Latent Indices in Assortative Matching Models*

William Diamond Nikhil Agarwal

February 21, 2017

Abstract

A large class of two-sided matching models that include both transferable and non-transferable utility result in positive assortative matching along a latent index. Data from matching markets, however, may not exhibit perfect assortativity due to the presence of unobserved characteristics. This paper studies the identification and estimation of such models. We show that the distribution of the latent index is not identified when data from one-to-one matches are observed. Remarkably, the model is non-parametrically identified using data in a single large market when each agent on one side has at least two matched partners. The additional empirical content in many-to-one matches can be illustrated using simulations and stylized examples. We then derive asymptotic properties of a minimum distance estimator as the size of the market increases, allowing estimation using dependent data from a single large matching market. The nature of the dependence requires modification of existing empirical process techniques to obtain a limit theorem.

Keywords: Matching, Identification, Estimation

JEL Codes: C78, C51

*Previously circulated as "Identification and Estimation in Two-Sided Matching Markets," *Cowles Foundation Discussion Paper No. 1905, Aug 2013 (rev. Feb 2014).* We thank Susan Athey, Gary Chamberlain, Victor Chernozhukov, Raj Chetty, Kirill Evdokimov, Jeremy Fox, Alfred Galichon, Phil Haile, Whitney Newey, Ariel Pakes, Parag Pathak, and Paulo Somaini for helpful discussions. Agarwal thanks the Cowles Foundation for Economic Research and the Department of Economics, Yale University for their generous hospitality. Agarwal: Department of Economics, Massachusetts Institute of Technology and NBER, email: `agarwaln@mit.edu`. Diamond: Department of Economics, Harvard University, email: `wdiamond@hbs.edu`.

1 Introduction

Assortative matching along a variety of dimensions has been well documented in many matching markets. There has been growing interest in estimating the underlying preferences that generate these patterns.[1] This is an important step for quantitatively evaluating economic questions such as equilibrium effects of policy interventions or changes in market structure. However, a researcher often has access only to data on matches instead of direct information on preferences and only on a limited set of characteristics. Unobserved characteristics result in deviations from the central assortative tendency observed in the data, and they can be important in understanding the distribution of preferences.

We study the identification and estimation of preferences in a large matching market in which the attractiveness of agents to the other side of the market can be summarized using a single dimensional index that aggregates an unobserved characteristic and multidimensional observed characteristics. We assume that the matching is positive assortative along this latent index. The positive assortative match is the unique pairwise stable match if utility is non-transferable, but also if utility is transferable and the total surplus is supermodular. While the single index model is canonical in the theoretical literature (see Becker, 1973), it is clearly restrictive as it rules out heterogeneity in preferences. At the cost of this restriction, compared to the large body of empirical work following Choo and Siow (2006), we present a non-parametric approach to identification that not only allows for unobserved agent characteristics that are valued by the other side,[2] but that is also agnostic about whether utility is transferable. This single-index assumption has been useful in empirical analyses of the marriage market (see Chiappori et al., 2009, 2012, for example). The model may also provide an approximation in labor or education markets in which workers or students are primarily differentiated by skill and firms or colleges are primarily differentiated by quality. Further, the insights and results from our analysis have proven useful to empirical approaches in related models (see Agarwal, 2015; Vissing, 2016;

[1] See Fox (2009) and Chiappori and Salanié (2015) for surveys.

[2] Choo and Siow (2006) assume that the (pre-transfer) utility of agent i for partner j is given by $u_{ij} = \phi(x_i, z_j) + \varepsilon_i(z_j)$, where x_i and z_j are observed. Therefore, characteristics of agent j that are not observed in the data do not directly affect the utility of agent i.

Jiang, 2016; Agarwal, 2017).

Estimates of the distribution of the latent index as a function of observables are useful for the analysis of many economic questions. For instance, quantitatively evaluating the trade-off for firms between workers' experience or education and unobserved productivity, or the trade-off for workers between wages and the value of amenities such as on-the-job training may require estimates that account for unobserved characteristics. Similarly, evaluating the consequences of a market reform (such as policies that place limits on college tuition) can require estimating the distribution of latent indices on both sides of the market. Identifying and estimating 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 a more preferred institution.

