IDENTIFICATION AND ESTIMATION IN TWO-SIDED MATCHING MARKETS

By

Nikhil Agarwal and William Diamond

August 2013
Revised February 2014

COWLES FOUNDATION DISCUSSION PAPER NO. 1905

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Identification and Estimation in Two-Sided Matching Markets*

Nikhil Agarwal William Diamond

February 27, 2014

Abstract

We study estimation and non-parametric identification of preferences in two-sided matching markets using data from a single market with many agents. We consider a model in which preferences of each side of the market are vertical, utility is non-transferable and the observed matches are pairwise stable. We show that preferences are not identified with data on one-to-one matches but are non-parametrically identified when data from many-to-one matches are observed. The additional empirical content in many-to-one matches is illustrated by comparing two simulated objective functions, one that does and the other that does not use information available in many-to-one matching. We also prove consistency of a method of moments estimator for a parametric model under a data generating process in which the size of the matching market increases, but data only on one market is observed. Since matches in a single market are interdependent, our proof of consistency cannot rely on observations of independent matches. Finally, we present Monte Carlo studies of a simulation based estimator.

*We thank Susan Athey, Gary Chamberlain, Victor Chernozhukov, Raj Chetty, Kirill Evdokimov, Jeremy Fox, Phil Haile, Whitney Newey, Ariel Pakes and Parag Pathak for helpful discussions. Agarwal: Cowles Foundation, Yale University and Department of Economics, Massachusetts Institute of Technology, email: agarwaln@mit.edu. Diamond: Department of Economics, Harvard University, email: wdiamond@hbs.edu.

1 Introduction

There has been growing interest in estimating preferences of agents in two-sided matching markets.[1] Since preferences determine outcomes in these markets, estimation these primitives is an important step for quantitatively evaluating economic questions such as counterfactual analysis of equilibrium effects of policy interventions or market structure. However, a researcher often has access only to data on final matches instead of stated preferences. In these cases, one approach is to use an equilibrium assumption on the observed matches to infer preferences of agents in the market as a function of agents characteristics, some of which may not be observed.

We consider estimation and non-parametric identification of preferences using a model in which utility is not transferable and the observed match is assumed to be pairwise stable. A non-transferable utility (NTU) model is an attractive assumption for markets where transfers are not individually negotiated or are prohibited. Examples include the matching of students to public schools or universities with free tuition or exogenous financial aid criteria; labor or apprenticeship markets with inflexible wages such as the medical residency market or a public-sector labor market with pre-determined wage scales. In these settings, preferences of agents on both sides of the market determine equilibrium outcomes: a student cannot simply show up to class in a college since she first needs to be admitted, and conversely, colleges compete to enroll the best students. Further, non-transferable utility may be a good assumption if financial aid is based on student characteristics and income criteria that are exogenous to the other colleges that admit that student.

Separately estimating preferences of agents on both sides of the market may be necessary for understanding the determinants of observed outcomes, and for analyzing certain economic questions. For instance, quanititatively evaluating the trade-off for firms between experience and education of the worker, the trade-off for workers between wages and the value of ammenities such as on-the-job training, or the welfare effects of alternative market

[1] See Fox (2009) for a survey.

structures can require separate preference estimates of agents on both sides of the market. Identifying preferences of agents on both sides of the market may be a challenging exercise because equilibrium matches are jointly determined by both sets of preferences: when we see a student enrolling at a particular college, it need not be the case that the college is her most preferred option because she may have not been accepted at her more preferred institutions.

We analyze the estimation and identification problem in a data environment in which a researcher observes matches in a single large market. This modelling choice is made for several reasons. First, data from several matching markets with the same underlying structure are rare compared to data from a few markets with many agents. For example, public high school markets, colleges, the medical residency market and marriage markets have several thousand participating agents. Second, recent papers in the theoretical matching literature have utilized large market approximations for analyzing strategic behaviour and the structure of equilibria.[2] Third, it highlights important interdependence between matches within a market in the asymptotic analysis of our estimator.

The main restriction in our model is that preferences on each side of the market are vertical, i.e., all agents agree on the relative ranking of any two agents. However, we allow for agents to have unobservable characteristics that may affect their desirability to the other side of the market. Therefore, two observationally identical agents need not be equally desirable. The model may provide an appropriate approximation for studying a labor market in which workers that have approximately homogeneous tastes for job-ammenities and firms value components of worker skill similarly. While the restriction that agents have homogeneous tastes is strong, the assumption is made for tractability given the well-known difficulty of expressing pairwise stable matches as a function of the preferences in a more general model. The assumption makes the model tractable because it implies that pairwise stable matches are perfectly assortative on the underlying desirabilities of the agents irrespective of their distributions. This simplification allows us to obtain identification results without making

[2] See Immorlica and Mahdian (2005); Kojima and Pathak (2009); Azevedo and Budish (2012), for example.

strong parametric assumptions on the latent utilities that represent the preferences of agents.

Even with this stark restriction on preferences, our first result on identification is negative. We show that the distribution of preferences on both sides of the market are not identified for a large class of models when only data from one-to-one matches are observed. Further, we compute an example using a family of parametric models of one-to-one matching that are observationally equivalent. The example highlights that our non-identification result is not a result of excessively weak conditions on the model. Intuitively, the non-identification arises because unobservable characteristics of either side of the market could be driving the observed joint distribution of agents and their match partners. These results imply limitations on what can be learned with data using data on one-to-one matches, and guide the use of empirical techniques. For instance, they weigh against estimating non-transferable utility models in marriage markets with data from a single market.

In contrast, we show that the distribution of preferences of agents on both sides of the market is non-parametrically identified from data on many-to-one matches. The result requires that each agent on one-side of the market is matched to at least two agents on the other side, a requirement that is likely satisfied in many education and labor markets. To the best of our knowledge, this is the first positive result on non-parametric identification of preferences of agents on both sides of the market in a non-transferable utility setting. Particularly, the difference between the empirical content of one-to-one matching and many-to-one matching has not been previously exploited to obtain identification results in our setting. Our proof is based on interpreting the matching model with two-to-one matches in terms of a measurement error model. This reinterpretation makes the additional empirical content of many-to-one matches is ex-post intuitive: the observable components of quality of a worker can provide a noisy measure of the overall quality of her colleagues.

We also illustrate the additional identifying information available in many-to-one matches using simulations and a parametrized family of models. Our simulations suggest that moments that only use information available in sorting patterns are not able to distinguish between a large set of parameter values. In the context of one-to-one matching, this is the only information observed in the dataset. In contrast, simulations also suggest that additional moments constructed from many-to-one matching can be used to distinguish parameter values that yield indistinguishable sorting patterns. An objective function constructed from both sets of moments has a global minimum near the true parameter. We therefore recommend using information from many-to-one matching, when available.

We then study asymptotic properties of a method of moments estimator for a parametric model based on a criterion function that uses moments from many-to-one matching as well as sorting patterns. The main limit theorem proves that under identifiable uniqueness of a parametric model, data on a single large market with many-to-one matches can be used to consistently estimate the true parameter. For simplicity, we restrict to the case with two-to-one matching. Our proof directly uses the properties of the equilibrium matches in which we account for interdependence between observations. Finally, we use Monte Carlo simulations to study the property of a simulation based estimator.

Related Literature: Most of the recent literature on identification and estimation of matching games studies the transferable utility (TU) model in which the equilibrium describes matches as well as the surplus split between the agents with quasi-linear preferences for money (Choo and Siow, 2006; Sorensen, 2007; Fox, 2010a; Gordon and Knight, 2009; Galichon and Salanie, 2010; Chiappori et al., 2011, among others). The equilibrium transfers imply that no two unmatched agents can find a profitable transfer in which they would like to match with each other. The typical goal in these studies is to recover a single aggregate surplus which determines the equilibrium matches. Often, these studies face a data constraint that monetary transfers between matched partners are not observed, so the hope of estimating two separate utility functions is limited. A branch of this literature, following the work of Choo and Siow (2006) proposes identification and estimation of a transferable utility model based on the assumption that individuals have unobserved tastes for observed groups of individuals. Agents in these models are indifferent between match partners of the same observable type, implying that the econometrician observes all agent characteristics that are valued by the other side of the market. Using this assumption, the papers propose estimation and identification of group-specific surplus functions (Choo and Siow, 2006; Galichon and Salanie, 2010; Chiappori et al., 2011). In our model, agents have homogeneous tastes for match partners, but agents are not indifferent between two partners with identical observables because of unobserved characteristics. A different approach to identification in transferable utility models, due to Fox (2010a), is based on assuming that the structural unobservables are such that the probability of observing a particular match is higher if the total systematic, observable component of utility is larger than an alternative match. In non-transferable utility models, such as ours, this approach is difficult to motivate since pairwise stable matchings need not maximize total surplus.

In many applications, inflexible monetary transfers or counterfactual analysis that require estimates of preferences for agents on both sides of the market motivate the use of a non-transferable utility model. A large theoretical literature (c.f. Roth and Sotomayor, 1992) has been devoted to the study of properties of stable matchings in NTU models. To our knowledge, however, there are only two previous papers that have formally considered identification and estimation in this setting. Both papers are complementary to our analysis and show partial identification of preferences. Hsieh (2011) follows Choo and Siow (2006) in assuming that agents belong to finitely many observed groups and that agents have idiosyncratic tastes for these groups. The main identification result in Hsieh (2011) shows that the model can rationalize any distribution of matchings in this setting, implying that the identified set is non-empty. Menzel (2013) studies identification and estimation in a non-transferable utility model in a large market where agent preferences are heterogeneous due to idiosyncratic match-specific tastes with a distribution in the domain of attraction of the Generalized Extreme Value (GEV) family, and observable characteristics have bounded support. By using the tail properties of the GEV class, Menzel (2013) finds that only the sum of the surplus of both sides obtained from matching is idenfied from data on one-to-one matching. The result is similar in spirit to our negative result on identification with one-to-one matches. While these papers focus on the one-to-one matching case, our results exploit data on many-to-one matches to non-parametrically identify preferences of both sides of the market, although at the cost of assuming homogeneous preferences.

There are now a few papers that have estimated non-transferable utility models. Agarwal (2013) estimates preferences in the market for medical residents using a method of simulated moments estimator based on moments that exploit information in many-to-one matching in addition to the sorting of resident and program characteristics. Logan et al. (2008) proposes a Bayesian technique for estimating the posterior distribution of preference parameters in marriage markets, and Boyd et al. (2013) estimates the preferences of teachers for schools and schools for teachers using a method of moments estimator. Both Logan et al. (2008) and Boyd et al. (2013) use only the sorting of observed characteristics of agents as given by the matches (sorting patterns) to recover primitives. Our result on non-identification with one-to-one matching implies that point estimates obtained with these approaches may be sensitive to parametric assumptions.

A few empirical papers estimate sets of preference parameters that are consistent with pairwise stability (Menzel, 2011; Uetake and Watanabe, 2012). The concern that preferences need not be point identified with one-to-one matches does not necessarily apply to these approaches. For example, Menzel (2011) uses two-sided matching to illustrate a Bayesian approach for estimating a set of parameters consistent with an incomplete structural model. Our results on non-identification and subsequent simulations that use information on sorting patterns suggest that a rather large set of parameters are observationally equivalent. While these results imply that the identified set may be large, these approaches may still be quite informative for certain questions of interest.

Our finding that data from many-to-one matching is important in identification is related to work by Fox (2010a) and Fox (2010b) on many-to-many matching. In these papers, many-to-many matching games allow identification of certain features of the observable component of the surplus function when agents share some but not all partners. This allows differencing the surplus generated from common match partners to learn valuations. In our setting, many-to-one matching plays a different role in that it allows us to learn about the extent to which unobservable characteristics of each side of the market drive the observed patterns. Finally, our results on identification with many-to-one matching are based on techniques for identifying non-linear measurement error models developed in Hu and Schennach (2008). These techniques have been applied to identify auction models with unobserved heterogeneity (Hu et al., 2013), and dynamic models with unobserved states (Hu and Shum, 2012). To our knowledge, these techniques have not been previously used to identify matching models.

Section 2 presents the model, Section 3 discusses identification, Section 4 discusses estimation results and Section 5 presents Monte Carlo results. All proofs are in the Appendix.

2 Model

We will consider a two-sided matching market with non-transferable utility. The two sides will be referred to as workers and firms, with individual agents indexed by i and j respectively. For simplicity, we assume that the total number of positions at firms equals the total number of workers. Additional firms could be introduced in order to capture unmatched workers.

2.1 Market Participants

The participants in the market are described by a pair of probability measures $m^e = (m^{x,\varepsilon}, m^{z,\eta})$. Here, $m^{x,\varepsilon}$ is the joint distribution of observable traits $x \in \chi \subseteq \mathbb{R}^{k_x}$ and unobsevable traits $\varepsilon \in \mathbb{R}$ for the workers. Likewise, $m^{z,\eta}$ is the joint distribution of observable traits $z \in \zeta \subseteq \mathbb{R}^{k_z} \times \mathbb{N}$ and unobservable traits $\eta \in \mathbb{R}$ for the firms.

In an economy with n agents $((X_1, \varepsilon_1), \ldots, (X_n, \varepsilon_n))$ and $((Z_1, \eta_1), \ldots, (Z_n, \eta_n))$ on each side, the measures will be of the form $m_n^{x,\varepsilon} = \frac{1}{n}\sum_{i=1}^n \delta_{(X_i,\varepsilon_i)}$ and $m_n^{z,\eta} = \frac{1}{n}\sum_{j=1}^n \delta_{\left(Z_j,\eta_j\right)}$ where δ_Y is the dirac delta measure at Y.

Assumption 1 *The population measures $m^{x,\varepsilon}$ and $m^{z,\eta}$ are such that*

(i) m^{η} and m^{ε} admit densities with full support on $\mathbb{R}$, and are absolutely continuous with respect to Lebesgue measure.

(ii) η and ε are independent of X and Z respectively

Assumption 1 (i) imposes a regularity condition on the support and distributions of the unobservables and Assumption 1 (ii) assumes independence. On its own, independence is not particularly strong, but a restriction on preferences to follow will make this a strong assumption.

2.2 Preferences

Each side of the economy has a utility function over observable and unobservable traits of the other side of the economy. That is, worker i's human capital index is given by the additively separable form

$$v\left(x_i, \varepsilon_i\right) = h\left(x_i\right) + \varepsilon_i. \tag{1}$$

Additive separability in the unobservables implies that the marginal value of observable traits does not depend on the unobservable. We also assume that the firm derives a value v_i from each worker that is separable from the other workers matched to it.

Likewise, the preference of workers for firm j is given by

$$u\left(z_j, \eta_j\right) = g\left(z_j\right) + \eta_j. \tag{2}$$

In addition to homogeneity of preferences, additivity of $v\left(x_i, \varepsilon_i\right)$ in ε_i and of $u\left(z_j, \eta_j\right)$ in η_j are strong assumptions when ε_i and η_j are independently distributed from x_i and z_j. While this may be difficult to motivate, it is commonly used in discrete choice literature. This paper is a first step towards providing theoretical results on identification and estimation in this market, and these assumptions significantly ease the analysis.

2.3 Pairwise Stability

Definition 1 *A* ***match*** *is a probability measure* μ *on* $(\chi \times \mathbb{R}) \times (\zeta \times \mathbb{R})$ *with marginals* $m^{x,\varepsilon}$ *and* $m^{z,\eta}$ *respectively. A* ***slot-match*** $\mu^{\mathbb{S}}$ *is a probability measure on* $(\chi \times \mathbb{R}) \times (\zeta \times \mathbb{R} \times \mathbb{S})$ *where* $\mathbb{S} \subseteq \mathbb{N}$ *indexes slots.*

A slot-match is relevant for considering a model with many-to-one matching, and induces a unique match μ by marginalizing over slots. The traditional definition of a match used in Roth and Sotomayor (1992) is based on a matching function $\mu^*(i) \mapsto J \cup \{i\}$ where J is the set of firms. With probability 1, such a function defines a unique counting measures of the form $\mu_n = \frac{1}{n}\sum_{i,j=1}^{n} \delta_{\left(X_i, \varepsilon_i, Z_j, \eta_j\right)}$ where $\delta_{\left(X_i, \varepsilon_i, Z_j, \eta_j\right)} > 0$ only if i is matched to j in a finite sample. This fact is a consquence of Assumption 1 (i), which implies that in a finite economy, (z, η) identifies a unique firm with probability 1. [3]

Definition 2 *A match* μ *is* ***pairwise stable*** *if there do not exist two (measurable) sets* $S_I \subseteq \chi \times \mathbb{R}$ *and* $S_J \subseteq \zeta \times \mathbb{R}$ *in the supports of* $m^{x,\varepsilon}$ *and* $m^{z,\eta}$ *respectively, such that* $\int_{S_I} v(X, \varepsilon)\, dm^{x,\varepsilon} > \int_{S_I} v(X, \varepsilon)\, d\mu(\cdot, S_J)$ *and* $\int_{S_J} u(Z, \eta)\, dm^{z,\eta} > \int_{S_J} u(Z, \eta)\, d\mu(S_I, \cdot)$*.*

This definition of pairwise stability is also equivalent to that for a finite market since μ is pairwise stable if and only if it does not have support on blocking pairs. Existence of a pairwise stable match follows in a finite market because preferences are *responsive* (Roth and Sotomayor, 1992) and uniqueness follows from *alignment of preferences* as disscussed in Clark (2006) and Niederle and Yariv (2009).

Remark 1 *In the model employed here the pairwise stable match* μ *has support on* $(x, \varepsilon, z, \eta)$ *only if* $F_U(u(z, \eta)) = F_V(v(x, \varepsilon))$ *where* F_U *and* F_V *are the cumulative distributions of* u *and* v *respectively.*

[3] In addition to a traditional matching function, in finite sample our definition also allows for fractional matchings. However, such realizations are not observed in typical datasets on matches.

Hence, this model exhibits full assortativity, so that the firm with the q-th quantile position of value to the worker is matched with the worker with the q-th quantile of desirability to the firm.

