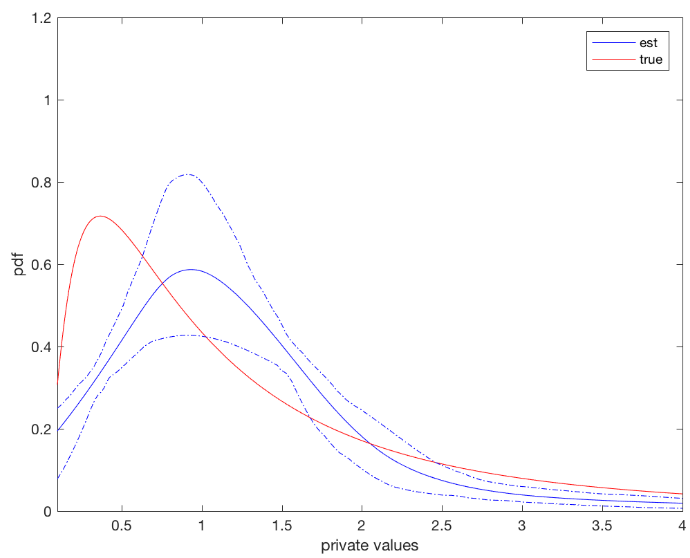


Figure 4.3: Kernel Density Fitting of Bidder's Value Function – Continuous Case (second specification)

Table 4.2: Estimation of the Distribution of Contest-wise Unobserved Heterogeneity – Continuous Case

	starting value	true value	bias	std.dev
First set of parameters				
$\mu_{\epsilon\|R=5}$	1	0.9	-0.1537	0.2135
$\mu_{\epsilon\|R=3}$	-1	-0.9	0.1532	0.2063
Second set of parameters				
$\mu_{\epsilon\|R=5}$	1	1	-0.1925	0.4180
$\mu_{\epsilon\|R=3}$	-1	-1	0.2365	0.4351
$\sigma_{\epsilon\|R=5}$	1	1	0.2534	0.7427
$\sigma_{\epsilon\|R=3}$	2	2	-0.2337	0.6918

Bibliography

[1] ACEMOGLU, D., MOSTAGIR, M., AND OZDAGLAR, A. Managing innovation in a crowd. Tech. rep., National Bureau of Economic Research, 2014.

[2] AGARWAL, N. An empirical model of the medical match. *NBER Working Paper*, w20767 (2014).

[3] AGARWAL, N., AND DIAMOND, W. Identification and estimation in two-sided matching markets.

[4] AN, Y., HU, Y., AND SHUM, M. Estimating first-price auctions with an unknown number of bidders: A misclassification approach. *Journal of Econometrics 157*, 2 (2010), 328–341.

[5] BACCARA, M., İMROHOROĞLU, A., WILSON, A. J., AND YARIV, L. A field study on matching with network externalities. *The American Economic Review* (2012), 1773–1804.

[6] Bajari, P., and Hortacsu, A. Economic insights from internet auctions: A survey. Tech. rep., National Bureau of Economic Research, 2003.

[7] Baye, M. R., Kovenock, D., and De Vries, C. G. The all-pay auction with complete information. *Economic Theory 8*, 2 (1996), 291–305.

[8] Berry, S. T., and Haile, P. A. Identification in differentiated products markets using market level data. Tech. rep., National Bureau of Economic Research, 2010.

[9] Bolton, G., Greiner, B., and Ockenfels, A. Engineering trust: reciprocity in the production of reputation information. *Management Science 59*, 2 (2013), 265–285.

[10] Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics 31*, 1 (2013), 83–117.

[11] Burdett, K., Shi, S., and Wright, R. Pricing and matching with frictions. *Journal of Political Economy 109*, 5 (2001), 1060–1085.

[12] Cantillon, E., Pesendorfer, M., et al. Auctioning bus routes: The london experience.

[13] Chiappori, P.-A., Iyigun, M., and Weiss, Y. Investment in schooling and the marriage market.

[14] Chiappori, P.-A., Oreffice, S., and Quintana-Domeque, C. Fatter attraction: anthropometric and socioeconomic matching on the marriage market. *Journal of Political Economy 120*, 4 (2012), 659–695.

[15] Chiappori, P. A., Salanié, B., and Weiss, Y. Partner choice and the marital college premium.