We study these problems assuming that the available data are from a large market. This approach is motivated by the fact that 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 at least several thousand participating agents. For similar reasons, recent papers in the theoretical matching literature have utilized large market approximations for analyzing strategic behavior and the structure of equilibria.[3] In our analysis, large market approximations highlight and account for important interdependence between matches within a market in the asymptotic analysis of our estimator.

Even with the stark restriction that preferences are homogeneous, our first result on identification is negative. We show that the distribution of the latent index is not identified from data from a single large market with one-to-one matches. Indeed, we construct an example parametric family of models of one-to-one matching that are observationally equivalent. This example illustrates that our non-identification result is not pathological. Intuitively, the observed joint distribution of agents and their match partners, which we refer to as sorting patterns, does

[3]See Immorlica and Mahdian (2005); Kojima and Pathak (2009); Azevedo and Budish (2013), for example.

not allow us to condition on unobservables. The non-identification arises because unobservable characteristics of either side of the market could be driving these sorting patterns. These results imply limitations on what can be learned using data on one-to-one matches and guide the use of empirical techniques. For instance, they weigh against estimating the distribution of the latent index in marriage markets using data from a single market. Nonetheless, data from one-to-one matches may still be useful for certain questions. We show that the relative value of various observed characteristics are identified with one-to-one matches. However, this limits the scope of questions that may be answered with such data.

In contrast to the non-identification result with one-to-one matches, we show that the distribution of latent indices on both sides of the market is non-parametrically identified from data on many-to-one matches. The key insight is that the same value of the unobservable characteristic of an agent determines multiple matches for that agent. The formal 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 difference between the empirical content of one-to-one matching and many-to-one matching has not been previously exploited to obtain non-parametric identification results of a model with unobserved characteristics. Our proof is based on interpreting the matching model with two-to-one matches in terms of a measurement error model (Hu and Schennach, 2008). This reinterpretation makes the additional empirical content of many-to-one matches ex-post intuitive: the observable components of a worker's quality provide a noisy measure of the overall quality of her colleagues. As in measurement error models, we use the repeated measurements made available when many workers match with the same firm to identify the model.

We also use simulations from a parametrized family of models to illustrate the additional identifying information available in many-to-one matches. 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 a dataset from a single large market. In contrast, our 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. These simulations suggest that using such information is important in empirical applications. For example, they suggest that moments such as the within-firm variance in worker observables contain information about primitives beyond what can be learned from the covariance between worker and firm observables. We therefore recommend empirical strategies that use information from many-to-one matching, when available.[4]

We then study the asymptotic properties of a minimum distance estimator for a parametric model based on a criterion function that uses moments from many-to-one matching as well as sorting patterns. As in the identification analysis, we develop an asymptotic theory based on data from a single market with the number of agents growing large. This approach requires us to deal with technical challenges that arise from the dependence of each match on the characteristics of all agents in the market. We prove both consistency and $\sqrt{N}-$asymptotic normality of the estimator. For simplicity, we restrict attention to the case with two-to-one matching. To our knowledge, ours is the first result on asymptotic theory for an estimator in a single large matching market.

Our asymptotic theory requires us to confront the fact that the observed matches, as well as the model predictions, are a non-linear function of the observables and unobservables in the entire market. We separately analyze the sampling distribution of the moments in the data and the map from the structural parameters to these moments. To prove a limit theorem for the sampling distribution of the moments in the data, we use the fact that the distribution of the observed characteristics of matched pairs depends only on the latent index. Hence, the conditional distribution of the observables given the latent indices are independent on the two sides of the market. This insight allows us to derive the asymptotic distribution of the moments

[4]If data from many-to-one matches is not available, it may be possible to use variation in market composition to identify the distribution of latent indices. We are not aware of any formal results that show that such variation is sufficient for identification. This approach may require assuming that the parameters governing the primitives are constant across the markets.

of the data.