3 Identification

This section presents conditions under which we can the identify the functions $h(x)$, $g(z)$ and the distributions of ε and η using the marginal distribution of the match μ on the observables, $\chi \times \zeta$. These objects allow determine the distribution of preferences, or the probabilities

$$\mathbb{P}\left(v\left(X_1, \varepsilon_1\right) > v\left(X_2, \varepsilon_2\right) | X_1 = x, X_2 = x'\right) \tag{3}$$

$$\text{and} \qquad \mathbb{P}\left(u\left(Z_1, \eta_1\right) > u\left(Z_2, \eta_2\right) | Z_1 = z, Z_2 = z'\right). \tag{4}$$

We make the following assumptions on $h(\cdot)$ and $g(\cdot)$

Assumption 2 *(i)* $h(\bar{x}) = 0$, $|\nabla h(\bar{x})| = 1$ *and* $g(\bar{z}) = 0$, $|\nabla g(\bar{z})| = 1$

(ii) ε *and* η *are median-zero*

(iii) $h(\cdot)$ *and* $g(\cdot)$ *have full support over* $\mathbb{R}$

(iv) $h(\cdot)$ *and* $g(\cdot)$ *are differentiable*

(v) The measures m^x *and* m^z *admit bounded densities* f_X *and* f_Z

(vi) The densities f_ε *and* f_η *are bounded, differentiable, and have a non-vanishing characteristic function*

Assumptions 2 (i) and (ii) impose scale and location normalizations that are necessary and common in the discrete choice literature since the latent variables are not observed. Such normalizations are necessary in single-agent discrete choice models and are without loss of generality. Assumption 2 (iii) is a support condition often necessary for non-paramteric identification. Assumption 2 (iv) - (vi) are regularity conditions.

Assumption 3 *At the pairwise stable match μ, the conditional distribution* $\mu^{(z,\eta)|(x,\varepsilon)} = \mu^{(z,\eta)|(x',\varepsilon')}$ *if* $h(x) + \varepsilon = h(x') + \varepsilon'$ *and* $\mu^{(x,\varepsilon)|z,\eta} = \mu^{(x,\varepsilon)|(z',\eta')}$ *if* $g(z) + \eta = g(z') + \eta'$*.*

This assumption requires that the desirability of an agent alone to determine the quality of their matches, and not directly through the underlying traits. In other words, the sorting observed in the data can depend only on the observable characteristics through their effect on the desirability to the other side of the market. Without this assumption, sorting patterns in the data may not be related to preferences.

3.1 Identification from Sorting Patterns

Data from one-to-one matches can be summarized as a joint distribution F_{XZ} of observed firm and worker traits. They allow assessing the sorting of worker observable traits to firm observables. This section shows that such data are sufficient for identifying certain preference features on both sides of the market, but not the entire distribution of preferences on both sides of the market. In fact, we show that the joint distribution is limited in its ability to identify preferences.

3.2 Identification of Indifference Curves and a Sign-restriction

Our first result shows that $h(\cdot)$ and $g(\cdot)$ are identified up to monotone transformations using only sorting patterns in the data.

Lemma 1 *Under Assumptions 1 and 3, and the representation of preferences in equations (1) and (2), the level sets of the functions* $h(\cdot)$ *and* $g(\cdot)$ *are identified a one-to-one observed match* μ*.*

Proof. See Appendix A.1. ■

We can detemine whether or not two worker types x and x' are equally desirable from the sorting patterns observed in one-to-one, hence, also in many-to-one matches. Intutively, if two worker types have equal values of $h(\cdot)$, then the distributions of their desirability to firms are identical. Consequently, the distribution of firms they match with are also identical. In a pairwise stable match, under the additive structure of equations (1) and (2), and independence of unobserved traits, the distribution of firm observable types these workers are matched with turns out to be identical. Conversely, if two worker types are matched with different distribution of firm observables, they cannot be identical in observable quality because of Assumption 3.

While the level-sets of $h(\cdot)$ and $g(\cdot)$ are known, we cannot yet determine $h(\cdot)$ and $g(\cdot)$ up to positive monotone transformations on either side of the market. In particular, it does not tell whether a any given worker trait is desirable or not. Intuitively, assortative matching between, say firm size and worker age, may result from either both traits being desirable or both traits being undesirable. The next result shows that under a **sign restriction** only on one side of the market, is sufficient for identifying both $h(\cdot)$ and $g(\cdot)$ up to positive monotone transformations.

Assumption 4 *The function $h(x)$ is strictly increasing in its first argument, x_1. Further, x_1 has full support in $\mathbb{R}$, and $\lim_{x_1 \to \infty} h(x) = \infty$, $\lim_{x_1 \to -\infty} h(x) = -\infty$.*

Proposition 1 *Assumption 4 and the conditions in Lemma 1 determine $h(\cdot)$ and $g(\cdot)$ up to positive monotone transformations.*

Proof. See Appendix A.2. ■

The sign restriction allows us to order the level sets of h. Workers at higher level sets of h also receive a more desirable distribution of firms. We can then use this to order the level sets of $g(\cdot)$ as well. A symmetric result would hold under a sign restriction on g.

3.3 Limitations of Sorting Patterns: A Negative Result

Our next result shows a limitation of empirical content in data from one-to-one matches. While these data are sufficient to determine whether two traits, for example, worker traits

x and x', are equally desirable or not, knowing the probability that the worker with trait x is chosen over x'. The approach is to show that the joint distribution F_{XZ} generated by a class of models satisfying Assumptions 1 - 4 can be rationalized with a modified model in which $\varepsilon \equiv 0$. Such non-identification can be problematic for counterfactuals relying on the probability of choices. For instance, the result implies that the data can be rationalized in a model in which any worker with trait x is chosen over any worker with trait x' if $h(x) > h(x')$ even if this is not the case.

The class of models we consider satisfy the following condition:

Condition 1 *The primitives h, g, $F_{X,\varepsilon}$, $F_{Z,\eta}$ satisfy the conditions*

(i) $F_{h(X)|V}(\bar{h}, v) = \mathbb{P}\left(h(X) \leq \bar{h} | h(X) + \varepsilon = v\right) = \gamma(\kappa\bar{h} - v)$ for some function γ and constant κ

(ii) $F_V^{-1} \circ F_U$ is a linear function

(iii) The functions h, g and $f_{h(X)|V}$ are twice continuously differentiable

While these conditions may seem restrictive at first glance, a large class of models satisfy this condition and constructing examples is fairly straightforward. As we will show in Example 2, these conditions are satisfied, for example, if $h(x)$ and $g(z)$ are linear, and X, ε, Z and η are independent and normally distributed.

Proposition 2 *Under Condition 1 and Assumptions 1 - 4, data from one-to-one matches can be rationalized in a matching model with $\varepsilon \equiv 0$.*

Proof. The proof proceeds by re-writing the matching model with $\varepsilon \equiv 0$ as a transformation model of Chiappori and Komunjer (2008), which they show is correctly specified. See Appendix A.3 for details. ■

The result shows that despite imposing additional regularity conditions, data from one-to-one matches can be rationalized using a model in which only one set of unobservables are present. In these data, only the joint distribution of observable characteristics given by the match or sorting patterns are known. Logan et al. (2008) and Boyd et al. (2013) employ empirical strategies that only use sorting patterns to estimate preferences. Our non-identification result implies these point estimates may be sensitive to parametric assumptions. As shown in the next section, data from many-to-one matching markets has additional information that is useful for identification. The dataset used by Boyd et al. (2013) contains this information, but their empirical strategy does not take advantage of it.

In the following example, we compute the joint distribution observed in the dataset for a family of models that satisfy Condition 1 and show that a transformation model produces the same data. Further, the same family of models can be used to compute a set of matching models that are observationally equivalent with data from one-to-one matches.

Example 2 *Let $h(x) = x$ and $g(z) = z$. Assume that X, Z are distributed as $N(0,1)$ and ε, η are distributed as $N(0,\sigma_\varepsilon^2)$ and $N\left(0,\sigma_\eta^2\right)$ respectively. The distributions of U and V are therefore $N\left(0,1+\sigma_\eta^2\right)$ and $N(0,1+\sigma_\varepsilon^2)$ respectively. It is straightforward to show that $X|V=v \sim N\left(\frac{1}{1+\sigma_\varepsilon^2}v, \frac{\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2}\right)$, that $U|Z \sim N\left(Z,\sigma_\eta^2\right)$, and that $F_V^{-1}\circ F_U = \left[\frac{1+\sigma_\varepsilon^2}{1+\sigma_\eta^2}\right]^{1/2}$. Condition 1 is therefore satisfied since the density $f_{X|V}(x,v)$ depends only on the difference $x-\frac{1}{1+\sigma_\varepsilon^2}v$, $F_V^{-1}\circ F_U$ is linear, and the density of a normal distribution is twice-continuously differentiable.*

We now compute the conditional distribution $F_{X|Z}$. Since the model exhibits full assortativity, the distribution of $X|Z=z$ is given by the distribution of

$$\frac{1}{1+\sigma_\varepsilon^2}F_V^{-1}\circ F_U(z+\eta)+\varepsilon_1$$

where $\varepsilon_1 \sim N\left(0,\frac{\sigma_\varepsilon^2}{1+\sigma_\varepsilon^2}\right)$ and $\eta \sim N\left(0,\sigma_\eta^2\right)$, independently of X and Z. Hence, $X|Z=z$ is distributed as

$$N\left(\frac{z}{\kappa^{1/2}},1-\frac{1}{\kappa}\right) \tag{5}$$

where $\kappa = (1+\sigma_\varepsilon^2)\left(1+\sigma_\eta^2\right)$. The joint distribution F_{XZ} produced by this model is identical *to one produced by the following tranformation model:*

$$X = \frac{1}{\kappa^{1/2}} Z + \eta_1$$

where $\eta_1 \sim N\left(0, 1 - \frac{1}{\kappa}\right)$.

Further note that the joint distribution as given in (5), produced by the above model is identical for all pairs $(\sigma_\varepsilon, \sigma_\eta)$ *with* $\left(1+\sigma_\eta^2\right)\left(1+\sigma_\varepsilon^2\right) = \kappa$. *Thus, the family of matching models with* $\left(1+\sigma_\eta^2\right)\left(1+\sigma_\varepsilon^2\right) = \kappa$ *are observationally equivalent with data from one-to-one matches.*

The example shows that Proposition 2 is not pathalogical. We further illustrate the nature of non-identification in Section 3.5 using a simulated objective function. The problem in these simulations is allieviated when data from many-to-one matching markets is observed and used.

3.4 Identification from Many-to-One Matches

We now show that data from many-to-one matching markets can be used to identify the model. We consider a limit dataset in which there are a large number of firms, and each firm has two workers. We assume that dataset does not distinguish directly between workers, and we arbitrarily label the slots occupied by each worker as slots 1 and 2, independently of the firm and worker characteristics. In such a dataset, with finitely many firms, the data consist of the joint distribution $F_{X_1,X_2,Z}$ where X_1 and X_2 are the observed characteristics of the two workers employed at a firm with observable characteristic Z. By the assumption that the slots are independently labelled, we have that $F_{X_1|Z} = F_{X_2|Z}$.

Our main result is that the distribution of $U|Z$ and $V|X$ is identified in such a dataset.

Theorem 3 *Under Assumptions 1 - 4, the functions* $h(\cdot)$, $g(\cdot)$ *and the densitites* f_η *and* f_ε *are identified from observing data from two-to-one matching, i.e.* f_{z,x_1,x_2}.

Proof. See appendix A.4. ■

The proof proceeds by interpretting our model in terms of a measurement error model, and employing techniques in Hu and Schennach (2008) to prove identification. More specifically, if we augment the joint distribution $f_{X_1,X_2,Z}$ by the latent quantile q of utility afforded by a specific firm to workers, then we can write

$$f_{X_1,X_2,Z,q}(x_1,x_2,z,q) = f_{X_1|q}(x_1|q)\ \ f_{X_2|q}(x_2|q)\ \ f_{Z|q}(z|q)\ \ f_q(q),$$

where $f_q(q) = 1$, $f_{X_1|q}(x_1|q)$ is the conditional density at x_1 given that $h(x_1) + \varepsilon = F_V^{-1}(q)$, and analogously for $f_{X_2|q}(x_2|q)$ and $f_{Z|q}(z|q)$. Intuitively, this simplification arises from the assumption that preferences are homogeneous, which implies perfect assortative matching on v and u. As noted earlier, the distribution of worker (firm) characteristics that a firm (worker) expects to match with under Assumption 3 only depends on its quality, and in this case the relevant quantile in the quality of all other agents on the same side. Hence, conditional on q, the random variables X_1, X_2 and Z are independent. The quantile therefore presents the role of the latent index that determines the joint distribution of the observables in a model similar to the measurement error model of Hu and Schennach (2008).[4]

A loose interpretation is that the observed worker/firm characteristics present a noisy measures of the true quality of the partners matched with each other. Agarwal (2013) discusses this intuition in the context of the medical residency market. The argument is that if the medical school quality of a resident is highly predictive of human capital, then the variation within program in human capital should be low. If unobservables such as test scores and recommendations are important, then residency programs should be matched with medical residents from varying medical school quality. Our results formally show the usefulness of data from many-to-one matching. We therefore recommend the use of this

[4]There is a technical difference in our model and Assumption 5 in Hu and Schennach (2008) related to the unique indexing of the latent eigenvalues. While Hu and Schennach (2008) assume that a known functional maps the eigenvalues, say $f(x_1|q)$ to an observable quantity, we do not need this assumption since the distribution of quantiles is known to be uniform. Details are in the appendix.

information when available.

3.5 Importance of Many-to-one Match Data: Simulation Evidence

The identification results presented in the previous section relied on observing data from many-to-one matching, and shows that the model is not identified using data from one-to-one matches. In this section, we present simulation evidence from a parametric version of the model to elaborate on the nature of non-identification and to illustrate the importance of using information from many-to-one matching in estimation. Although the formal identification results were based on many workers per firm, we will be able to illustrate the empirical information in many-to-one matching even in simulations with only a few workers per firm.

We simulate a dataset of pairwise stable matches from a simple model and then compare objective functions of a method of simulated moments estimator that is constructed from moments only using information present in sorting patterns to another that also use information from many-to-one matching. The dataset of paiwise stable matches is simulated from a model of the form

$$v_i = x_i\alpha + \varepsilon_i$$
$$u_j = z_j\beta + \eta_j$$

where $x_i,\ z_j,\ \varepsilon_i,\ \eta_j$ are distributed as standard normal random variables. These parametric assumptions are identical to those used in Example 2. We generate a sample using $J = 500$ firms, each firm j has capacity q_j drawn uniformly at random from $\{1, \ldots, 10\}$. The number of workers in the simulation is $N = \sum c_j$. A pairwise stable match $\mu : \{1, \ldots, N\} \rightarrow \{1, \ldots, J\}$ is computed for $\alpha = 1$ and $\beta = 1$. Using the same draw of observables and firm capacities, the variables ε_i and η_j are simulated $S = 1000$ times, and a pairwise stable match μ_s^θ can be computed for each $s \in \{1, \ldots, S\}$ as a function of $\theta = (\alpha, \beta)$. We then compute two sets of moments

$$\hat{\psi}_{ov} = \frac{1}{N}\sum_i x_i z_{\mu(i)} \tag{6a}$$

$$\hat{\psi}^S_{ov}(\theta) = \frac{1}{S}\sum_s \frac{1}{N}\sum_i x_i z_{\mu^\theta_s(i)} \tag{6b}$$

and

$$\hat{\psi}_w = \frac{1}{N}\sum_i \left(x_i - \frac{1}{|\mu^{-1}(\mu(i))|} \sum_{i' \in \mu^{-1}(\mu(i))} x_{i'} \right)^2 \tag{7a}$$

$$\hat{\psi}^S_w(\theta) = \frac{1}{S}\sum_s \frac{1}{N}\sum_i \left(x_i - \frac{1}{\left|\left(\mu^\theta_s\right)^{-1}(\mu^\theta_s(i))\right|} \sum_{i' \in \left(\mu^\theta_s\right)^{-1}(\mu^\theta_s(i))} x_{i'} \right)^2. \tag{7b}$$

The first set, $\hat{\psi}_{ov}$ and $\hat{\psi}^S_{ov}(\theta)$, captures the degree of assortativity between the characteristics x and z in the pairwise stable matches in the generated data, and as a function of θ. For a given $\alpha > 0$ (likewise $\beta > 0$), this covariance should be increasing in β (likewise α). The second set, $\hat{\psi}_w$ and $\hat{\psi}^S_w(\theta)$ capture the within-firm variation in the characteristic x. If the value of α is large, we can expect that workers with very different values of x are unlikely to be of the same quantile. Hence, the within-firm variation in x will be small. Using both sets of moments, we construct an objective function $\hat{Q}(\theta) = \left\|\hat{\psi} - \hat{\psi}^S(\theta)\right\|_W$ where $\hat{\psi} = \left(\hat{\psi}_{ov}, \hat{\psi}_w\right)'$, $\hat{\psi}^S(\theta) = \left(\hat{\psi}^S_{ov}(\theta), \hat{\psi}^S_w(\theta)\right)'$ and W indexes the norm.

Figure 1(a) presents a contour plot of an objective function that only penalizes deviations of $\hat{\psi}_{ov}$ from $\hat{\psi}^S_{ov}(\theta)$. This objective function only using information in the sorting between x and z to differentiate values of θ. We see that pairs of parameters, α and β, with large values of α and small values of β yield identical values of the objective function. These contour sets result from identical values of $\hat{\psi}^S_{ov}(\theta)$, illustrating that this moment cannot distinguish between values along this set. In particular, the figure shows that the objective function has a trough containing the true parameter vector with many values of θ yielding similar values of the objective function.

In Figure 1(b), we consider an objective function that only penalizes deviations of $\hat{\psi}_w$ from $\hat{\psi}_w^S(\theta)$. The vertical contours indicate that the moment is able to clearly distinguish values of α because the moment $\hat{\psi}_w^S(\theta)$ is strictly decreasing in α. However, the shape of the objective function indicates that this moment cannot distinguish different values of β.

Finally, the plots of an objective function that penalizes deviations from both $\hat{m}_w$ and $\hat{m}_{ov}$ (Figure 1(c)) show that we can combine information from both sets of moments to identify the true parameter. Unlike the other two figures, this objective function displays a unique minimum close to the true parameter. Together, Figures 1(a)-(c) illustrate the importance of using both these types of moments in estimating our model.