[16] Chiappori, P.-A., and Weiss, Y. Divorce, remarriage, and welfare: A general equilibrium approach. *Journal of the European Economic Association 4*, 2-3 (2006), 415–426.

[17] Choo, E., and Siow, A. Who marries whom and why. *Journal of political Economy 114*, 1 (2006), 175–201.

[18] Clark, S. The uniqueness of stable matchings. *Contributions in Theoretical Economics 6*, 1 (2006), 1–28.

[19] Cullen, Z., and Farronato, C. Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms.

[20] Dinerstein, M., Einav, L., Levin, J., and Sundaresan, N. Consumer price search and platform design in internet commerce. Tech. rep., National Bureau of Economic Research, 2014.

[21] DiPalantino, D., and Vojnovic, M. Crowdsourcing and all-pay auctions. In *Proceedings of the 10th ACM conference on Electronic commerce* (2009), ACM, pp. 119–128.

[22] Einav, L., Farronato, C., and Levin, J. Peer-to-peer markets. Tech. rep., National Bureau of Economic Research, 2015.

[23] Einav, L., Kuchler, T., Levin, J., and Sundaresan, N. Assessing sale strategies in online markets using matched listings. *American Economic Journal: Microeconomics 7*, 2 (2015), 215–247.

[24] Fox, J. T. Identification in matching games. *Quantitative Economics 1*, 2 (2010), 203–254.

[25] Fox, J. T., and Gandhi, A. Nonparametric identification and estimation of random coefficients in nonlinear economic models. Tech. rep., Citeseer, 2010.

[26] Fox, J. T., il Kim, K., Ryan, S. P., and Bajari, P. The random coefficients logit model is identified. *Journal of Econometrics 166*, 2 (2012), 204–212.

[27] Fox, J. T., and Yang, C. Unobserved heterogeneity in matching games. Tech. rep., National Bureau of Economic Research, 2012.

[28] Fradkin, A. Search frictions and the design of online marketplaces. Tech. rep., Tech. rep., Working paper, Stanford University, 2013.

[29] Galichon, A., Kominers, S. D., and Weber, S. An empirical framework for matching with imperfectly transferable utility. *Available at SSRN 2535394* (2014).

[30] Galichon, A., and Salanié, B. Cupids invisible hand: Social surplus and identification in matching models. *Available at SSRN 1804623* (2012).

[31] Gentry, M. L., Komarova, T., and Schiraldi, P. Simultaneous first-price auctions with preferences over combinations: Identification, estimation and application. *Estimation and Application (January 2016)* (2016).

[32] Graham, B. S. Comparative static and computational methods for an empirical one-to-one transferable utility matching model. *Advances in Econometrics 31* (2013), 153–181.

[33] Graham, B. S. Errata in econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers.

[34] Guerre, E., Perrigne, I., and Vuong, Q. Optimal nonparametric estimation of first-price auctions. *Econometrica 68*, 3 (2000), 525–574.

[35] Harsanyi, J. C. Games with incomplete information played by" bayesian" players, i-iii. part i. the basic model. *Management Science* (1967), 159–182.

[36] Horton, J. J. Misdirected search effort in a matching market: Causes, consequences and a partial solution. In *Proceedings of the fifteenth ACM conference on Economics and computation* (2014), ACM, pp. 357–357.

[37] Horton, J. J. Supply constraints as a market friction: Evidence from an online labor market.

[38] Horton, J. J., and Chilton, L. B. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on Electronic commerce* (2010), ACM, pp. 209–218.

[39] Howe, J. The rise of crowdsourcing. *Wired magazine 14*, 6 (2006), 1–4.

[40] Hsieh, Y. Understanding mate preferences from two-sided matching markets: Identification, estimation and policy analysis. Tech. rep., Working Paper, University of Southern California, 2012.

[41] Hu, Y. Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics 144*, 1 (2008), 27–61.

[42] Hu, Y., and Schennach, S. M. Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*, 1 (2008), 195–216.

[43] Kim, S. W., Olivares, M., and Weintraub, G. Y. Measuring the performance of large-scale combinatorial auctions: A structural estimation approach. *Management Science 60*, 5 (2014), 1180–1201.

[44] Kvasov, D. Contests with limited resources. *Journal of Economic Theory 136*, 1 (2007), 738–748.

[45] Lewbel, A. Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics 97*, 1 (2000), 145–177.