Then, we study the model's prediction for the moments as a function of the structural parameters and the observables in the data. Analyzing this map is challenging because the matches depend on the characteristics of all agents in the market. This generates dependency that cannot be analyzed using standard empirical process techniques for i.i.d. data (e.g. van der Vaart and Wellner, 2000). In particular, deriving the sensitivity of the matches between extremely desirable or extremely undesirable agents to the parameter requires controlling the tail behavior of the latent index. We make progress by first showing that this map, ignoring the tails of the latent utilities, is smooth – specifically, Hadamard differentiable – in the sampled observed characteristics. This allows us to use continuous mapping theorems and the functional delta method to show convergence properties, except at the tails. When the tails are negligible, the limit as the size of the tails we ignore goes to zero yields large sample properties of the moment function.

The dependence inherent in the model also complicates the analysis of these tails. We show that the tails are negligible by adapting a chaining argument from the empirical process literature (Pollard, 2002), replacing a tail bound for i.i.d. data used in the existing proof with a concentration of measure inequality (Boucheron et al., 2003) suitable to the dependent data in our problem. This method allows us to prove the equicontinuity results necessary for the limit theorem. For simplicity of exposition in the main text, the technical regularity conditions on the primitives that justify this approximation are detailed in the appendix. Finally, we use Monte Carlo simulations to study the property of a simulated minimum distance estimator.

The paper starts with a brief discussion of the related literature, after which we present the model (Section 2). Section 3 discusses identification with one-to-one and many-to-one matching, Section 4 presents our asymptotic analysis of the estimator, and Section 5 presents Monte Carlo results. All proofs are in the Appendix.

Related Literature: Most of the recent literature on identification and estimation of matching games studies the transferable utility (TU) model, in which the equilibrium governs the 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, 2012; Chiappori et al., 2015, among others). The equilibrium transfers are such 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 that determines the equilibrium matches. 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 each agent's utility depends only on observed characteristics and an unobserved taste shock drawn from a specified distribution. Using this assumption, the papers propose estimation and identification of group-specific surplus functions (Choo and Siow, 2006; Galichon and Salanie, 2012; Chiappori et al., 2015). 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. Compared to these approaches, our study is restricted to a single index model but incorporates both TU and NTU matching in a non-parametric framework. We also allow for unobserved characteristics of the partner to affect agent preferences and are interested in identifying the distribution of unobservable characteristics, which are not considered in the maximum score approach by Fox (2010a).

In many applications, inflexible monetary transfers or counterfactual analyses that require estimates of preferences for agents on both sides of the market motivate the use of a non-transferable utility model (c.f. Roth and Sotomayor, 1992). Previous analysis of NTU models have resulted in only partial identification. 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 (2015) 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 in which observable characteristics have bounded support. Menzel (2015) finds that only the sum of the surplus of both sides obtained from matching is identified from data on one-to-one matching. The result that identification is incomplete with one-to-one matching is similar in spirit to our negative result on identification. 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 our results come at the cost of assuming homogeneous preferences.

With the exception of Chiappori et al. (2012) and Galichon et al. (2014), previous models are typically restricted to either non-transferable or transferable utility. The objective in Galichon et al. (2014) is to generalize the Choo and Siow (2006) framework to models of imperfectly transferable utility. Our framework is closer to that of Chiappori et al. (2012), which studies a marriage market with positive assortative matching. They also assume a single index model and allow for both transferable and non-transferable utility matching. They show that the marginal rates of substitution between two observable characteristics is identified using data on one-to-one matching. Our identification results with data on one-to-one matching are consistent with their results, but may also explain why Chiappori et al. (2012) may not have estimated the distribution of the latent index with their data. Specifically, we show that a many-to-one matching market is needed for such identification. Agarwal (2015) and Vissing (2016) use our insight on the information in many-to-one matching to respectively estimate preferences in the market for medical residents and the market for oil drilling contracts using simulated minimum distance estimators. This approach is different from work by Logan et al. (2008) and Boyd et al. (2013), who propose techniques that use only the sorting of observed characteristics of agents as given by the matches (sorting patterns) to recover primitives. Our result on non-identification of a single-index model with data only on sorting patterns implies that a more general model with heterogenous preferences will also not be identified. Therefore, our results suggest that point estimates obtained using only information in sorting patterns 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, 2013). 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 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,b) 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 the extent to which unobservable characteristics of each side of the market drive the observed patterns.

The 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.