4 Estimation

In this section, we assume that the latent utilities of workers for firms and vice-versa are known up to a finite dimensional parameter $\theta \in \Theta \subseteq \mathbb{R}^{K_\theta}$. The utilities are generated by

$$\begin{aligned} u(z,\eta;\theta) &= g(z;\theta)+\eta \\ v(x,\varepsilon;\theta) &= h(x;\theta)+\varepsilon \end{aligned}$$

where $g : \zeta \times \Theta \to \mathbb{R}$ and $h : \chi \times \Theta \to \mathbb{R}$ are known-functions that are Lipschitz-continuous in each of their arguments. We assume that the densities f_ε and f_η are known. Similar parametric assumptions are common in the discrete choice literature where one may assume a normal or an extreme value type I distribution for the unobservable ε.

Our results are for a sample of J firms, each with $\bar{c} = 2$ slots each, and consider the properties of an estimator as $J \to \infty$. The number of workers is $N = \bar{c}J$. The characteristics of each worker are sampled from the measure $m^{x,\varepsilon}$ and the characteristics of the firm are sampled from $m^{z,\eta}$.

We will study the estimator defined by

$$\theta_N = \arg\min_{\theta\in\Theta} \|\psi_N - \psi_N(\theta)\|_W , \tag{8}$$

where ψ_N are finite-dimensional moments computed from the sample, $\psi_N(\theta)$ are computed from the observed sample of firms and workers as a function of θ, and W defines a norm. For instance, we may use a quadratic form in the difference $(\psi_N - \psi_N(\theta))$, yielding a GMM estimator.

Let m^u, m^v be the image measures of $m^{z,\eta}$ and $m^{x,\varepsilon}$ under $u(z,\eta;\theta_0)$ and $v(x,\varepsilon;\theta_0)$ respectively, and m_N^u, m_J^v be their empirical analogues. We will treat these densities as known functions. Let $\Psi : \chi \times \chi \times \zeta \to \mathbb{R}^{K_\Psi}$ be a moment function. We make the following assumption:

Assumption 5 *(i) $\Psi(x_1,x_2,z)$ is bounded with bounded partial derivatives and symmetric in x_1 and x_2*

(ii) The densities f_ε and f_η are bounded with bounded derivatives and have full support on the real line

(iii) m^x and m^z admit densities f_X and f_Z

(iv) The conditional densities $f_{X|v}(x)$ and $f_{Z|u}(z)$ have uniformly bounded derivatives.

The symmetry of $\Psi(x_1,x_2,z)$ in the first two arguments ensures that there is no distinction between the two positions at any given firm. The remaining conditions are regularity conditions.

The data consist of $N = 2J$ matches, $\{(x_{2j-1}, x_{2j}, z_j)\}_{j=1}^J$, and can be used to construct empirical moments of the form

$$\psi_N = \frac{1}{N}\sum_j \Psi(x_{2j-1}, x_{2j}, z_j) . \tag{9}$$

The moments discussed previously, in equations (6a) and (7a), are given by particular choices for Ψ.

Instead of writing the sampling process as drawing pairs of (x_i, ε_i) and $\left(z_j, \eta_j\right)$, it will be convenient to first sample N and J draws from m^v and m^u respectively, and then sample $x_i|v_i$ and $z_j|u_j$ from their respective conditional distributions. This sampling process has an identical distribution for (x_i, ε_i) and $\left(z_j, \eta_j\right)$ as sampling directly from $m^{x,\varepsilon}$ and $m^{z,\eta}$ directly.

In the population dataset, firms with the q-th quantile of m^v are matched with workers on the q-th quantile of m^u. With the sampling process described above in mind, the moment can be written as

$$\psi = \int_0^1 \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q); \theta_0\right) dq$$

where F_v and F_u are the cdf corresponding to m^v and m^u respectively and $\tilde{\psi}(v_1, v_2, u)$ is the expectation of $\Psi(X_1, X_2, Z)$ given that X_1 and X_2 are drawn from $m^{x|v}$ and Z is drawn from $m^{z|u}$. The term $\tilde{\psi}(v_1, v_2, u)$ can be written as

$$\begin{aligned}
\tilde{\psi}(v_1, v_2, u; \theta) &= \int \Psi(X_1, X_2, Z) f_{X|v_1}(X_1; \theta) f_{X|v_2}(X_2; \theta) f_{Z|u}(X_2; \theta)\, dm^x dm^x dm^z \qquad (10)\\
&= \frac{\int \Psi(X_1, X_2, Z) f_\varepsilon(v - h(X_1; \theta)) f_\varepsilon(v - h(X_2; \theta)) f_\eta(u - g(Z; \theta))\, dm^x dm^x dm^z}{\int f_\varepsilon(v - h(X_1; \theta)) f_\varepsilon(v - h(X_2; \theta)) f_\eta(u - g(Z; \theta))\, dm^x dm^x dm^z}.
\end{aligned}$$

Our first result shows that the empirical analog in equation (9) converge at the true parameter θ_0 to ψ.

Proposition 3 *Let ψ^k and ψ_N^k denote the k-th dimensions of ψ and ψ_N respectively. If Assumption 5 is satisfied, then for each $k \in \{1, \ldots, K_\Psi\}$, $\psi_N^k - \psi^k$ converges in probability to* 0.

Proof. See Appendix B.2. ■

The primary technical difficulty arises from the dependent data nature of the observed matches. By re-writing the sampling process as one in which the utilities u and v are drawn first, we can condition on the utility-matches in the data. The observed characteristics of the matched agents are then sampled conditional on this utility draw. This sampling process, although identical to drawing the characteristics directly from $m^{x,\varepsilon}$ and $m^{z,\eta}$, allows a more tractable approach to proving consistency of the moments. The proof technique is based on leveraging the triangular array structure implied by this process: the individual components of the triple (X_1, X_2, Z) are independent conditional on the utilities drawn.

For estimation, we also need to consider the population and empirical analogs of ψ evaluated at values of θ other than θ_0. The population analog is given by

$$\psi(\theta) = \int_0^1 \tilde{\psi}\left(F_{v;\theta}^{-1}(q), F_{v;\theta}^{-1}(q), F_{u;\theta}^{-1}(q); \theta\right) dq$$

where $F_{v;\theta}(v) = \int_{-\infty}^{v} F_\varepsilon(v - h(X;\theta))\, dm^x$, $F_{u;\theta}(u) = \int_{-\infty}^{u} F_\eta(u - g(Z;\theta))\, dm^z$ and $\tilde{\psi}$ is defined in equation (10) above. We study an estimator that uses the following sample analog of $\psi(\theta)$ as a function of θ,

$$\begin{aligned} \psi_N(\theta) &= \int_0^1 \tilde{\psi}\left(F_{N,v;\theta}^{-1}(q), F_{N,v;\theta}^{-1}(q), F_{J,u;\theta}^{-1}(q); \theta_0\right) dq \qquad (11) \\ \tilde{\psi}_N(v_1, v_2, u; \theta) &= \frac{\int \Psi(X_1, X_2, Z) f_\varepsilon(v_1 - h(X;\theta)) f_\varepsilon(v_2 - h(X;\theta)) f_\eta(u - g(Z;\theta))\, dm_N^x dm_N^x dm_J^z}{\int f_\varepsilon(v_1 - h(X;\theta)) f_\varepsilon(v_2 - h(X;\theta)) f_\eta(u - g(Z;\theta))\, dm_N^x dm_N^x dm_J^z}, \end{aligned}$$

where $F_{N,v;\theta}$ and $F_{J,u;\theta}$ are empirical cdf functions from a random sample from $F_{v;\theta,m_N^x}(v) = \int_{-\infty}^{v} F_\varepsilon(v - h(X;\theta))\, dm_N^x$ and $F_{v;\theta,m_J^z}(v) = \int_{-\infty}^{u} F_\eta(u - g(Z;\theta))\, dm_J^z$ respectively. $\psi_N(\theta)$ can be computed by first drawing ε and η to simulate $F_{N,v;\theta}$ and $F_{J,u;\theta}$, and then using the expression in equation (11) to compute $\psi_N(\theta)$. It may also be possible to create a simulation analog of $\psi_N(\theta)$, that uses a second simulation step to approximate the integral. More specifically, we may independently sample from the conditional distributions of X and Z given the measures m_N^x and m_J^z and simulated values of v_i and u_j.

The next result proves uniform convergence of the difference $\psi(\theta) - \psi_N(\theta; m_N^x, m_J^z)$, a result required for consistency of the estimator.

Proposition 4 *Let $\psi^k(\theta)$ and $\psi_N^k(\theta; m_N^x, m_J^z)$ denote the k-th dimensions of $\psi(\theta)$ and $\psi_N(\theta; m_N^x, m_J^z)$ respectively. If Assumptions 5(i) - (iii) are satisfied, then for each $k \in$*

$\{1,\ldots,K_\Psi\}$, $\left|\psi^k(\theta)-\psi_N^k(\theta;m_N^x,m_J^z)\right|$ *converges in outer probability to* 0 *uniformly in* θ.

Proof. See Appendix B.3. ■

Again, the proof leverages the triangular sampling structure, conditioning on the drawn utilities. The first step is to prove that the cumulative distribution of utilities converge uniformly in θ as the sampled observed characteristics, m_N^x and m_J^z converge to their population analogs. Given this, we can take advantage of the triangular structure to construct the expected value of Ψ given an empirical distribution of sampled utilities by computing $\tilde{\psi}_N(v_1,v_2,u;\theta)$ along the quantiles of $F_{N,v;\theta}$ and $F_{J,u;\theta}$.

The proof is not a direct extension of techniques in Proposition 3. Intuitively, the particular triples (X_1,X_2,Z) that are matched are less tractable across values of θ. This is because the pairings depend on the ranks of given observations, which depends on both the overall utility of the observation and the utilities of other observations. Both these vary with the value of θ, and depend on the realized unobservables ε or η paired in the sample with the observed worker or firm characteristic. The quantity $\tilde{\psi}_N(v_1,v_2,u;\theta)$, on the other hand, does not depend on this pairing and makes an analysis across values of θ tractable. To show that ψ_N satisfies a law of large numbers, however, we did not need to consider the behaviour of the sample at values of θ other than θ_0.

Finally, we use the following standard assumptions to prove consistency of the estimator defined in equation (8).

Assumption 6 *(i) The parameter space* Θ *is compact*

(ii) There is a unique $\theta_0\in\Theta$ *for which* $\psi(\theta)=\psi(\theta_0)$

(iii) For any $\varepsilon>0$, *there exists a* $\delta>0$ *such that* $\|\psi(\theta)-\psi(\theta_0)\|_W\Rightarrow\|\theta-\theta_0\|<\varepsilon$

(iv) The norm $\|\cdot\|_W$ *is continuous in its argument*

Theorem 4 *Let* $\widehat{\theta}_N=\arg\min_{\theta\in\Theta}\|\psi_N-\psi_N(\theta)\|_W$. *If Assumptions 5 and 6 are satisfied, then* $\widehat{\theta}_N$ *converges in probability to* θ_0.

Proof. See Appendix B.4. ■

5 Monte Carlo Evidence

This section presents Monte Carlo evidence of a simulation based estimator from synthetic datasets of varying size and models of varying complexity to assess the properties of a method of simulated moments estimator. The results are presented for a simulation based estimator of the form

$$\hat{\theta}_N = \arg\min_{\theta\in\Theta} \left\| \psi_N - \psi_{N,S}(\theta) \right\|_W \quad (12)$$

$$= \arg\min_{\theta\in\Theta} \left[\left(\psi_N - \psi_{N,S}(\theta)\right)' W \left(\psi_N - \psi_{N,S}(\theta)\right) \right]^{1/2} \quad (13)$$

where ψ_N is as defined in equation (9) and $\psi_{NS}(\theta)$ is computed by averaging over $S = 100$ simulations as follows. For each simulation s, we sample the unobservables ε_i and η_j, compute the unique pairwise stable match and compute $\psi_{N,s}(\theta)$ for the simulated matches. The quantity $\psi_{N,S}(\theta) = \frac{1}{S}\sum_s \psi_{N,s}(\theta)$. The moments used are as defined in equations (6a) and (7a). One within moment is included for each observed component of x. The overall moments include each component of x interacted with each component of z.

We begin by assessing the performance of the estimator for the double-vertical model which is the focus of the theoretical results presented in the paper. We also present Monte Carlo evidence on models with workers having heterogeneous preferences for firms although we do not have formal theory on those models.

5.1 Design of Monte Carlo Experiments

Our Monte Carlo experiments vary the number of programs, $J \in \{100, 500\}$, and the maximum number of residents matched with each program $\bar{c} \in \{5, 10\}$. For each program j, the capacity c_j is chosen uniformly at random from $\{1, \ldots, \bar{c}\}$. The number of residents is a random variable set at $N = \sum c_j$. We will use up to three characteristics for residents and up to four characteristics for programs. The characteristics z_j of program j are distributed as

$$z_j = (z_{j1}, z_{j2}, z_{j3}) \sim N(a, I_3)$$

where $a = (1, 2, 3, 4)$ and I_3 is a 3×3 identity matrix. The characteristics of the residents, x_i are distributed as

$$x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4}) \sim N(b, I_4)$$

where $b = (1, 2, 3)$ and I_4 is a 4×4 identity matrix.

For each model specification, we generate 500 samples indexed by b and parameter estimates $\hat{\theta}_b$.[5] The confidence intervals are generated by using a parametric bootstrap described in Appendix C.

5.2 Results

5.2.1 The Double Vertical Model

We present Monte Carlo evidence from a model with no preference heterogeneity. The preferences are of the form,

$$v_i = x_i\alpha + \varepsilon_i \tag{14a}$$

$$u_j = z_j\beta + \eta_j \tag{14b}$$

where $\varepsilon_i \sim N(0, 1)$ and $\eta_j \sim N(0, 1)$. Table 1 presents results from two specifications. The specification in Column (1) has a single observable characteristic on each side of the market and column (2) has two observable characteristics. With few exceptions, the bias, the root mean squared error (RMSE) and the standard error fall with J and $\bar{q}$ for both specifications. The coverage ratios of 95% confidence intervals constructed from the proposed bootstrap approximation are mostly between 90% and 98%, particularly for simulations with a larger sample sizes, particularly for estimates of α. Also notice that estimates for α are more precise

[5] The $b-$th (pseudo-random) sample is generated from a Mersenne Twister algorithm with the seed b.

than estimates of β in both specfciations and all sample sizes.

5.2.2 Heterogeneous Preferences

Preference models without heterogeneity may be quite restrictive for some empirical applications. While we do not have formal results on identification or estimation of models with preference heterogeneity, we present Monte Carlo evidence from a model in which workers have heterogeneous preferences for firms. We conduct Monte Carlo simulations for pairwise stable matches using preference models of the form

$$v_i = x_i\alpha + \varepsilon_i \tag{15a}$$

$$u_{ij} = z_j\beta + \sum_{k,l}\gamma_{kl} \times x_{i,k} \times z_{j,l} + \eta_j \tag{15b}$$

where $\varepsilon_i \sim N(0,1)$ and $\eta_j \sim N(0,1)$. In this model, workers have varying preferences for the firm characteristic z based on their characteristic x. Table 2 presents results from two specifications, one with one interaction term and another with two interactions. As in the model with no preference heterogeneity, the bias, root mean square error falls with J and $\bar{c}$, and the coverage ratios are close to correct. The result suggests that a simulation based estimator for the model with preference heterogeneity may have desirable large sample properties as well.

6 Conclusion

This paper provides results on the identification and estimation of preferences from data from a matching market described by pairwise stability and non-transferable utility, when data only on final matches are observed. Our results are restricted to the case when preferences on both sides are homogeneous. We show that using information available in many-to-one matching is necessary and sufficient for non-parametric identification if data on a single large market is observed. We also prove consistency of an estimator for a parametric class of models. Finally, we present Monte Carlo evidence on a simulation based estimator.

There are several avenues for future research on both identification and estimation for similar models. While we show that it is necessary to use information from many-to-one matching for identification with data on a single large market, it may be possible to use variation in the characteristics of market participants for identification. This can be particularly important for the emprical study of marriage markets in the non-transferable utility framework. Our formal results are also restricted to the case with homogeneous preferences on both sides of the market. Extending this domain of preferences is particularly important. A treatment of heterogeneous preferences on both sides of the market may be of particular interest, but may need to confront difficulties arising from the multiplicity of equilibria. Finally, we have also left the exploration of computationally more tractable estimators for future research.

References

Agarwal, N. (2013). An Empirical Model of the Medical Match.

Azevedo, E. and E. Budish (2012). Strategyproofness in the large.

Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff (2013). Analyzing the Determinants of the Matching Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets. *Journal of Labor Economics 31*(1), 83–117.

Chiappori, P.-A. and I. Komunjer (2008, December). Correct Specification and Identification of Nonparametric Transformation Models. Technical report.

Chiappori, P.-A., B. Salanié, and Y. Weiss (2011). Partner Choice and the Marital College Premium.

Choo, E. and A. Siow (2006). Who Marries Whom and Why. *Journal of Political Economy 114*(1), 175–201.

Clark, S. (2006). The Uniqueness of Stable Matchings. *The B.E. Journal of Theoretical Economics contr. 6*(1), 8.

Fox, J. T. (2009). Structural Empirical Work Using Matching Models. In S. N. Durlauf and L. E. Blume (Eds.), *New Palgrave Dictionary of Economics* (Online ed.).

Fox, J. T. (2010a). Estimating Matching Games with Transfers.

Fox, J. T. (2010b). Identification in matching games. *Quantitative Economics 1*(2), 203–254.

Galichon, A. and B. Salanie (2010). Matching with Trade-offs: Revealed Preferences over Competing Characteristics.

Gordon, N. and B. Knight (2009). A spatial merger estimator with an application to school district consolidation. *Journal of Public Economics*, 752 – 765.

Hsieh, Y.-W. (2011). Understanding Mate Preferences from Two-Sided Matching Markets: Identification, Estimation and Policy Analysis.

Hu, Y., D. McAdams, and M. Shum (2013). Identification of first-price auctions with non-separable unobserved heterogeneity. *Journal of Econometrics 174*(2), 186–193.