[46] Logan, J. A., Hoff, P. D., and Newton, M. A. Two-sided estimation of mate preferences for similarities in age, education, and religion. *Journal of the American Statistical Association 103*, 482 (2008), 559–569.

[47] Matzkin, R. L. Heterogeneous choice. *Econometric Society Monographs 43* (2007), 75.

[48] Menzel, K. Large matching markets as two-sided demand systems. Tech. rep., working paper, 2013.

[49] Milgrom, P. R., and Weber, R. J. A theory of auctions and competitive bidding. *Econometrica: Journal of the Econometric Society* (1982), 1089–1122.

[50] Milgrom, P. R., and Weber, R. J. Distributional strategies for games with incomplete information. *Mathematics of Operations Research 10*, 4 (1985), 619–632.

[51] Moldovanu, B., and Sela, A. The optimal allocation of prizes in contests. *American Economic Review* (2001), 542–558.

[52] Moldovanu, B., and Sela, A. Contest architecture. *Journal of Economic Theory 126*, 1 (2006), 70–96.

[53] Ni, J., and Srinivasan, K. Matching in the sourcing market: A structural analysis of the upstream channel. *Marketing Science 34*, 5 (2015), 722–738.

[54] Niederle, M., and Yariv, L. Decentralized matching with aligned preferences. Tech. rep., National Bureau of Economic Research, 2009.

[55] Nosko, C., and Tadelis, S. The limits of reputation in platform markets: An empirical analysis and field experiment. Tech. rep., National Bureau of Economic Research, 2015.

[56] Resnick, P., and Zeckhauser, R. Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. *The Economics of the Internet and E-commerce 11*, 2 (2002), 23–25.

[57] Roth, A. E. On the allocation of residents to rural hospitals: a general property of two-sided matching markets. *Econometrica: Journal of the Econometric Society* (1986), 425–427.

[58] Roth, A. E., and Sotomayor, M. The college admissions problem revisited. *Econometrica: Journal of the Econometric Society* (1989), 559–570.

[59] Rothkopf, M. H., Pekeč, A., and Harstad, R. M. Computationally manageable combinational auctions. *Management science 44*, 8 (1998), 1131–1147.

[60] Sandholm, T., Suri, S., Gilpin, A., and Levine, D. Cabob: A fast optimal algorithm for combinatorial auctions. In *International Joint Conference on Artificial Intelligence* (2001), vol. 17, LAWRENCE ERLBAUM ASSOCIATES LTD, pp. 1102–1108.

[61] Shapley, L., and Shubik, M. *Game Theory in Economics: Chapter 3: the" rules of the Games"*. Rand, 1972.

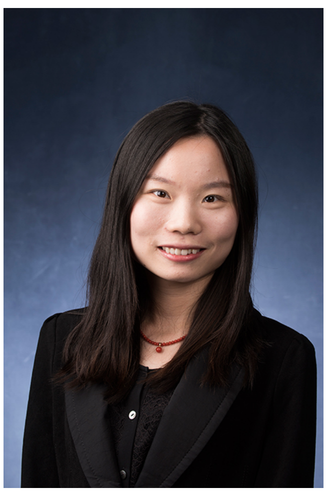
[62] Sinha, S. Identification and estimation in one-to-one matching models with nonparametric unobservables. Tech. rep., mimeo, Northwestern University, 2014.

[63] SØRENSEN, M. How smart is smart money? a two-sided matching model of venture capital. *The Journal of Finance 62*, 6 (2007), 2725–2762.

[64] Train, K. E. Em algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling 1*, 1 (2008), 40–69.

[65] Uetake, K., and Watanabe, Y. Entry by merger: Estimates from a two-sided matching model with externalities. *Available at SSRN 2188581* (2014).

Vita



Yajing Jiang was born in Shandong Province, China on April 28, 1989. She received her B. Sc. degrees in Finance and Statistics from Peking University, Beijing, China in 2010, and enrolled in the Economics Ph.D. program at Johns Hopkins University in the same year. Her research interest is empirical industrial organization, micro-econometrics and applied micro-economics. Starting in August 2016, Yajing works as a Senior Associate in the Antitrust and Competition practice of Charles River Associate International Inc., where she performs economic and econometric analyses for professional clients to quantify the competitive effects on both firm- and market-level.