Finally, we use a novel approach for dependent data to prove our limit theorems because standard empirical process theories for i.i.d. data are not applicable in our context. This feature of our model may be shared with other contexts, such as network formation models (c.f. Graham, 2014; Boucher and Mourifie, 2012; Chandrasekhar and Jackson, 2015; Leung, 2015). A common technique in the asymptotic analysis of network models is based on assuming that dependence across links decays with a notion of distance between two nodes. Our application of concentration of measure inequalities removes the need for an analogous assumption in our model.

2 Model

We consider a two-sided matching market with one side labelled as workers and the other labeled firms. Although these labels are suggestive of a labor market, the model may be applied to other two-sided matching markets, including matching of students to schools, and the marriage market. Our model does not presume a monetary transfer between the two sides of the market, and will include both non-transferable and transferable utility cases. We first describes the latent indices that will be the object of interest in our identification and estimation analysis before discussing their interpretation in transferable and non-transferable utility models. Finally, we discuss questions of interest that may be answered in this framework.

2.1 Latent Indices

Most datasets have information on a limited number of characteristics on each side of the market. Let the observable characteristics of worker i be x_i and the observable characteristics of firm j be z_j. Given our focus on positive assortative matching, we posit two latent quality indices, v_i and u_j, one for each side of the market. These indices simply order workers and firms by quality and do not impose cardinal restrictions. For instance, firms may have heterogeneous production functions that take human capital (v_i) as an input. The latent indices can depend on both observable characteristics as well as unobserved characteristics. Specifically, we assume that worker i's human capital index is given by the additively separable form

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

where we set the location normalizations $h\left(\bar{x}\right) = 0$ for some $\bar{x}$, assume that ε is median zero, and set the scale normalization $\left|\nabla h\left(\bar{x}\right)\right| = 1$. Because an additively separable representation of preferences is unique up to a positive affine transformation, the scale and location normalizations are without loss of generality. These normalizations ensure that the latent indices in our model are well defined.

The scalar unobservable ε_i aggregates the effect of all relevant determinants worker quality that are not observed in the dataset. Additive separability in ε_i implies that the marginal value of observable traits does not depend on the unobservable.

As for the model for the human capital index, we assume that quality of firm j is given by

$$u_j = g(z_j) + \eta_j, \tag{2}$$

with the normalizations $g(\bar{z}) = 0$, η is median zero and $|\nabla g(\bar{z})| = 1$. The quality of the firm may reflect productivity differences or on-the-job amenities for workers. For instance, one may also include wages in this model through one of the characteristics z_j if they are not negotiated during the matching process. This approach may be used to model medical residency markets or colleges/schools in countries with non-negotiable tuition rates.

We make the following assumptions on the model:

Assumption 1 *(i) ε and η are independent of X and Z respectively*

(ii) ε and η have bounded, differentiable densities, f_ε and f_η, with full support on $\mathbb{R}$, and non-vanishing characteristic functions

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

(iv) The random variables $h(X)$ and $g(Z)$ admit bounded continuous densities f_h and f_g

Assumption 1 (i) assumes independence of the unobservables. On its own, independence is not particularly strong, but the restriction of additive separability makes this restrictive. Additive separability with independence is commonly used in discrete choice literature as it significantly eases the analysis. Assumption 1 (ii) requires that ε and η have large support and imposes technical regularity conditions on their distributions that will be useful in our identification analysis. The support conditions in Assumption 1 (iii) ensures that the observables can trace out the distribution of ε and η in the tails as well, and Assumption 1 (iv) requires at least one covariate to be sufficiently smooth while others may be discrete.

2.2 Positive Assortative Matching

The composition of the market is described by a pair of probability measures, $\mu_{X,\varepsilon}$ and $\mu_{Z,\eta}$. Here, $\mu_{X,\varepsilon}$ is the joint distribution of workers' observable traits $x \in \chi \subseteq \mathbb{R}^{k_x}$ and unobservable traits $\varepsilon \in \mathbb{R}$. Likewise, $\mu_{Z,\eta}$ is the joint distribution of firms' observable traits $z \in \zeta \subseteq \mathbb{R}^{k_z}$ and unobservable traits $\eta \in \mathbb{R}$. This formulation allows us to consider large but finite economies, as well as a continuum limit in a unified notational framework. For instance, an economy with N agents on each side can be represented with the measures $\mu_{(X,\varepsilon)_N} = \frac{1}{N}\sum_{i=1}^{n} \delta_{(X_i,\varepsilon_i)}$ and $\mu_{(Z,\eta)_N} = \frac{1}{N}\sum_{j=1}^{n} \delta_{\left(Z_j,\eta_j\right)}$, where δ_Y is the dirac delta measure at Y.