Hu, Y. and S. M. Schennach (2008, January). Instrumental Variable Treatment of Nonclassical Measurement Error Models. *Econometrica 76*(1), 195–216.

Hu, Y. and M. Shum (2012). Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics 171*(1), 32–44.

Immorlica, N. and M. Mahdian (2005, January). Marriage, honesty, and stability. pp. 53–62.

Kojima, F. and P. A. Pathak (2009). Incentives and Stability in Large Two-Sided Matching Markets. *American Economic Review 99*(3), 608 – 27.

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

Menzel, K. (2011). Robust Decisions For Incomplete Structural Models Of Social Interactions.

Menzel, K. (2013). Large Matching Markets As Two-Sided Demand Systems.

Niederle, M. and L. Yariv (2009). Decentralized Matching with Aligned Preferences.

Roth, A. E. and M. A. O. Sotomayor (1992). *Two-Sided Matching.* Cambridge University Press.

Sorensen, M. (2007). How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital. *Journal of Finance 62*(6), 2725 – 2762.

Uetake, K. and Y. Watanabe (2012). Entry by Merger: Estimates from a Two-Sided Matching Model with Externalities.

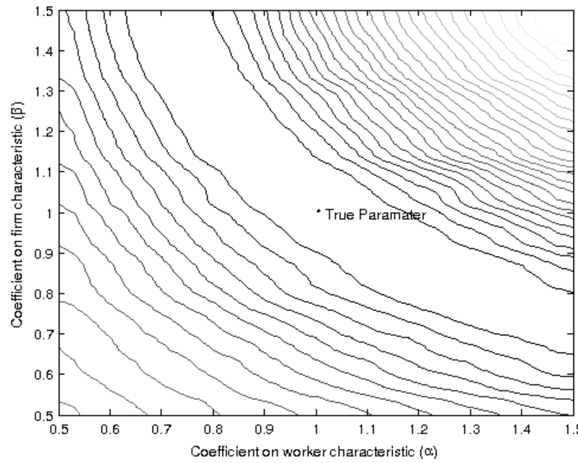
van der Vaart, A. and J. Wellner (2000). *Weak Convergence and Empirical Processes: With Applications to Statistics (Springer Series in Statistics)* (Corrected ed.). Springer.

Table 1: Monte Carlo Evidence: Double-Vertical Model

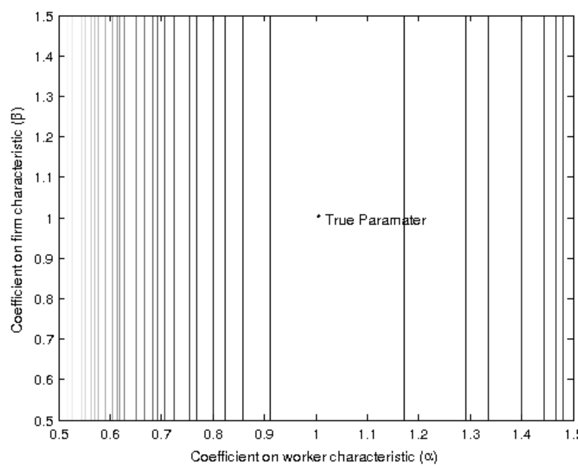
	One Characteristic (1)		Two Characteristics (2)			
	$\alpha_1\,(x_1)$	$\beta_1\,(z_1)$	$\alpha_1\,(x_1)$	$\alpha_2\,(x_2)$	$\beta_1\,(z_1)$	$\beta_2\,(z_2)$
$J = 100,\ \bar{c} = 5$						
True Par.	1	1	1	2	-1	2
Bias	0.005	0.053	0.033	0.042	-0.593	1.134
RMSE	0.093	0.239	0.270	0.412	1.063	1.938
SE	0.131	0.403	0.151	0.225	0.419	0.803
Coverage	0.954	0.970	0.868	0.850	0.760	0.750
$J = 100,\ \bar{c} = 10$						
True Par.	1	1	1	2	-1	2
Bias	0.002	0.046	-0.028	-0.036	-0.478	0.970
RMSE	0.063	0.196	0.143	0.211	0.932	1.703
SE	0.073	0.341	0.119	0.166	0.392	0.768
Coverage	0.972	0.978	0.894	0.914	0.800	0.820
$J = 500,\ \bar{c} = 5$						
True Par.	1	1	1	2	-1	2
Bias	0.000	0.002	-0.018	-0.022	-0.066	0.134
RMSE	0.042	0.086	0.052	0.069	0.207	0.383
SE	0.057	0.153	0.062	0.093	0.093	0.172
Coverage	0.934	0.978	0.950	0.954	0.808	0.848
$J = 500,\ \bar{c} = 10$						
True Par.	1	1	1	2	-1	2
Bias	0.000	0.004	-0.012	-0.017	-0.043	0.088
RMSE	0.027	0.080	0.037	0.049	0.159	0.296
SE	0.033	0.141	0.055	0.076	0.094	0.177
Coverage	0.968	0.992	0.966	0.976	0.888	0.894

Table 2: Monte Carlo Evidence: Observable Heterogeneity

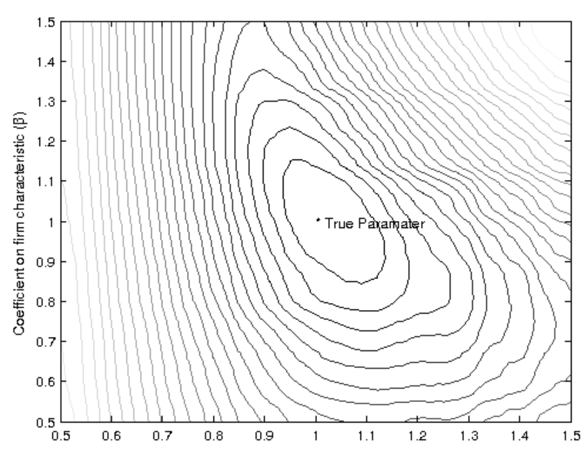
	One Interaction (1)			Two Interactions (2)					
	$\alpha_1\,(x_1)$	$\beta_1\,(z_1)$	$\gamma_1\,(z_1 \times x_3)$	$\alpha_1\,(x_1)$	$\alpha_2\,(x_2)$	$\beta_1\,(z_1)$	$\beta_2\,(z_2)$	$\gamma_{1,3}\,(z_1 \times x_3)$	$\gamma_{3,3}\,(z_3 \times x_3)$
$J = 100,\ \bar{c} = 5$									
True Par.	1	1	1	1	2	-1	2	-1	2
Bias	0.013	-0.196	0.195	0.027	0.030	-0.090	0.184	-0.112	0.175
RMSE	0.086	1.335	1.066	0.201	0.332	0.396	0.462	0.292	0.405
SE	0.136	1.861	1.389	0.114	0.175	0.219	0.293	0.179	0.284
Coverage	0.988	0.950	0.944	0.934	0.908	0.802	0.854	0.862	0.906
$J = 100,\ \bar{c} = 10$									
True Par.	1	1	1	1	2	-1	2	-1	2
Bias	0.005	-0.204	0.158	-0.005	-0.003	-0.067	0.110	-0.067	0.104
RMSE	0.061	1.002	0.765	0.114	0.187	0.317	0.378	0.235	0.290
SE	0.086	1.539	1.129	0.103	0.146	0.218	0.290	0.175	0.278
Coverage	0.984	0.940	0.964	0.960	0.966	0.884	0.912	0.906	0.964
$J = 500,\ \bar{c} = 5$									
True Par.	1	1	1	1	2	-1	2	-1	2
Bias	-0.001	0.025	0.007	-0.015	-0.018	-0.012	0.028	-0.018	0.026
RMSE	0.039	0.666	0.371	0.040	0.059	0.107	0.131	0.074	0.085
SE	0.056	0.885	0.498	0.056	0.082	0.102	0.150	0.089	0.145
Coverage	0.978	0.926	0.960	0.962	0.984	0.956	0.978	0.978	0.994
$J = 500,\ \bar{c} = 10$									
True Par.	1	1	1	1	2	-1	2	-1	2
Bias	0.000	-0.043	0.032	-0.008	-0.011	-0.010	0.021	-0.014	0.020
RMSE	0.027	0.559	0.302	0.026	0.038	0.099	0.120	0.074	0.081
SE	0.035	0.750	0.408	0.052	0.072	0.103	0.154	0.089	0.139
Coverage	0.976	0.952	0.988	0.986	0.988	0.962	0.984	0.970	0.990



(a) Sorting Moments Only



(b) Within-program Moments Only



(c) All Moments

Figure 1: Importance of Many-to-one Matches: Objective Function Contours

A Proofs: Identification

A.1 Proof of Lemma 1

As noted in Remark 1, the q-th quantile of each side matches with the q-th quantile of the other. If F_U and F_V are the distributions of utilities on each side, the match operator on utilities, which determines the worker quality index v that a firm of quality u is matched with, is given by $F_V^{-1} \circ F_U$. Note that these are both monotonically increasing.

Define T_ε be a convolution with F_ε :

$$T_\varepsilon \left(F_{h(X)}\right)(v) = \int_{-\infty}^{\infty} F_{h(X)}(v-\varepsilon)\, dF_\varepsilon, \tag{16}$$

where $F_{h(X)}$ is the cdf of the random variable $h(X)$. Note that T_ε is strictly increasing (wrt the partial order induced by First Order Stochastic Dominance) in $F_{h(x)}$. Let T_ε^{-1} be the associated inverse transform, which exists since ε admits a density and is strictly increasing since it is the inverse of a monotone operator. Let T_η be a convolution with F_η.

The distribution of U that $h(x)$ is matched with is then given by $T_\varepsilon\left(\delta_{h(x)}\right) \circ F_V^{-1} \circ F_U$ where $\delta_{h(x)}$ is the dirac delta measure at $h(x)$. The distribution on $g(Z)$ that a given $h(x)$ is matched with is $T_\varepsilon\left(\delta_{h(x)}\right) \circ F_V^{-1} \circ F_U \circ T_\eta^{-1}$. Since a composition of strictly increasing functions is strictly increasing, $T_\varepsilon\left(\delta_{h(x)}\right) \circ F_V^{-1} \circ F_U \circ T_\eta^{-1}$ is strictly increasing in $\delta_{h(x)}$.

Now consider two values $x_1, x_2 \in \operatorname{supp} m^x$ such that $\mu(x_1, \cdot) = \mu(x_2, \cdot)$ where μ is the stable match, and, wlog, $h(x_1) > h(x_2)$. Monotonicity of $T_\varepsilon\left(\delta_{h(x)}\right) \circ F_V^{-1} \circ F_U \circ T_\eta^{-1}$ implies that $T_\varepsilon\left(\delta_{h(x_1)}\right) \circ F_V^{-1} \circ F_U \circ T_\eta^{-1} >_{FOSD} T_\varepsilon\left(\delta_{h(x_2)}\right) \circ F_V^{-1} \circ F_U \circ T_\eta^{-1}$. Therefore, it must be that $\mu(x_1, \cdot) \neq \mu(x_2, \cdot)$ since the associated cdfs satisfy $F_{g(Z)|X=x_1} >_{FOSD} F_{g(Z)|X=x_2}$.

Conversely, if $h(x_1) = h(x_2)$, we have that the distributions of $g(z)$ matched with x_1 and x_2 are the same: $T_\varepsilon\left(\delta_{h(x_1)}\right) \circ F_V^{-1} \circ F_U \circ T_\eta^{-1} = T_\varepsilon\left(\delta_{h(x_2)}\right) \circ F_V^{-1} \circ F_U \circ T_\eta^{-1}$. Assumption 3 implies that the distribution of z that x_1 and x_2 are matched with are also identical, i.e. $\mu(x_1, \cdot) = \mu(x_2, \cdot)$.

Hence, $h(x_1) = h(x_2)$ if and only if $\mu(x_1, \cdot) = \mu(x_2, \cdot)$, which is known from the data. A symmetric argument identifies the level sets of $g(\cdot)$.

A.2 Proof of Proposition 1

Identification of h up to a positive monotone transformation follows immediately from Lemma 1 and Assumption 4.

To show identification of $g(\cdot)$, note that the proof of Proposition 1, although stated for h, can be replicated to conclude that $g(z_1) > g(z_2)$ implies that $T_\eta\left(\delta_{g(z_1)}\right) \circ F_U^{-1} \circ F_V \circ T_\varepsilon^{-1} >_{FOSD} T_\eta\left(\delta_{g(z_2)}\right) \circ F_U^{-1} \circ F_V \circ T_\varepsilon^{-1}$ i.e. the distribution of $h(x)$ matching with z_1 dominates the distribution for z_2. This notion of dominance is invariant to positive transformations of $h(x)$. Hence, we can order the level sets of $g(\cdot)$.

A.3 Proof of Proposition 2

The proof proceeds by rewriting the matching model with $\varepsilon \equiv 0$ in terms of the transformation model of Chiappori and Komunjer (2008). We will appeal to Chiappori and Komunjer (2008), Proposition 2 stating that the transformation model is correctly specified. We verify that the conditional distribution $F_{X|Z}$ produced by the matching model satisfies the hypothesis of the proposition.

In what follows, we will treat X and Z as known scalars with $h(\cdot)$ and $g(\cdot)$ as increasing functions of them respectively. This simplification is without loss of generality since under the hypotheses of the result, Proposition 1 guarantees that $h(\cdot)$ and $g(\cdot)$ are known up to positive monotone transformations. Since X and Z are uni-dimensional, Assumption A3 of Chiappori and Komunjer (2008) is then equivalent to strict exogeneity of ε and η from X and Z respectively, as maintained under the hypotheses of Proposition 2.

In the matching model, quantiles of $h(X)+\varepsilon$ are matched with quantiles of $g(Z)+\eta$. We will use Proposition 2 of Chiappori and Komunjer (2008) to show that there exist increasing functions $\bar{\Gamma}$, $\bar{g}$, $F_{\bar{\eta}}$ such that the the tranformation model

$$h(X) = \bar{\Gamma}\left(\bar{g}(Z) + \bar{\eta}\right)$$

rationalizes any joint distribution F_{XZ} from a matching model satisfying Assumptions 1 - 4 and Condition 1. This is model is equivalent to a matching model with $\bar{h} = \bar{\Gamma}^{-1} \circ h$, $\varepsilon \equiv 0$, and $F_{\bar{\eta}}$, $\bar{g}$.

Let the probability that a firm with obserable trait z is matched with workers with $h(X)$ at most $\bar{h}$ be denoted

$$\Phi\left(\bar{h}, z\right) = F_{h(X)|Z}\left(\bar{h}, z\right) \tag{17}$$

Note that

$$\begin{aligned}
\Phi\left(\bar{h}, z\right) &= \int F_{h(X)|V}\left(\bar{h}, F_V^{-1} F_U\left(g\left(z\right) + \eta\right)\right) dF_\eta \\
&= \int \gamma\left(\kappa \bar{h} - A\left(g\left(z\right) + \eta\right)\right) dF_\eta,
\end{aligned}$$

for some constant A. The first equality is derived from the quantile-quantile matching of workers and firms and the second equality follows from Conditions 1 (i) and 1 (ii).

First, we ensure that Φ has continuous third order partial derivatives $\partial^3\Phi(\bar{h}, z)/\partial\bar{h}\partial^2 z$ and $\partial^3\Phi(\bar{h}, z)/\partial^2\bar{h}\partial z$, and that $\partial\Phi(\bar{h}, z)/\partial\bar{h} > 0$. Twice-differentiability of h, g and $f_{X|V}$ guarantee the existence of the required partial derivatives. Further, since $F_{h(X)|V}\left(\bar{h}, v\right)$ is strictly increasing in $\bar{h}$, we have that $\partial\Phi(\bar{h}, z)/\partial\bar{h} > 0$.

We will now verify that $\Phi\left(\bar{h}, z\right)$ satisfies Condition C in Chiappori and Komunjer (2008), i.e. we need to show that

$$\frac{\partial^2}{\partial\bar{h}\partial z_k}\left(\log\left|\frac{\partial\Phi\left(\bar{h}, z\right)/\partial\bar{h}}{\partial\Phi\left(\bar{h}, z\right)/\partial z_1}\right|\right) = 0.$$

The partial derivatives of $\Phi\left(\bar{h}, z\right)$ with respect to $\bar{h}$ and z_1 are given by:

$$\begin{aligned}
\frac{\partial\Phi\left(\bar{h}, z\right)}{\partial\bar{h}} &= \kappa\int \gamma'\left(\kappa\bar{h} - A\left(g\left(z\right) + \eta\right)\right) dF_\eta \\
\frac{\partial\Phi\left(\bar{h}, z\right)}{\partial z_1} &= -A\frac{\partial g\left(z\right)}{\partial z_1}\int \gamma'\left(\bar{h} - A\left(g\left(z\right) + \eta\right)\right) dF_\eta.
\end{aligned}$$

Note that γ' exists since the existence of densities f_X, f_ε and differentiability of $h(\cdot)$ implies that the derivatives of $F_{h(X)|V}\left(\bar{h}, v\right)$ exist.

Using the expressions above, rewrite

$$
\begin{aligned}
& \frac{\partial^2}{\partial \bar{h} \partial z_k}\left(\log\left|\frac{\partial \Phi\left(\bar{h}, z\right)/\partial \bar{h}}{\partial \Phi\left(\bar{h}, z\right)/\partial z_1}\right|\right) \\
= & \frac{\partial^2}{\partial \bar{h} \partial z_k}\left(\log|\kappa| - \log\left|A\frac{\partial g(z)}{\partial z_1}\right|\right) = 0.
\end{aligned}
$$

The last equality follows since $\log\left|\frac{\partial g(z)}{\partial z_1}\right|$ and $\log|\kappa|$ do not depend on $\bar{h}$.

We now show that equations (4) and (5) in Chiappori and Komunjer (2008) are satisfied. Since $F_{h(X)|V}\left(\bar{h}, A\left(g(z)+\eta\right)\right)$ is a cdf, it is bounded, $\lim_{\bar{h}\to-\infty} F_{X|V}\left(\bar{h}, A\left(g(z)+\eta\right)\right) = 0$ and $\lim_{x\to\infty} F_{h(X)|V}\left(\bar{h}, A\left(g(z)+\eta\right)\right) = 1$ for each z and η. Hence, $\lim_{x\to-\infty}\Phi(x,z) = 0$ and $\lim_{x\to\infty}\Phi(x,z) = 1$.

To verify (5), note that

$$
\begin{aligned}
& \int_0^{\bar{h}} \frac{\partial \Phi(a,z)/\partial x}{\partial \Phi(a,z)/\partial z_1}\frac{\partial \Phi(0,z)/\partial z_1}{\partial \Phi(0,z)/\partial x} da \\
= & \int_0^{\bar{h}} \frac{\kappa}{-\partial g(z)/\partial z_1}\frac{-\partial g(z)/\partial z_1}{\kappa} da \\
= & \int_0^{\bar{h}} 1 da \\
= & \bar{h}.
\end{aligned}
$$

Condition (5) follows since $h(X)$ has full support on $\mathbb{R}$.

By Proposition 2 of Chiappori and Komunjer (2008), there exist $\bar{\Gamma}$, $\bar{g}$, $F_{\bar{\eta}}$ that rationalize Φ.

A.4 Preliminaries for Theorem 3

In what follows we treat x and z as single dimensional variable that are uniformly distributed on $[0,1]$. As noted earlier in the proof of Proposition 2, due to Proposition 1, this simplification is without loss of generality.

Let $v = h(x) + \varepsilon$, where $h(x)$ is strictly increasing with $h(\bar{x}) = 0$, $h'(\bar{x}) = 1$ and let ε be median zero with density f_ε. For quantile $\tau \in [0,1]$, let $f_{\tau|X}(\tau, x)$ be the density on $v = F_V^{-1}(\tau)$ given x, or $f_{V|X=x}\left(F_V^{-1}(\tau)|x\right) = f_\varepsilon\left(F_V^{-1}(\tau) - h(x)\right)$, where $F_V(v) = \int F_\varepsilon(v - h(x))\, dF_X$.

Lemma 2 *The function $h(x)$ and the density f_ε are identified from $f_{\tau|x}(\tau)$ if $h(x)$ is differentiable.*

Proof. Let $\phi(x, x')$ be the probability that $h(x) + \varepsilon > h(x') + \varepsilon'$ given x and x'. $\phi(x, x')$ is identified from $f_{\tau|x}(\tau)$ since it can be written as

$$\phi(x, x') = \int_0^1 \int_{\tau > \tau'} f_{\tau|X}(\tau, x)\, f_{\tau|X}(\tau', x')\, d\tau d\tau'.$$

However, $\phi(x, x')$ can also be written in terms of the primitives $h(\cdot)$ and f_ε as

$$\phi(x, x') = \int F_\varepsilon(h(x) + \varepsilon - h(x'))\, f_\varepsilon(\varepsilon)\, d\varepsilon.$$

Taking the derivative with respect to x and x', we get

$$\begin{aligned} \frac{\partial \phi(x, x')}{\partial x} &= h'(x) \int f_\varepsilon(h(x) + \varepsilon - h(x'))\, f_\varepsilon(\varepsilon)\, d\varepsilon \\ \frac{\partial \phi(x, x')}{\partial x} &= -h'(x') \int f_\varepsilon(h(x) + \varepsilon - h(x'))\, f_\varepsilon(\varepsilon)\, d\varepsilon. \end{aligned}$$

The ratio $\frac{\partial \phi(x,x')}{\partial x} / \frac{\partial \phi(x,x')}{\partial x}$ is identified and is equal to

$$-\frac{h'(x)}{h'(x')}.$$

Since $h'(\bar{x})$ is known, $h'(x)$ can be determined everywhere. The boundary conditions $h(0) = 1$ provides the unique solution to the resulting differential equation determining $h(\cdot)$.

We now need to show that F_ε is identified. Let $R_x(t)$ be the (utility-) rank distribution of x, i.e. the probability that the utility of x is below the t-quantile in utility distributions, $\mathbb{P}\left(h(x) + \varepsilon \leq F_V^{-1}(t)|x\right)$. $R_x(t)$ is known since it is equal to $\int_0^t f_{\tau|x}(\tau)\,d\tau$. Since F_V^{-1} is continuous and ε admits a density, $R_x(t)$ is continuous and strictly increasing in t. Let τ^* be the median rank of $\bar{x}$, i.e. $R_{\bar{x}}(t^*) = \frac{1}{2}$. Since ε is median-zero, $h(\bar{x}) = 0$ and $\mathbb{P}\left(h(\bar{x}) + \varepsilon \leq F_V^{-1}(t^*)|\bar{x}\right) = \frac{1}{2}$, we have that $F_V^{-1}(t^*) = 0$. For any x, $R_x(t^*)$ is therefore the probability that $h(x) + \varepsilon \leq 0$ given x, i.e. $R_x(t^*) = F_\varepsilon(-h(x))$. Since $h(x)$, $R_x(t^*)$ is known and has full support on $\mathbb{R}$, F_ε is identified. ■

Lemma 3 *If f_ε has a non-vanishing characteristic function and $h(x)$ has full-support on $\mathbb{R}$, then for any function $m(v)$, we have that $\int f_\varepsilon(v - h(x))\,m(v)\,dv = 0$ for all x implies that $m(v) = 0$. Hence, $f_{v|x}(v|x) = f_\varepsilon(v - h(x))$ is complete.*

Proof. Note that

$$\int f_\varepsilon(v - h(x))\,m(v)\,dv = \int f_\varepsilon(\varepsilon)\,m(h(x) + \varepsilon)\,d\varepsilon$$

is a convolution of $m(\cdot)$ with ε. Since f_ε has a non-vanishing characteristic, so does $f_{-\varepsilon}$. Therefore, completeness follows from Mattner (1993), Theorem 2.1. ■

For a function m, define the operator $L_{x|q}$ as $L_{x|q}(m) = \int f(x|q)\,m(q)\,dq$ where $f(x|q)$ is the conditional distribution of X given $Q = q$.

Lemma 4 *$L_{x_1|q}$ is injective if (i) f_ε has a non-vanishing characteristic function (ii) F_V is continuous and strictly increasing and (iii) $h(x)$ has full-support on $\mathbb{R}$*

Proof. We first rewrite the operator $L_{x|q}$ as a convolution:

$$\begin{aligned} L_{x|q}(m) &= \int f(x|q)\, m(q)\, dq \\ &= \int f(x|v)\, m(F_V(v))\, f_V(v)\, dv \\ &= \int f_\varepsilon(v - h(x))\, M(v)\, dv \end{aligned}$$

where the second equality follows from a change of variables. Let $M(v) = m(F_V(v))$. By Lemma 3, $f_\varepsilon(v - h(x))$ is complete. Since F_V is bijective, $m(q) = 0$ for all q. Therefore, $f(x|q)$ is complete, and as noted in HS, implies that $L_{x|q}$ is injective. ∎

Lemma 5 *$L_{x_1|x_2}$ is injective if (i) f_ε has a non-vanishing characteristic function and (ii) $h(x)$ has full-support on $\mathbb{R}$*

Proof. Note that

$$\begin{aligned} \int f_{x_1|x_2}(x_1|x_2)\, m(x_2)\, dz_2 &= \int \left(\int f(x_1, q|x_2)\, dq \right) m(x_2)\, dx_2 \\ &= \int \left(\int f(x_1|q)\, f(q|x_2)\, dq \right) m(x_2)\, dx_2 \\ &= \int \left(\int f(x_1|v)\, f(v|x_2)\, f_V(v)\, dv \right) m(x_2)\, dx_2 \\ &= \int \int f_\varepsilon(v - h(x_1))\, f(v|x_2)\, m(x_2)\, dx_2 dv \\ &= \int f_\varepsilon(v - h(x_1)) \left(\int f_\varepsilon(v - h(x_2))\, m(x_2)\, dx_2 \right) dv \end{aligned}$$

where we use (i) $f(x_1|q, x_2) = \frac{f(x_1,q|x_2)}{f(q|x_2)}$, (ii) a change of variables $F_V(v) = q$, (iii) $f(x_1|v) = \frac{f_\varepsilon(v-h(x_1))}{f_V(v)}$ and (iv) a change in the order of integration by Fubini's theorem.

By Lemma 3, $f_\varepsilon(v - h(x_1))$ is complete, and consequently, $\int f_{x_1|x_2}(x_1|x_2)\, m(x_2)\, dx_2 = 0$ implies $\int f_\varepsilon(v - h(x_2))\, m(x_2)\, dx_2 = 0$ for all v. A second application of Lemma 3 implies that $m(x) = 0$ for all x (since $f_\varepsilon(v - h(x_2))$ is complete). Hence, $f_{x_1|x_2}(x_1|x_2)$ is complete, and as noted in Hu and Schennach, completeness implies injectivity. ∎

Lemma 6 *$\int_0^1 f_{x|q}(x|q)\, m(q)\, dq = 0$ for all x, implies that $m(q) = 0$ if (i) f_ε has a non-vanishing characteristic function and (ii) $h(x)$ has full-support on $\mathbb{R}$*

Proof. Note that

$$\begin{aligned}
\Delta_{x|q} m(q) &= \int_0^1 f_{x|q}(x|q)\, m(q)\, dq \\
&= \int_0^1 f_\varepsilon\left(F_V^{-1}(q) - h(x)\right) m(q)\, dq \\
&= \int_0^1 f_\varepsilon(v - h(x))\, m(F_V(v))\, f_V(v)\, dv \\
&= \int_0^1 f_\varepsilon(v - h(x))\, \tilde{m}(v)\, dv
\end{aligned}$$

where the first equality follows from the definition of q, and the second equality follows from a change of variables $q = F_V(v)$. The last expression is a convolution of $\tilde{m}$ with respect to $f_{-\varepsilon}$. Since f_ε has a non-vanishing characteristic, so does $f_{-\varepsilon}$. Therefore, by Mattner (1993), Theorem 2.1, $\int_0^1 f_\varepsilon(v - h(x))\, \tilde{m}(v)\, dv = 0$ implies that $\tilde{m}(v) = 0$. Since $0 < f_V(v) < \infty$ for all $v \in \mathbb{R}$, we have that $\tilde{m}(v) = 0$ implies $m(F_V(v))\, f_V(v) = 0$, and $m(F_V(v))$ for all v. Hence, $\Delta_{x|q} m(q) = 0$ implies that $m(q) = 0$ for all q. ■

Our final preliminary result will use the following condition defining a set of eigenvalues $f_{x|q}(x|q)$:

Condition 2 *(i) For all $q_1^* \neq q_2^* \in [0,1]$, the set $\left\{x : f_{x|q}(x|q_1^*) \neq f_{x|q}(x|q_2^*)\right\}$ has positive probability under f_x.*

(ii) $f_{x|q}(x|q)$ continuously differentiable in q

(iii) for all $q \in (0,1)$, there exists and x, such that $\frac{\partial f_{x|q}(x|q)}{\partial q} \neq 0$

(iv) $f_q(q) = 1\{q \in [0,1]\}$

(v) $\int_0^1 f_{x|q}(x|q)\, m(q)\, dq = 0$ for all x, implies that $m(q) = 0$

Lemma 7 *Consider two conditional distributions $\tilde{f}_{\tilde{q}}(x|\tilde{q})$ and $f_q(x|q)$ satisfying Condition 2. If there exists a bijection $Q : [0,1] \to [0,1]$ such that $\tilde{f}_{\tilde{q}}(x|Q(q)) = f_q(x|q)$ then Q is the identity.*

Proof. Note that the function $f_{x|q}(x|q) = f_\varepsilon\left(F_U^{-1}(q) - g(x)\right)$ is a particular family of eigenfunctions that satisfy the assumptions of the theorem. We will show that if there exists a reindexing $\tilde{f}_{x|\tilde{q}}(x|\tilde{q})$ of q via a bijection $Q : [0,1] \to [0,1]$ such that $\tilde{f}_{x|\tilde{q}}(x|Q(q)) = f_{x|q}(x|q)$ (Assumption HS.4 requires $Q(q)$ to be injective on $[0,1]$ and the support assumption in the hypothesis implies surjectivity). If $f_q(q) = 1$ and $f_{\tilde{q}}(\tilde{q}) = 1$, then $Q(\cdot)$ is the identity.

By the assumptions of the theorem,

$$f(x) = \int_0^1 f_{x|q}(x|q)\,dq = \int_0^1 \tilde{f}_{x|q}(x|q)\,dq = \int_0^1 f_{x|q}(x|Q(q))\,dq.$$

A change of variables, $v = Q(q)$ yields that

$$\begin{aligned}\int_0^1 f_{x|q}(x|Q(q))\,dq &= \int_0^1 f_{x|q}(x|v)\,dQ^{-1}(v) \\ &= \int_0^1 f_{x|q}(x|q)\,\frac{1}{Q'\left(Q^{-1}(q)\right)}dq.\end{aligned}$$

The second inequality follows from the inverse function theorem. Differentiability of Q follows from the implicit function theorem: $Q(q)$ is defined implicitly from $\tilde{f}_{x|\tilde{q}}(x|Q) - f_{x|q}(x|q) = 0$, where for every Q there exists x such that $\tilde{f}_{x|\tilde{q}}(x|Q)$ has a non-zero derivative under the hypotheses of the theorem.

Hence,

$$\begin{aligned}\int_0^1 f_{x|q}(x|q)\,dq - \int_0^1 f_{x|q}(x|Q(q))\,dq &= 0 \\ \Rightarrow \int_0^1 f_{x|q}(x|q)\left(1 - \frac{1}{Q'\left(Q^{-1}(q)\right)}\right)dq &= 0.\end{aligned}$$

By Lemma 6, for all $q \in [0,1]$

$$\begin{aligned}&\left(1 - \frac{1}{Q'\left(Q^{-1}(q)\right)}\right) = 0 \\ \Rightarrow\ & Q'(q) = 1.\end{aligned}$$

Therefore $Q(\cdot)$ is the identity. ■

A.5 Proof of Theorem 3

In what follows we treat x and z as single dimensional variable that are uniformly distributed on $[0,1]$, and $h(\cdot)$ and $g(\cdot)$ are increasing. This simplification is without loss of generality given identification of $g(x)$ and $h(z)$ up to a positive monotone transformation by Proposition 1.

The proof follows from recasting the matching model in terms of the non-classical measurement error model similar to Hu and Schennach (2008), (henceforth HS) to first identify $f_{x|q}(x|q)$ and $f_{z|q}(z|q)$, which are the conditional distributions of x and z respectively given $h(x)+\varepsilon = F_U^{-1}(q)$ and $g(z)+\eta = F_V^{-1}(q)$.[6] Then,....

We begin by verifying Assumptions HS.1-HS.4. We prove Assumption HS.1 by showing that the joint density of $z,\ x_1,\ x_2,\ q$ is bounded wrt. Lebesgue measure and has bounded marginals. By Assumption 3, $f_{z,\ x_1,\ x_2,\ q}(z,\ x_1,\ x_2,\ q) = f_{Z|q}(z|q)\, f_{X|q}(x_1|q)\, f_{X|q}(x_2|q)\, f_q(q)$ where $f_q(q)=1$, $f_{z|q}(z|q) = f_\eta\left(F_U^{-1}(q) - g(z)\right) f_Z(z)$, and $f_{X|q}(x|q) = f_\eta\left(F_V^{-1}(q) - g(x)\right) f_X(x)$. Assumption 2(v) and (vi) imply that both $f_{z|q}(z|q)$ and $f_{X|q}(x|q)$ are bounded, hence satisfying Assumption HS.1. Assumption 4 implies Assumption HS.2. Assumption HS.3 requires that $L_{x|q}$ and $L_{x_1|x_2}$ are injective, where, for variables a,b,

$$L_{a|b}(g) = \int f_{a|b}(a|b)\, g(b)\, db.$$

Lemmas 5 and 4 imply that under Assumptions 2(iii) and 2(vi), $L_{x|q}$ and $L_{x_1|x_2}$ are injective. Finally, Assumption HS.4 is satisfied since the conditional distributions of Z and X given q are strictly increasing (in the sense of First Order Stochastic Dominance) in q.

[6] For the double-vertical matching model, the latent variable x^* in HS will be labelled q, the outcome y in HS is instead z, x in HS is x_1 and z in HS is x_2.

For a function $m(\cdot)$, and variables x and q, define the operator

$$\Delta_{x;q} m(q) = f_{x|q}(x|q)\, m(q)$$

as in HS. Since $f(z, x_1|x_2)$ is observed, for any real valued function m, we can compute

$$\int f(z, x_1|x_2)\, m(x_2)\, dx_2 = L_{x_1|q} \circ \Delta_{z;q} \circ L_{q|x_2}(m)$$

as shown in HS. Further, HS show that under assumptions HS.1 - HS.4 the conditional densities $f_{Z|q}(z|q)$ and $f_{X|q}(x|q)$ are identified upto a reindexing via an injection $Q(\cdot)$ where $\tilde{q} = Q(q)$.

If Assumption HS.5 were true, we could use it to show that Q is the identity, completing the proof. However, instead of imposing an additional assumption, we use Lemma 7 to show directly, that Q must be the identity under Assumption 2(vi). To apply Lemma 7, we need to show that $f_{x|q}(x|q) = f_\varepsilon\left(F_V^{-1}(q) - h(x)\right)$ and $f_{z|q}(z|q) = f_\eta\left(F_U^{-1}(q) - g(z)\right)$ where q are quantiles satisfies Condition 2. Since the proof is symmetric, we show this only for $f_{x|q}(x|q)$. Condition 2(i) is satisfied, as noted earlier, since the conditional distributions of X given q is strictly increasing (in the sense of First Order Stochastic Dominance) in q. Condition 2(ii) follows from Assumption 2(vi) since $\frac{\partial f_{x|q}(x|q)}{\partial q} = \frac{1}{f_V\left(F_V^{-1}(q)\right)} f'_\varepsilon\left(F_V^{-1}(q) - h(x)\right)$ and $f_V(v) > 0$. To show Condition 2(iii), we first argue by contradiction to show that there exists an $\varepsilon^* \in (-\infty, \infty)$ such that $f_\varepsilon(\varepsilon^*) \neq 0$ and then show that for all q, there exists an x, with $\frac{\partial f_{x|q}}{\partial q}(x|q) \neq 0$. = Assume that $f'_\varepsilon(\varepsilon) = 0$ for all $\varepsilon^* \in (-\infty, \infty)$. Since $f_\varepsilon(\varepsilon)$ is a density, it must therefore be a non-negative constant function with bounded support. However, this contradicts Assumption 2(vi), which requires $f_\varepsilon(\varepsilon)$ to have a non-vanishing characteristic. Hence, it must be that $f'_\varepsilon(\varepsilon^*) \neq 0$ for some $\varepsilon^* \in (-\infty, \infty)$. For q, pick x_{q,ε^*} such that $F_V^{-1}(q) - h(x_{q,\varepsilon^*}) = \varepsilon^*$ (which can be done since $h(x)$ has full-support) and note that $\frac{\partial f_{x|q}\left(x_{q,\varepsilon^*}|q\right)}{\partial q} = \frac{1}{f_V\left(F_V^{-1}(q)\right)} f'_\varepsilon\left(F_V^{-1}(q) - h(x_{q,\varepsilon^*})\right) \neq 0$. Condition 2(iv) follows from Lemma 4 under Assumptions 2(iii) and 2(vi). We have thus verified Condition 2 for $f_{x|q}$.