A one-to-one **match** is a probability measure μ on $(\chi \times \mathbb{R}) \times (\zeta \times \mathbb{R})$ with marginals $\mu_{X,\varepsilon}$ and $\mu_{Z,\eta}$ respectively. The measure μ could be used to represent a continuum limit as well as a finite-economy match. The traditional definition of a finite-market match used in Roth and Sotomayor (1992) is based on a matching function $\mu^*(i) \mapsto \mathcal{J} \cup \{i\}$, where $\mathcal{J}$ is the set of firms. For an economy of size N, with probability 1, such a function defines a unique counting measure 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. When η and ε admit a density, in a finite economy, (z,η) (respectively (x,ε)) identifies a unique firm (respectively worker) with probability 1.[5] A many-to-one match with M partners on one side is defined analogously as a measure μ on $(\chi \times \mathbb{R})^M \times (\zeta \times \mathbb{R})$.

The match μ is **positive assortative** 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 $\mu_{X,\varepsilon}$ and $\mu_{Z,\eta}$ respectively, such that $\int_{S_I} (h(X) + \varepsilon)\, d\mu_{X,\varepsilon} > \int_{S_I} (h(X) + \varepsilon)\, d\mu(\cdot, S_J)$ and $\int_{S_J} (g(Z) + \eta)\, d\mu_{Z,\eta} > \int_{S_J} (g(Z) + \eta)\, d\mu(S_I, \cdot)$. This definition considers two potential sets of agents S_I and S_J. If $\int_{S_I} (h(X) + \varepsilon)\, d\mu_{X,\varepsilon} > \int_{S_I} (h(X) + \varepsilon)\, d\mu(\cdot, S_J)$, then the expected value of the latent indices of agents in S_I are larger than those matched with S_J. The analogous inequality for agents in S_J yields the second condition. Hence, there are no such sets if these inequalities are not simultaneously satisfied for any pair S_I and S_J, and the matching is assortative.[6]

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

[6] We do not consider unmatched agents for two reasons. First, different equilibrium notions matching (TU or NTU) impose different restrictions on preferences for unmatched agents. Using the implications of positive

This formulation presents a unified definition for assortativity in continuum markets as well as markets with a finite number of agents. In the finite market case, consider a match in which an agent with characteristics (x, ε) (respectively (x', ε')) is matched with an agent with characteristics (z, η) (respectively (z', η')). Now, consider singleton sets $S_I = \{(x, \varepsilon)\}$ and $S_J = \{(z', \eta')\}$. The inequalities above imply that either $h(x) + \varepsilon \leq h(x') + \varepsilon'$ or that $g(z') + \eta' \leq g(z) + \eta$. Therefore, there are no such pairs of sets in the finite markets if the conditions of our definition are satisfied. In what follows, we simply assume that the market is characterized by positive assortative matching. As we discuss in the next few sections, this assumption encompasses both transferable and non-transferable utility models.

Further, our model requires that the matching only depends on the latent index. This assumption is vacuous in finite samples because ties are zero-probability events. Shi and Shum (2014) formalize this as "random matching" in a continuum version of the Beckerian marriage model. They note that without this assumption, the distribution of observed characteristics of matched partners is indeterminate. Our consistency results imply that the moments of the finite sample data naturally converge to a population analog with this property. Therefore, the data generating process we consider has the following property in a positive assortative match:

Remark 1 *A positive assortative 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. Further, the latent index is a sufficient statistic for the distribution of match partners.*