An identical argument follows for $f_{z|q}$.

Therefore, by Lemma 7, Q is the identity, the indexing of eigenvalues satisfying Condition 2 is unique. Hence, we have identified the quantities $f_{x|q}$ and $f_{z|q}$. The proof is completed by appealing to Lemma 2 to claim that the primitives $h(\cdot)$, $g(\cdot)$, f_η and f_ε, are identified from $f(x|q)$ and $f(z|q)$.

B Proofs: Estimation

B.1 Preliminaries

Lemma 8 *Let F_X be the cdf of a random variable X and $F_{N,X}$ be the empirical analog. For $\delta \in \left(0, \frac{1}{2}\right)$, if the density f_X exists, is continuous and is bounded away from zero on $\left[F_X^{-1}(\delta), F_X^{-1}(1-\delta)\right]$, then $\sup_{q\in[\delta,1-\delta]}\left|F_{N,X}^{-1}(q) - F_X^{-1}(q)\right| \to 0$ in outer probability.*

Proof. Follows from van der Vaart and Wellner (2000), Example 3.9.21 as a consequence of the continuous mapping theorem. ■

Let $v(x,\varepsilon;\theta) = h(x;\theta) + \varepsilon$ be Lipschitz continuous in (x,θ) and let ε have a coninuous density f_ε. For a measure m^x on X, let

$$F_{v;\theta}(v) = \int_{-\infty}^{v} F_\varepsilon(v - h(x;\theta))\, dm^x$$

and denote the corresponding empirical cdf of a sample size N with $F_{N,v;\theta}$. This is the cdf of the image $m_N^{x,\varepsilon}$ Assume $\theta \in \Theta$, compact.

Lemma 9 *If $f_{v;\theta}$ is bounded away from 0 for every compact set, then for each $\delta \in (0,1)$, $\sup_{\theta\in\Theta, q\in[\delta,1-\delta]}\left|F_{N,v;\theta}^{-1}(q) - F_{v;\theta}^{-1}(q)\right| \to 0$ in outer probability.*

Proof. The result follows from the fact that the collection of sets, $\{(x,\varepsilon,\theta) : h(x;\theta) + \varepsilon \leq v\}$, indexed by (v,θ) are Glivenko-Cantelli since ε admits a density and $h(x,\theta)$ is Lipschitz continuous. Note that $m_N^{x,\varepsilon}$ converges uniformly to $m^x \times m^\varepsilon$ over the collection of sets $\{(x,\varepsilon,\theta) : h(x;\theta) + \varepsilon \leq v\}$. We now prove continuity of $F_{N,v;\theta}^{-1}(q) - F_{v;\theta}^{-1}(q)$ with respect to $m_N^{x,\varepsilon} - m^x \times m^\varepsilon$.

By definition,

$$q = \int_{-\infty}^{F_{v,\theta}^{-1}(q)} dm^{v,\theta} = \int_{-\infty}^{F_{N,v;\theta}^{-1}(q)} dm_N^{v,\theta}$$

where $m^{v,\theta}$ and $m_n^{v,\theta}$ are image measures of $m^{x,\varepsilon}$ and $m_N^{x,\varepsilon}$ under $v(x,\varepsilon;\theta)$. Hence, for $\lambda \in (0, 1-\delta)$ and measure $m_N^{x,\varepsilon}$ such that

$$\|m_N^{x,\varepsilon} - m^x \times m^\varepsilon\| < \lambda,$$

we have

$$\begin{aligned} \left| \int_{F_{v,\theta}^{-1}(q)}^{F_{N,v;\theta}^{-1}(q)} dm^{v,\theta} \right| &\leq \left| \int_{-\infty}^{F_{N,v;\theta}^{-1}(q)} dm_N^{v,\theta} - \int_{-\infty}^{F_{N,v;\theta}^{-1}(q)} dm^{v,\theta} \right| \\ &\leq \lambda \end{aligned}$$

Since $m^{v;\theta}$ has a density bounded away from 0 on every compact interval uniformly over θ, we have that

$$\left|\left| F_{N,v;\theta}^{-1}(q) - F_{v;\theta}^{-1}(q) \right|\right| \leq \lambda \left(\inf_{\theta \in \Theta, v \in \left[F_{v;\theta}^{-1}(\delta - \lambda), F_{v;\theta}^{-1}(1-\delta+\lambda) \right]} f_{v,\theta} \right)^{-1}.$$

Hence, the result follows by the continuous mapping theorem since $\|m_N^{x,\varepsilon} - m^x \times m^\varepsilon\| \to 0$ by the law of large numbers. ■

Lemma 10 *If Assumption 5 is satisfied, then* $\left\| \nabla \tilde{\psi} \right\|_\infty < \infty$.

Proof. Note that

$$\begin{aligned}
& \tilde{\psi}(v_1, v_2, u) \\
= & \int \Psi(X_1, X_2, Z)\, dm^{x|v_1} dm^{x|v_2} dm^{z|v} \\
= & \int \Psi(X_1, X_2, Z)\ \tilde{f}_{v,x}(v_1, X_1)\ \tilde{f}_{v,x}(v_2, X_2)\ \tilde{f}_{u,z}(u, Z)\ dm^x dm^x dm^z \\
& \text{where } \tilde{f}_{v,x}(v, x) = \frac{f_\varepsilon(v - h(x; \theta_0))}{\int f_\varepsilon(v - h(X; \theta_0))\, dm^x} \text{ and } \tilde{f}_{u,z}(u, z) = \frac{f_\eta(u - g(z, \theta_0))}{\int f_\eta(u - g(Z, \theta_0))\, dm^z}
\end{aligned}$$

We will only show $\tilde{\psi}(v_1, v_2, u)$ has a bounded derivative in v_1 as the proof for the other two arguments are identical. Note that

$$\begin{aligned}
& \frac{\partial}{\partial v} \frac{f_\varepsilon(v - h(x; \theta_0))}{\int f_\varepsilon(v - h(X; \theta_0))\, dm^x} \\
= & \frac{f'_\varepsilon(v - h(x; \theta_0))}{\int f_\varepsilon(v - h(X; \theta_0))\, dm^x} - \frac{f_\varepsilon(v - h(x; \theta_0)) \int f'_\varepsilon(v - h(X; \theta_0))\, dm^x}{\left(\int f_\varepsilon(u - h(X; \theta_0))\, dm^x\right)^2} \qquad (18)
\end{aligned}$$

If the expression in equation (18) is m^x integrable in x, and the terms $\frac{f_\varepsilon(v-h(x;\theta_0))}{\int f_\varepsilon(v-h(X;\theta_0))dm^x}$ and $\frac{f_\eta(u-g(z,\theta_0))}{\int f_\eta(u-g(Z,\theta_0))dm^z}$ are respectively m^x and m^z integrable, then the Dominated Convergence Theorem implies that the derivative $\frac{\partial}{\partial v_1}\tilde{\psi}(v_1, v_2, u)$ exists and is given by

$$\int \Psi(X_1, X_2, Z) \frac{\partial}{\partial v_1} \tilde{f}_{v,x}(v_1, X_1)\ \tilde{f}_{v,x}(v_2, X_2)\ \tilde{f}_{u,z}(u, Z)\, dm^x dm^x dm^z.$$

To proceed, we will show that

$$\begin{aligned}
\sup_v \left| \int \left(\frac{f'_\varepsilon(v - h(x; \theta_0))}{\int f_\varepsilon(v - h(X; \theta_0))\, dm^X} - \frac{f_\varepsilon(v - h(x; \theta_0)) \int f'_\varepsilon(v - h(X; \theta_0))\, dm^X}{\left(\int f_\varepsilon(v - h(X; \theta_0))\, dm^X\right)^2} \right) dm^x \right| & < \infty \\
\sup_u \left| \int \left(\frac{f'_\eta(u - g(z; \theta_0))}{\int f_\eta(u - g(Z, \theta_0))\, dm^z} - \frac{f_\eta(u - g(z, \theta_0)) \int f'_\eta(u - g(Z, \theta_0))\, dm^z}{\left(\int f_\eta(u - g(Z, \theta_0))\, dm^z\right)^2} \right) dm^z \right| & < \infty
\end{aligned}$$

for the first expression since the proof of the other expression is identical. Note that

$$
\begin{aligned}
& \sup_{v} \left| \int \left(\frac{f_{\varepsilon}'\left(v-h\left(x;\theta_0\right)\right)}{\int f_{\varepsilon}\left(v-h\left(X;\theta_0\right)\right)dm^X} - \frac{f_{\varepsilon}\left(v-h\left(x;\theta_0\right)\right)\int f_{\varepsilon}'\left(v-h\left(X;\theta_0\right)\right)dm^X}{\left(\int f_{\varepsilon}\left(v-h\left(X;\theta_0\right)\right)dm^X\right)^2} \right) dm^x \right| \\
\leq \; & \sup_{v,x} \left| \frac{\partial}{\partial v} \frac{f_{\varepsilon}\left(v-h\left(x;\theta_0\right)\right) f_X\left(x\right)}{f_v\left(v\right)} \right| \\
= \; & \sup_{v,x} \left| \frac{\partial}{\partial v} f_{X|v}\left(x\right) \right| < \infty
\end{aligned}
$$

by Assumption 5 (iii) and (iv).

Since $\left\|\Psi\right\|_{\infty} < \infty$, and $\tilde{f}_{v,x}\left(v_2, X_2\right)\ \tilde{f}_{u,z}\left(u, Z\right) = \frac{f_{\varepsilon}\left(v_1-h\left(X_1;\theta_0\right)\right)}{\int f_{\varepsilon}\left(v_1-h\left(X_1;\theta_0\right)\right)dm^x} \frac{f_{\eta}\left(u-g\left(z,\theta_0\right)\right)}{\int f_{\eta}\left(u-g\left(Z,\theta_0\right)\right)dm^z} \leq 1$,

$$
\begin{aligned}
& \frac{\partial}{\partial v_1} \tilde{\psi}\left(v_1, v_2, u\right) \\
\leq \; & \left\|\Psi\right\|_{\infty} \left| \int \left(\frac{f_{\varepsilon}'\left(v-h\left(x;\theta_0\right)\right)}{\int f_{\varepsilon}\left(v-h\left(X;\theta_0\right)\right)dm^X} - \frac{f_{\varepsilon}\left(v-h\left(x;\theta_0\right)\right)\int f_{\varepsilon}'\left(v-h\left(X;\theta_0\right)\right)dm^X}{\left(\int f_{\varepsilon}\left(v-h\left(X;\theta_0\right)\right)dm^X\right)^2} \right) dm^x \right| \\
< \; & \infty
\end{aligned}
$$

■

B.2 Proof of Proposition 3

We will use the following lemmata for the result, which are stated, for each dimension of ψ. We omit the dimension index for notational simplicity.

Lemma 11 *If Assumption 5 is satisfied,* $E\left(\psi_N | m_N^v, m_J^u\right) - \psi$ *converges in probability to* 0 *as* $N \to \infty$.

Proof. The quantity $E\left(\psi_N | m_N^v, m_J^u\right)$ can be computed from the fact that for all $1 \leq k \leq J$, the k'th most desirable firm is occupied by the $2k$-th and the $(2k-1)$-th most desirable workers. By definition, the conditional expectation of $\Psi\left(x_1, x_2, z\right)$ for the k'th desirable job is $\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{k}{J}\right)\right)$ where $F_{N,v}$ and $F_{J,u}$ are the cdfs representing the empirical measures m_N^v and m_J^u respectively:

$$\begin{aligned} E\left(\psi_N | m_N^v, m_J^u\right) &= \frac{1}{J}\sum_{k=1}^{J}\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{k}{J}\right)\right) \\ &= \frac{1}{2J}\sum_{i=1}^{2J}\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{i}{N}\right), F_{N,v}^{-1}\left(\frac{i}{N}\right), F_{J,u}^{-1}\left(\frac{i}{N}\right)\right) + R. \end{aligned} \quad (19)$$

We will show that

$$\frac{1}{2J}\sum_{i=1}^{2J}\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{i}{N}\right), F_{N,v}^{-1}\left(\frac{i}{N}\right), F_{J,u}^{-1}\left(\frac{i}{N}\right)\right) - \psi \to 0 \quad (20)$$

and that $R \to 0$.

First, consider $|R|$ for large N. By the expression above,

$$\begin{aligned} R &= \frac{1}{J}\sum_{k=1}^{J}\Bigg[\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{k}{J}\right)\right) \\ &\quad -\frac{1}{2}\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{J,u}^{-1}\left(\frac{2k-1}{N}\right)\right) \\ &\quad -\frac{1}{2}\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\Bigg]. \end{aligned}$$

By the triangle ineuqality, the absolute value of the k-th term in the summation is at most

$$\begin{aligned} &\frac{1}{2}\Bigg|\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{k}{J}\right)\right) \\ &-\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{J,u}^{-1}\left(\frac{2k-1}{N}\right)\right)\Bigg| \\ &+\frac{1}{2}\Bigg|\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{k}{J}\right)\right) \\ &-\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\Bigg| \end{aligned}$$

Hence,

$$\begin{aligned}|R| &\leq \frac{1}{2J}\sum_{k=1}^{J}\left|\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\right. \\ &\left. -\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\right| \\ &+\frac{1}{2J}\sum_{k=1}^{J}\left|\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\right. \\ &\left. -\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\right|\end{aligned}$$

For $\delta \in \left(0, \frac{1}{2}\right)$,

$$\begin{aligned}|R| &\leq \frac{1}{2J}\sum_{\lceil J\delta\rceil < k < \lfloor J(1-\delta)\rfloor}\left|\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\right. \\ &\left. -\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\right| \\ &+\frac{1}{2J}\sum_{\lceil J\delta\rceil < k < \lfloor J(1-\delta)\rfloor}\left|\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\right. \\ &\left. -\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{N,v}^{-1}\left(\frac{2k-1}{N}\right), F_{J,u}^{-1}\left(\frac{2k}{N}\right)\right)\right| \\ &+4\delta\left\|\tilde{\psi}\right\|_{\infty} \\ &= R_1 + R_2 + 4\delta\left\|\tilde{\psi}\right\|_{\infty} \qquad (21)\end{aligned}$$

By a first order Taylor expansion,

$$R_1+R_2 \leq \sup_{\lceil J\delta\rceil < k < \lfloor J(1-\delta)\rfloor}\left\|\nabla\tilde{\psi}\right\|_{\infty}\left[\left|F_{N,v}^{-1}\left(\frac{2k-1}{N}\right) - F_{N,v}^{-1}\left(\frac{2k}{N}\right)\right| + \left|F_{N,v}^{-1}\left(\frac{2k}{N}\right) - F_{N,v}^{-1}\left(\frac{2k}{N}\right)\right|\right] \qquad (22)$$

and since $R_1, R_2 \geq 0$ and $\left\|\nabla\tilde{\psi}\right\|_{\infty} < \infty$ (Lemma 10), the sum R_1+R_2 converges in probability to 0 by the uniform convergence of the quantile process of v on $[\delta, 1-\delta]$ (Lemmata 9, 14).

Now, consider the difference in equation (20). Note that $F_{J,u}$ is constant on each interval $[\frac{k-1}{J}, \frac{k}{J})$ and $F_{N,v}$ is constant on $[\frac{i-1}{N}, \frac{i}{N})$. Hence,

$$\begin{aligned}
& \frac{1}{2J}\sum_{i=1}^{2J}\tilde{\psi}\left(F_{N,v}^{-1}\left(\frac{i}{N}\right), F_{N,v}^{-1}\left(\frac{i}{N}\right), F_{J,u}^{-1}\left(\frac{i}{N}\right)\right) - \psi \\
= & \int_0^1 \tilde{\psi}\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) dq - \int_0^1 \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right) dq \\
= & \int_\delta^{1-\delta} \left[\tilde{\psi}\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right)\right] dq \\
& + \left(\int_0^\delta + \int_{1-\delta}^1\right) \left[\tilde{\psi}\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right)\right] dq \\
= & T_1 + T_2
\end{aligned} \tag{23}$$

where $\delta \in \left(0, \frac{1}{2}\right)$.

We now bound for T_1 and T_2 in terms of $\delta > 0$. Since $\|\psi\|_\infty \leq \|\Psi\|_\infty < \infty$, $|T_2| \leq 4\delta \left\|\tilde{\psi}\right\|_\infty$. To bound T_1, note that

$$\begin{aligned}
T_1 = & \left|\int_\delta^{1-\delta} \left[\tilde{\psi}\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right)\right] dq\right| \\
\leq & \int_\delta^{1-\delta} \left|\tilde{\psi}\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right)\right| dq
\end{aligned}$$

and for all $q \in [\delta, 1-\delta]$, we have a bound via a pointwise Taylor expansion,

$$\begin{aligned}
T_1 \leq & \left|\tilde{\psi}\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \tilde{\psi}\left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right)\right| \\
\leq & \left\|\nabla\tilde{\psi}\right\|_\infty \sup_{q\in[\delta,1-\delta]} \left\|\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right)\right\|_\infty
\end{aligned} \tag{24}$$

Note that Lemma 10 implies that $\left\|\nabla\tilde{\psi}\right\|_\infty$ exists. Combining equations (19) - (24) and the bound on T_2, we have that

$$\begin{aligned} & \left|E\left(\psi_N|m_N^v, m_J^u\right) - \psi\right| \\ \leq & \left\|\nabla\tilde{\psi}\right\|_\infty \sup_{q\in[\delta,1-\delta]} \left|\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right)\right| + 8\delta\left\|\psi\right\|_\infty \end{aligned}$$

where $8\delta\left\|\psi\right\|_\infty$ is the contribution from equation (21) and the bound on T_2.

We now show that $\left|E\left(\psi_N|m_N^v, m_J^u\right) - \psi\right| \to 0$ in probability as $N \to \infty$. Fix $\epsilon > 0$ and choose $\delta = \frac{\epsilon}{16\|\psi\|_\infty}$. Since

$$\sup_{q\in[\delta,1-\delta]} \left|\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right)\right|$$

converges in probability to 0, for sufficiently large N we have by Lemma 9,

$$P\left(\left\|\nabla\tilde{\psi}\right\|_\infty \sup_{Q'\epsilon[\delta,1-\delta]} \left|\left(F_{N,v}^{-1}(q), F_{N,v}^{-1}(q), F_{J,u}^{-1}(q)\right) - \left(F_v^{-1}(q), F_v^{-1}(q), F_u^{-1}(q)\right)\right| > \frac{\epsilon}{2}\right) < \epsilon.$$

This implies $P\left(\left|E\left(\psi_N|m_N^v, m_J^u\right) - \psi\right| > \epsilon\right) < \epsilon$, proving the desired convergence in probability to 0. ■

Lemma 12 *$\psi_N - E\left(\psi_N|m_N^v, m_J^u\right)$ converges in probability to 0 if $\|\Psi\|_\infty < \infty$.*

Proof. Let $v^{(k)}$ and $u^{(k)}$ be k'th order statistics of worker and firm desirability and let $X^{(k)}$ and $Z^{(k)}$ be the corresponding observations drawn from $m^{x|v^{(k)}}$ and $m^{z|u^{(k)}}$ respectively. Note that the second moment of the function

$$\psi_N - E\left(\psi_N|m_N^u, m_N^v\right) = \frac{1}{J}\left(\sum_{k=1}^{J} \Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(i)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right)\right).$$

conditional on (m_N^v, m_J^u) is

$$
\begin{aligned}
& \frac{1}{J^2} E\left(\sum_{i=1}^{J} \Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) | m_N^v, m_J^u\right)^2 \\
= & \frac{1}{J^2} \sum_{i=1}^{J} E\left(\Psi\left(X^{(2k-1)}, X^{(2k)}, Z^{(k)}\right) - \tilde{\psi}\left(v^{(2k-1)}, v^{(2k)}, u^{(k)}\right) | m_N^v, m_J^u\right)^2 \\
\leq & \frac{1}{J}\left\|\sigma^2\right\|_\infty,
\end{aligned}
$$

where the first equality follows from independence and $\left\|\sigma^2\right\|_\infty$ is defined as the supremum of the function

$$
\sigma^2(v_1, v_2, u) = Var\left(\Psi(X_1, X_2, Z) | h(X_1; \theta_0) + \varepsilon_1 = v_1, h(X_2; \theta_0) + \varepsilon_2 = v_2, g(Z; \theta_0) + \eta = u\right).
$$

The quantity $\left\|\sigma^2\right\|_\infty$ is well defined and finite since $\Psi(x_1, x_2, z)$ is bounded.

However, since $\psi_N - E(\psi_N | m_N^v, m_J^u)$ is by definition mean zero, it follows that the unconditional variance of $\psi_N - E(\psi_N | m_N^v, m_J^u)$ is bounded above by $\frac{1}{J}\left\|\sigma^2\right\|_\infty$, by the law of total variance.. This proves $\sqrt{N}\left(\psi_N - E(\psi_N | m_N^v, m_J^u)\right) = O_p(1)$ and thus, by Chevyshev's inequality, $\psi_N - E(\psi_N | m_N^v, m_J^u) = o_p(1)$. ■

Proposition 3 *Let ψ^k and ψ_N^k denote the k-th dimensions of ψ and ψ_N respectively. If Assumption 5 is satisfied, then for each $k \in \{1, \ldots, K_\Psi\}$, $\psi_N^k - \psi^k$ converges in probability to 0.*

Proof. The proof is identical for each component k. For notational simplicity we drop the index k. Since $\psi_N - \psi = (\psi_N - E(\psi_N | m_N^v, m_J^u)) + (E(\psi_N | m_N^v, m_J^u) - \psi)$ is the sum of two terms that converge in probability to 0, this follows directly from Slutsky's theorem. ■

B.3 Proof of Proposition 4

For ease of notation, define the quantities

$$\begin{aligned}
\varepsilon(q,x;\theta) &= F_{v;\theta}^{-1}(q) - h(x;\theta) \\
\varepsilon_N(q,x;\theta) &= F_{N,u;\theta}^{-1}(q) - h(x;\theta) \\
\eta(q,z;\theta) &= F_{u;\theta}^{-1}(q) - g(z;\theta) \\
\eta_J(q,z;\theta) &= F_{J,u;\theta}^{-1}(q) - g(z;\theta).
\end{aligned}$$

We first prove three preliminary results

Lemma 13 *If Assumption 5(i) - (iii) are satisfied, then for each* $\delta \in \left(0, \frac{1}{2}\right)$, *the quantity*

$$T_{1,\delta} = \sup_{\theta} \int_{\delta}^{1-\delta} \left| t_1(q; m^x, m^z) - t_{N,1}(q; m_N^x, m_J^z) \right| dq$$

where

$$\begin{aligned}
t_1(q; m^x, m^z, \theta) &= \int \Psi(X_1, X_2, Z) f_\varepsilon(\varepsilon(q, X_1; \theta)) f_\varepsilon(\varepsilon(q, X_2; \theta)) f_\eta(\eta(q, Z; \theta)) dm^x dm^x dm^z \\
t_{N,1}(q; m^x, m^z, \theta) &= \int \Psi(X_1, X_2, Z) f_\varepsilon(\varepsilon_N(q, X_1; \theta)) f_\varepsilon(\varepsilon_N(q, X_2; \theta)) f_\eta(\eta_J(q, Z; \theta)) dm^x dm^x dm^z
\end{aligned}$$

converges in outer probability to 0.

Proof. We first split $T_{1,\delta}$ into two terms R_1 and R_2 as follows:

$$\begin{aligned}
& \sup_{\theta} \int_{\delta}^{1-\delta} \left| t_1(q; m^x, m^z, \theta) - t_{N,1}(q; m_N^x, m_N^z, \theta) \right| dq \\
\leq\ & \sup_{\theta} \int_{\delta}^{1-\delta} \left| t_1(q; m^x, m^z, \theta) - t_{N,1}(q; m^x, m^z, \theta) \right| dq \\
& + \sup_{\theta} \int_{\delta}^{1-\delta} \left| t_{N,1}(q; m^x, m^z, \theta) - t_{N,1}(q; m_N^x, m_J^z, \theta) \right| dq \\
=\ & R_1 + R_2
\end{aligned}$$

Note that R_2 is bounded in absolute value by

$$\begin{aligned}
& \sup_{\theta,\ q\in[\delta,1-\delta]} && \left|t_{N,1}\left(q;m^{x},m^{z},\theta\right)-t_{N,1}\left(q;m^{x}_{N},m^{z}_{J},\theta\right)\right| \\
= & \sup_{\theta,\ v\in[v_1,\ v_2],\ u\in[u_1,\ u_2]} && \left|\int \Psi\left(X_1,X_2,Z\right) f_{\varepsilon}\left(u-h\left(X_1;\theta\right)\right) f_{\varepsilon}\left(u-h\left(X_2;\theta\right)\right) f_{\eta}\left(v-g\left(Z;\theta\right)\right)\right. \\
& && \left.\left(dm^{x}dm^{x}dm^{z}-dm^{x}_{N}dm^{x}_{N}dm^{z}_{J}\right)\right|
\end{aligned}$$

where $[u_1,\ u_2]=\left[F^{-1}_{N,v;\theta}\left(\delta\right),F^{-1}_{N,v;\theta}\left(1-\delta\right)\right]$ and $[v_1,\ v_2]=\left[F^{-1}_{J,u;\theta}\left(\delta\right),F^{-1}_{J,u;\theta}\left(1-\delta\right)\right]$. The expression above converges in outer probability to 0 by the Glivenko-Cantelli theorem:

$$\left\{\Psi\left(X_1,X_2,Z\right) f_{\varepsilon}\left(u-h\left(x;\theta\right)\right) f_{\varepsilon}\left(u-h\left(x;\theta\right)\right) f_{\eta}\left(v-g\left(z;\theta\right)\right)\right\}_{u\in[u_1,\ u_2],\ v\in[v_1,v_2],\ \theta\in\Theta}$$

is a Glivenko-Cantelli class for all u_1,u_2,v_1,v_2 since

$$\Psi\left(X_1,X_2,Z\right) f_{\varepsilon}\left(u-h\left(x;\theta\right)\right) f_{\varepsilon}\left(u-h\left(x;\theta\right)\right) f_{\eta}\left(v-g\left(z;\theta\right)\right)$$

is uniformly Lipschitz in $u,\ v$, and θ, and $\Theta\times[u_1,\ u_2]\times[v_1,\ v_2]$ is compact.

To bound R_1, note that

$$\begin{aligned}
& \sup_{\theta}\int_{\delta}^{1-\delta}\left|t_1\left(q;m^{x},m^{z}\right)-t_{N,1}\left(q;m^{x},m^{z}\right)\right|dq \\
= & \int_{\delta}^{1-\delta}\left|\int \Psi\left(X_1,X_2,Z\right)\left[f_{\varepsilon}\left(\varepsilon\left(q,X_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,X_2;\theta\right)\right) f_{\eta}\left(\eta\left(q,Z;\theta\right)\right)\right.\right. \\
& \left.\left.-f_{\varepsilon}\left(\varepsilon_N\left(q,X_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon_N\left(q,X_2;\theta\right)\right) f_{\eta}\left(\eta_J\left(q,Z;\theta\right)\right)\right] dm^{x}dm^{x}dm^{z}dq\right| \\
\leq & \left\|\Psi\right\|_{\infty}\sup_{\theta,q\epsilon[\delta,1-\delta]}\left|f_{\varepsilon}\left(\varepsilon\left(q,X_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,X_2;\theta\right)\right) f_{\eta}\left(\eta\left(q,Z;\theta\right)\right)-\right. \\
& \left. f_{\varepsilon}\left(\varepsilon_N\left(q,X_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon_N\left(q,X_2;\theta\right)\right) f_{\eta}\left(\eta_J\left(q,Z;\theta\right)\right)\right|.
\end{aligned}$$

We now show that

$$\sup_{\theta,q\epsilon[\delta,1-\delta]}\left|f_{\varepsilon}\left(\varepsilon\left(q,X_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,X_2;\theta\right)\right) f_{\eta}\left(\eta\left(q,Z;\theta\right)\right)-f_{\varepsilon}\left(\varepsilon_N\left(q,X_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon_N\left(q,X_2;\theta\right)\right) f_{\eta}\left(\eta_J\left(q,Z;\theta\right)\right)\right|$$

converges to 0 in outer probability. First rewrite the difference

$$
\begin{aligned}
&f_{\varepsilon}\left(\varepsilon\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,x_2;\theta\right)\right) f_{\eta}\left(\eta\left(q,z;\theta\right)\right) - f_{\varepsilon}\left(\varepsilon_N\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon_N\left(q,x_2;\theta\right)\right) f_{\eta}\left(\eta_J\left(q,z;\theta\right)\right) \\
&\quad = f_{\varepsilon}\left(\varepsilon\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,x_2;\theta\right)\right) \left[f_{\eta}\left(\eta\left(q,z;\theta\right)\right) - f_{\eta}\left(\eta_J\left(q,z;\theta\right)\right)\right] \\
&+ f_{\eta}\left(\eta\left(q,z;\theta\right)\right) \left[f_{\varepsilon}\left(\varepsilon\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,x_2;\theta\right)\right) - f_{\varepsilon}\left(\varepsilon_N\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon_N\left(q,x_2;\theta\right)\right)\right] \\
&+ \left[f_{\eta}\left(\eta_J\left(q,z;\theta\right)\right) - f_{\eta}\left(\eta\left(q,z;\theta\right)\right)\right] \left[f_{\varepsilon}\left(\varepsilon\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,x_2;\theta\right)\right) - f_{\varepsilon}\left(\varepsilon_N\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon_N\left(q,x_2;\theta\right)\right)\right].
\end{aligned}
$$

Hence, by the triangle inequality,

$$
\begin{aligned}
&\left|f_{\varepsilon}\left(\varepsilon\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,x_2;\theta\right)\right) f_{\eta}\left(\eta\left(q,z;\theta\right)\right) - f_{\varepsilon}\left(\varepsilon_N\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon_N\left(q,x_2;\theta\right)\right) f_{\eta}\left(\eta_J\left(q,z;\theta\right)\right)\right| \\
&\quad \leq f_{\varepsilon}\left(\varepsilon\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,x_2;\theta\right)\right) \left|f_{\eta}\left(\eta\left(q,z;\theta\right)\right) - f_{\eta}\left(\eta_J\left(q,z;\theta\right)\right)\right| \\
&\quad + f_{\eta}\left(\eta\left(q,z;\theta\right)\right) \left|f_{\varepsilon}\left(\varepsilon\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,x_2;\theta\right)\right) - f_{\varepsilon}\left(\varepsilon_N\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon_N\left(q,x_2;\theta\right)\right)\right| \\
&\quad + \left|f_{\eta}\left(\eta_J\left(q,z;\theta\right)\right) - f_{\eta}\left(\eta\left(q,z;\theta\right)\right)\right| \left|f_{\varepsilon}\left(\varepsilon\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,x_2;\theta\right)\right) - f_{\varepsilon}\left(\varepsilon_N\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon_N\left(q,x_2;\theta\right)\right)\right|
\end{aligned}
$$

Further, since f_{ε} and f_{η} are bounded,

$$
\begin{aligned}
&\left|f_{\varepsilon}\left(\varepsilon\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,x_2;\theta\right)\right) f_{\eta}\left(\eta\left(q,z;\theta\right)\right) - f_{\varepsilon}\left(\varepsilon_N\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon_N\left(q,x_2;\theta\right)\right) f_{\eta}\left(\eta_N\left(q,z;\theta\right)\right)\right| \\
\leq\quad &\left\|f_{\varepsilon}\right\|_{\infty}^{2} \left|f_{\eta}\left(\eta\left(q,z;\theta\right)\right) - f_{\eta}\left(\eta_N\left(q,z;\theta\right)\right)\right| \\
&+ \left\|f_{\eta}\right\|_{\infty} \left|f_{\varepsilon}\left(\varepsilon\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,x_2;\theta\right)\right) - f_{\varepsilon}\left(\varepsilon_N\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon_N\left(q,x_2;\theta\right)\right)\right| \\
&+ 2\left\|f_{\eta}\right\|_{\infty} \left|f_{\varepsilon}\left(\varepsilon\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon\left(q,x_2;\theta\right)\right) - f_{\varepsilon}\left(\varepsilon_N\left(q,x_1;\theta\right)\right) f_{\varepsilon}\left(\varepsilon_N\left(q,x_2;\theta\right)\right)\right| \\
&\quad \leq \left\|f_{\varepsilon}\right\|_{\infty}^{2} \left\|f_{\eta}'\right\|_{\infty} \left|\eta\left(q,z;\theta\right) - \eta_N\left(q,z;\theta\right)\right| \\
&+ 3\left\|f_{\eta}\right\|_{\infty} \left\|f_{\varepsilon}\right\|_{\infty} \left\|f_{\varepsilon}'\right\|_{\infty} \left|\varepsilon\left(q,x_1;\theta\right) - \varepsilon_N\left(q,x_1;\theta\right)\right|
\end{aligned}
$$

where the last inequality follows from a Taylor expansion. Hence, we have that

$$\begin{aligned}
& \|\Psi\|_\infty \sup_{\theta,q\epsilon[\delta,1-\delta]} \left|f_\varepsilon\left(\varepsilon\left(q,x_1;\theta\right)\right) f_\varepsilon\left(\varepsilon\left(q,x_2;\theta\right)\right) f_\eta\left(\eta\left(q,z;\theta\right)\right)\right. \\
& \left.- f_\varepsilon\left(\varepsilon_N\left(q,x_1;\theta\right)\right) f_\varepsilon\left(\varepsilon_N\left(q,x_2;\theta\right)\right) f_\eta\left(\eta_N\left(q,z;\theta\right)\right)\right| \\
\leq\ & \|\Psi\|_\infty \|f_\varepsilon\|^2 \left\|f_\eta'\right\| \sup_{\theta\in\Theta} \left|\eta\left(q,z;\theta\right) - \eta_N\left(q,z;\theta\right)\right| \\
& +3\|\Psi\|_\infty \|f_\eta\| \|f_\varepsilon\|_\infty \|f_\varepsilon'\| \sup_{\theta\in\Theta} \left|\varepsilon\left(q,x_1;\theta\right) - \varepsilon_N\left(q,x_1;\theta\right)\right|.
\end{aligned}$$

By Lemma 9, $\sup_{\theta\in\Theta,q\in[\delta,1-\delta]} \left|\varepsilon\left(q,x;\theta\right) - \varepsilon_N\left(q,x;\theta\right)\right|$ and $\sup_{\theta\in\Theta,q\in[\delta,1-\delta]} \left|\eta\left(q,z;\theta\right) - \eta_N\left(q,z;\theta\right)\right|$ converge in outer probability to 0. Hence, $|R_1|$ is bounded by a function that converges in outer probability to 0.

Since $T_{1,\delta}$ is bounded above by the sum of elements which converge in outer probability to 0, it does so as well. ■

Lemma 14 *If Assumptions 5(ii) and (iii) are satisfied, then for every* $\delta \in \left(0, \frac{1}{2}\right)$ *and* $q \in [\delta, 1-\delta]$*, the quantities*

$$\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right) dm^x \ \text{and} \ \int f_\eta\left(\eta\left(q,Z;\theta\right)\right) dm^z$$

are bounded away from 0*, uniformly in* θ*. In particular,*

$$\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right) f_\varepsilon\left(\varepsilon\left(q,X_2;\theta\right)\right) f_\eta\left(\eta\left(q,Z;\theta\right)\right) dm^x dm^x dm^z$$

is also bounded away from zero.

Proof. Note that

$$\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right) f_\varepsilon\left(\varepsilon\left(q,X_2;\theta\right)\right) f_\eta\left(\eta\left(q,Z;\theta\right)\right) dm^x dm^x dm^z = f_{v;\theta}\left(F_{v;\theta}^{-1}(q)\right) f_{v;\theta}\left(F_{v;\theta}^{-1}(q)\right) f_{u;\theta}\left(F_{u;\theta}^{-1}(q)\right)$$

where $f_{u;\theta}\left(F_{u;\theta}^{-1}(q)\right)$ and $f_{v;\theta}\left(F_{v;\theta}^{-1}(q)\right)$ are the densities of u and v at their q-th quantiles respectively. Assumptions 5(ii) and (iii) require that f_ε and f_η are continuous and strictly positive, hence they bounded away from zero on any compact set. Consequently, $f_{u;\theta}$ and $f_{v;\theta}$ are also bounded away from zero on any compact set. Since $F_{u;\theta}^{-1}$ is jointly continuous, the image $F_{u;\theta}^{-1}([\delta, 1-\delta])$ is compact and $f_{v;\theta}(F_v^{-1}(q))\, f_{v;\theta}(F_v^{-1}(q))\, f_{u;\theta}(F_u^{-1}(q))$ is bounded away from 0 for all $q \in [\delta, 1-\delta]$. ■

Lemma 15 *If Assumptions 5(ii) and (iii) are satisfied, the quantity*

$$\begin{aligned} T_{2,\delta} &= \sup_{\theta\in\Theta} \int_{\delta}^{1-\delta} \left| \frac{1}{\int f_\varepsilon(\varepsilon(q, X_1;\theta))\, f_\varepsilon(\varepsilon(q, X_2;\theta))\, f_\eta(\eta(q, Z;\theta))\, dm^x dm^x dm^z} \right. \\ &\left. -\frac{1}{\int f_\varepsilon(\varepsilon_N(q, X_1;\theta))\, f_\varepsilon(\varepsilon_N(q, X_2;\theta))\, f_\eta(\eta_J(q, Z;\theta))\, dm_N^x dm_N^x dm_J^z} \right| dq \end{aligned}$$

converges in outer probability to 0.

Proof. As in proof of Lemma 13,

$$\begin{aligned} L_N &= \sup_{\theta, q\in[\delta,1-\delta]} \left| \int f_\varepsilon(\varepsilon(q, X_1;\theta))\, f_\varepsilon(\varepsilon(q, X_2;\theta))\, f_\eta(\eta(q, Z;\theta))\, dm^x dm^x dm^z \right. - \\ &\left. \int f_\varepsilon(\varepsilon_N(q, X_1;\theta))\, f_\varepsilon(\varepsilon_N(q, X_2;\theta))\, f_\eta(\eta_N(q, Z;\theta)) \right| dm_N^x dm_N^x dm_J^z \end{aligned}$$

converges in outer probability to 0. Since $\int f_\varepsilon(\varepsilon(q, X_1;\theta))\, f_\varepsilon(\varepsilon(q, X_2;\theta))\, f_\eta(\eta(q, Z;\theta))\, dm^x dm^x dm^z$ is bounded away from 0 over $q \in [\delta, 1-\delta]$ and all θ (Lemma 14), a tailor expansion of the function $1/x$ implies that for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$\begin{aligned} &L_N < \delta \\ &\Rightarrow \sup_{\theta, q\epsilon[\delta,1-\delta]} \left| \frac{1}{\int f_\varepsilon(\varepsilon(q, X_1;\theta))\, f_\varepsilon(\varepsilon(q, X_2;\theta))\, f_\eta(\eta(q, Z;\theta))\, dm^x dm^x dm^z} - \right. \\ &\left. \frac{1}{\int f_\varepsilon(\varepsilon_N(q, X_1;\theta))\, f_\varepsilon(\varepsilon_N(q, X_2;\theta))\, f_\eta(\eta_N(q, Z;\theta))\, dm_N^x dm_N^x dm_J^z} \right| < \epsilon. \end{aligned}$$

Therefore, $T_{2,\delta}$ converges in outer probability to 0. ■

Proposition 4 *Let $\psi^k(\theta)$ and $\psi_N^k(\theta; m_N^x, m_J^z)$ denote the k-th dimensions of $\psi(\theta)$ and $\psi_N(\theta; m_N^x, m_J^z)$ respectively. If Assumptions 5(i) - (iii) are satisfied, then for each $k \in$*

$\{1,\ldots,K_\Psi\}$, $\left|\psi^k(\theta)-\psi_N^k(\theta;m_N^x,m_J^z)\right|$ *converges in outer probability to* 0 *uniformly in* θ.

Proof. The proof is identical for each dimention of $\psi^k(\theta)$ and $\psi_N^k(\theta;m_N^x,m_J^z)$. For notational simplicity, we drop the index for the dimension k.

Rewrite

$$\psi(\theta)-\psi_N(\theta;m_N^x,m_N^z)=\int_0^1 t(q,\theta;m_N^x,m_N^z)\,dq$$

where

$$\begin{aligned}t(q,\theta;m_N^x,m_N^z) &= \frac{\int\Psi(X_1,X_2,Z)\,f_\varepsilon(\varepsilon(q,X_1;\theta))\,f_\varepsilon(\varepsilon(q,X_2;\theta))\,f_\eta(\eta(q,Z;\theta))\,dm^x dm^x dm^z}{\int f_\varepsilon(\varepsilon(q,X_1;\theta))\,f_\varepsilon(\varepsilon(q,X_2;\theta))\,f_\eta(\eta(q,Z;\theta))\,dm^x dm^x dm^z}\\ &\quad-\frac{\int\Psi(X_1,X_2,Z)\,f_\varepsilon(\varepsilon_N(q,X_1;\theta))\,f_\varepsilon(\varepsilon_N(q,X_2;\theta))\,f_\eta(\eta_J(q,Z;\theta))\,dm_N^x dm_N^x dm_J^z}{\int f_\varepsilon(\varepsilon_N(q,X_1;\theta))\,f_\varepsilon(\varepsilon_N(q,X_2;\theta))\,f_\eta(\eta_J(q,Z;\theta))\,dm_N^x dm_N^x dm_J^z}.\end{aligned}$$

For $\delta\in\left(0,\frac{1}{2}\right)$, we have that

$$\begin{aligned}\psi(\theta)-\psi_N(\theta;m_N^x,m_N^z) &= \int_\delta^{1-\delta} t(q,\theta;m_N^x,m_N^z)\,dq+\left(\int_0^\delta+\int_{1-\delta}^1\right)t(q,\theta;m_N^x,m_N^z)\,dq\\ &= R_1+R_2+\left(\int_0^\delta+\int_{1-\delta}^1\right)t(q,\theta;m_N^x,m_N^z)\,dq \qquad (25)\end{aligned}$$

where

$$\begin{aligned}R_1 &= \int_\delta^{1-\delta}\frac{\int\Psi(X_1,X_2,Z)\,f_\varepsilon(\varepsilon(q,X_1;\theta))\,f_\varepsilon(\varepsilon(q,X_2;\theta))\,f_\eta(\eta(q,Z;\theta))\,dm^x dm^x dm^z}{\int f_\varepsilon(\varepsilon(q,X_1;\theta))\,f_\varepsilon(\varepsilon(q,X_2;\theta))\,f_\eta(\eta(q,Z;\theta))\,dm^x dm^x dm^z}dq\\ &\quad-\int_\delta^{1-\delta}\frac{\int\Psi(X_1,X_2,Z)\,f_\varepsilon(\varepsilon_N(q,X_1;\theta))\,f_\varepsilon(\varepsilon_N(q,X_2;\theta))\,f_\eta(\eta_J(q,Z;\theta))\,dm_N^x dm_N^x dm_J^z}{\int f_\varepsilon(\varepsilon(q,X_1;\theta))\,f_\varepsilon(\varepsilon(q,X_2;\theta))\,f_\eta(\eta(q,Z;\theta))\,dm^x dm^x dm^z}dq\end{aligned}$$

and

$$\begin{aligned}R_2 &= \int_\delta^{1-\delta}\frac{\int\Psi(X_1,X_2,Z)\,f_\varepsilon(\varepsilon_N(q,X_1;\theta))\,f_\varepsilon(\varepsilon_N(q,X_2;\theta))\,f_\eta(\eta_J(q,Z;\theta))\,dm_N^x dm_N^x dm_J^z}{\int f_\varepsilon(\varepsilon(q,X_1;\theta))\,f_\varepsilon(\varepsilon(q,X_2;\theta))\,f_\eta(\eta(q,Z;\theta))\,dm^x dm^x dm^z}dq\\ &\quad-\int_\delta^{1-\delta}\frac{\int\Psi(X_1,X_2,Z)\,f_\varepsilon(\varepsilon_N(q,X_1;\theta))\,f_\varepsilon(\varepsilon_N(q,X_2;\theta))\,f_\eta(\eta_J(q,Z;\theta))\,dm_N^x dm_N^x dm_J^z}{\int f_\varepsilon(\varepsilon_N(q,X_1;\theta))\,f_\varepsilon(\varepsilon_N(q,X_2;\theta))\,f_\eta(\eta_J(q,Z;\theta))\,dm_N^x dm_N^x dm_J^z}dq.\end{aligned}$$

Hence, by the triangle inequality,

$$\left|\psi(\theta)-\psi_N(\theta;m_N^x,m_N^z)\right|\leq|R_1|+|R_2|+\left|\left(\int_0^\delta+\int_{1-\delta}^1\right)t(q,\theta;m_N^x,m_N^z)\,dq\right|. \tag{26}$$

We now bound each of the terms. Note that the third term is bounded since M is:

$$\left|\left(\int_0^\delta+\int_{1-\delta}^1\right)t(q,\theta;m_N^x,m_N^z)\,dq\right|\leq 4\delta\left\|M\right\|_\infty. \tag{27}$$

The term R_1 is bounded in terms of $T_{1,\delta}$ defined in Lemma 13:

$$\begin{aligned}
|R_1| &\leq \sup_{\theta,q\in[\delta,1-\delta]}\frac{1}{\int f_\varepsilon(\varepsilon(q,X_1;\theta))\,f_\varepsilon(\varepsilon(q,X_2;\theta))\,f_\eta(\eta(q,Z;\theta))\,dm^x dm^x dm^z}\\
&\quad\times\int_\delta^{1-\delta}\left|\int\Psi(X_1,X_2,Z)\,f_\varepsilon(\varepsilon(q,X_1;\theta))\,f_\varepsilon(\varepsilon(q,X_2;\theta))\,f_\eta(\eta(q,Z;\theta))\,dm^x dm^x dm^z\right.\\
&\quad\left.-\int\Psi(X_1,X_2,Z)\,f_\varepsilon(\varepsilon_N(q,X_1;\theta))\,f_\varepsilon(\varepsilon_N(q,X_2;\theta))\,f_\eta(\eta_J(q,Z;\theta))\,dm_N^x dm_J^z\right|dq\\
&= \sup_{\theta,q\varepsilon[\delta,1-\delta]}\frac{1}{\int f_\varepsilon(\varepsilon(q,X_1;\theta))\,f_\varepsilon(\varepsilon(q,X_2;\theta))\,f_\eta(\eta(q,Z;\theta))\,dm^x dm^x dm^z}T_{1,\delta}
\end{aligned} \tag{28}$$

and finally, the term R_2 is bounded in terms of $T_{2,\delta}$ defined in Lemma 15:

$$\begin{aligned}
|R_2| &\leq \left\|\Psi\right\|_\infty\int_\delta^{1-\delta}\left|\frac{1}{\int f_\varepsilon(\varepsilon(q,X_1;\theta))\,f_\varepsilon(\varepsilon(q,X_2;\theta))\,f_\eta(\eta(q,Z;\theta))\,dm^x dm^x dm^z}\right.\\
&\quad\left.-\frac{1}{\int f_\varepsilon(\varepsilon_N(q,X_1;\theta))\,f_\varepsilon(\varepsilon_N(q,X_2;\theta))\,f_\eta(\eta_J(q,Z;\theta))\,dm_N^x dm_N^x dm_J^z}\right|dq\\
&= \left\|\Psi\right\|_\infty T_{2,\delta}
\end{aligned} \tag{29}$$

For any ϵ, δ can be chosen to make $4\delta\left\|\Psi\right\|_\infty\leq\epsilon$ without reference to θ. Hence, we need to show that the terms $|R_1|$ and $|R_2|$ can also be made small enough, uniformly in θ. By equations (26), (27), (28) and (29),

$$
\begin{aligned}
&\sup_{\theta\in\Theta}\left|\psi\left(\theta\right)-\psi_N\left(\theta;m_N^x,m_N^z\right)\right| \\
\leq\; &\left(\sup_{\theta,q\in[\delta,1-\delta]}\frac{1}{\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right)f_\varepsilon\left(\varepsilon\left(q,X_2;\theta\right)\right)f_\eta\left(\eta\left(q,Z;\theta\right)\right)dm^x dm^x dm^z}\right)T_{1,\delta} \\
&+\left\|\Psi\right\|_\infty T_{2,\delta}+4\delta\left\|\Psi\right\|_\infty,
\end{aligned}
$$

where $T_{1,\delta}$ and $T_{2,\delta}$ each converge in outer probability to 0.

The term

$$\sup_{\theta,q\in[\delta,1-\delta]}\frac{1}{\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right)f_\varepsilon\left(\varepsilon\left(q,X_2;\theta\right)\right)f_\eta\left(\eta\left(q,Z;\theta\right)\right)dm^x dm^x dm^z}$$

is bounded since $\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right)f_\varepsilon\left(\varepsilon\left(q,X_2;\theta\right)\right)f_\eta\left(\eta\left(q,Z;\theta\right)\right)dm^x dm^x dm^z$ is bounded away from 0 (Lemma 14).

Fix $\epsilon>0$, and pick $\delta=\frac{\epsilon}{8\|\Psi\|_\infty}$. For sample size N sufficiently large, we have that

$$P\left(T_{1,\delta}>\left(\sup_{\theta,q\in[\delta,1-\delta]}\frac{4}{\int f_\varepsilon\left(\varepsilon\left(q,X_1;\theta\right)\right)f_\varepsilon\left(\varepsilon\left(q,X_2;\theta\right)\right)f_\eta\left(\eta\left(q,Z;\theta\right)\right)dm^x dm^x dm^z}\right)^{-1}\epsilon\right)<\epsilon$$

and $P\left(T_{2,\delta}>\frac{\epsilon}{4\|\Psi\|_\infty}\right)<\epsilon$. This implies that

$$P\left(\sup_{\theta}\left|\psi\left(\theta\right)-\psi_N\left(\theta;m_N^x,m_J^z\right)\right|>\epsilon\right)<\epsilon$$

proving the desired uniform convergence in probability. ■

B.4 Proof of Theorem 4

Fix $\epsilon>0$, and choose δ such that $\left\|\psi\left(\theta\right)-\psi\left(\theta_0\right)\right\|_W<\frac{\delta}{2}\Rightarrow\left\|\theta-\theta_0\right\|<\epsilon$.

By Propositions 3 and 4 and the continuous mapping theorem, $\left\|\psi_N-\psi_N\left(\theta\right)\right\|_W$ converges in (outer) probability to $\left\|\psi-\psi\left(\theta\right)\right\|_W$ uniformly in θ.

Note that $\|\psi - \psi(\theta_0)\|_W = 0$. It follows that for sufficiently large N,

$$P\left(\sup_{\theta\in\Theta} |\|\psi_N - \psi_N(\theta)\|_W - \|\psi - \psi(\theta)\|_W| > \frac{\delta}{2}\right) < \epsilon,$$

so with probability at least $1-\epsilon$,

$$\begin{aligned}\inf_{\theta\in\Theta} \left\|\psi_N - \psi_N\left(\hat{\theta}_N\right)\right\|_W &\leq \inf_{\theta\in\Theta} \|\psi - \psi(\theta)\|_W + \frac{\delta}{2} \\ \left\|\psi_N - \psi_N\left(\hat{\theta}_N\right)\right\|_W &\leq \frac{\delta}{2}\end{aligned}$$

and

$$\left\|\psi - \psi\left(\hat{\theta}_N\right)\right\|_W \leq \left\|\psi_N - \psi_N\left(\widehat{\theta}_N\right)\right\|_W + \frac{\delta}{2}.$$

Hence,

$$\left\|\psi - \psi\left(\hat{\theta}_N\right)\right\|_W \leq \delta$$

However, by Assumption 6, this implies $\left\|\widehat{\theta}_N - \theta_0\right\| < \epsilon$. It follows that $\hat{\theta}_N$ converges in probability to θ_0, proving consistency of the estimator.

C Parametric Boostrap

Let $\{z_j\}_{j=1}^J$ be a sample of firm characteristics and $\{x_i\}_{i=1}^N$ denote a sample of worker characteristics. The parametric bootstrap for the estimate $\hat{\theta} = \arg\min_{\theta\in\Theta} \hat{Q}_N(\theta)$ is constructed by the following procedure for $b = \{1, \ldots, 500\}$

1. Sample J firms with replacement from the empirical sample $\{z_j\}_{j=1}^J$. Denote this sample with $\left\{z_j^b\right\}_{j=1}^J$.

2. Draw N^b workers with replacement from the empirical sample $\{x_i\}_{i=1}^N$, where $N^b = \sum c_j^b$ and c_j^b is capacity of the j-th sampled firm in the bootstrap sample.

3. Simulate the unobservables ε_j^b and η_i^b.

4. Compute the quantities v_i^b and u_j^b at $\hat{\theta}$ from equations (14a) and (14b). For the model with preference heterogeneity, compute u_{ij}^b as in equation (15b).

5. Compute a pairwise stable match for the bootstrap sample.

6. Compute $\hat{\theta}_b = \arg\min_{\theta \in \Theta} \hat{Q}_N^b(\theta)$ using the bootstrap pairwise stable match and an independent set of simulations for $\hat{Q}_N^b(\theta)$.