Hence, 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. The dependence only on the latent index, in the one-to-one case, implies that $\mu_{X,\varepsilon|Z,\eta} = \mu_{X,\varepsilon|Z',\eta'}$ if $g(Z) + \eta = g(Z') + \eta'$ and $\mu_{Z,\eta|X,\varepsilon} = \mu_{Z,\eta|X',\varepsilon'}$ if $h(X) + \varepsilon = h(X') + \varepsilon'$. Our paper studies identification and estimation of the latent utility indices using data from a matching market with this property. As described

assortative matching alone allows us to be agnostic about the nature of transfers. Second, many datasets do not have information on unmatched agents. For example, typical employer-employee matched datasets do not contain the number of job openings, and Agarwal (2015) does not have information on medical residents that were not placed at residency programs.

below, it turns out that positive assortative matching on v and u can result from both non-transferable and transferable utility models.

2.2.1 Non-transferable Utility (NTU) Matching

Models of matching markets in which transfers between the parties are prohibited or restricted are commonly used in the theoretical literature (c.f. Roth and Sotomayor, 1992). Motivating examples include marriage markets, public schooling, and colleges. In such a model, the latent indices v_i and u_j are interpreted as representing the *ordinal* preference relation for their match partners. Because these indices are ordinal, the framework allows for each firm j to have a separate production function $\Phi_j(v)$ as long as Φ_j is strictly increasing in v. In the many-to-one matching case, a focus of this paper, we will assume that Φ_j is increasing in each of its components. Specifically, $\Phi_j(v_1, v_2)$ is increasing in both v_1 and v_2 when a firm is matched with two workers.

The typical equilibrium assumption is that of pairwise stability, which makes two restrictions. First, there is no worker-firm pair such that both agents prefer matching with each other to their current match (where the firm can fire a currently matched worker, if necessary). Second, no worker or firm is matched with an unacceptable partner. 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 discussed in Clark (2006) and Niederle and Yariv (2009). It is easy to see that the unique pairwise stable match is positive assortative on the latent indices v_i and u_j. Given our focus on positive assortative matching, we assume that all workers and firms are acceptable to the other side.

Although the models are referred to as non-transferable utility models, the model can incorporate transfers that are not simultaneously determined with the matching. In this case, one of the observables includes the salary offered by program j. Estimating the latent index allows one to measure the willingness to pay for various on-the-job amenities by assuming a functional form, say

$$u_j = z_j\beta + w_j + \eta_j. \tag{3}$$

For instance, Agarwal (2015) uses a similar model to quantify the value for various attributes of medical residency programs such as size, prestige, and patient mix.

An important restriction in the latent index framework is that agents have homogeneous ordinal preferences over their match partners. While the theoretical literature assumes very general preferences when studying the existence of stable matchings, formal identification and estimation analysis is yet to incorporate significant heterogeneity in preferences.

2.2.2 Transferable Utility (TU) Matching

Our latent index framework fits well into the classical Beckerian model of the marriage market. This matching model posits men and women differentiated by a one-dimensional characteristics that split a surplus from marriage given by $\Phi(u_j, v_i)$. A matching is pairwise stable if there are transfers t_{ij} (possibly negative), such that no man-woman pair find it mutually beneficial to agree to a transfer and match with each other. As is well known, the unique pairwise stable match is positive assortative on u and v if $\Phi(u_j, v_i)$ is supermodular. This elegant model has received a considerable amount of attention, and patterns of positive assortative matching observed along age, income, and education in various marriage markets have been well documented.

A thrust of our paper is the consideration of many-to-one matching. In this case, we assume separability of the surplus function across matches in order to maintain positive assortative matching on the latent indices. Specifically, we assume that the surplus generated by firm with index u_j that is matched with workers v_i and v_k is given by

$$\Phi(v_i, v_k, u_j) = \tilde{\Phi}(v_i, u_j) + \tilde{\Phi}(v_k, u_j), \tag{4}$$

where $\tilde{\Phi}.$ is supermodular. The assumption rules out complementarities across matches but retains positive assortativity in a pairwise stable match. It also assumes that the multiple matches for an agent are symmetric. For example, in the worker-firm context, the model is best suited for a market in which firms are hiring multiple workers with the same job description.

2.3 Unobserved Characteristics

The lack of perfect positive assortative matching on observable characteristics may be attributable to unaccounted preference heterogeneity or unobserved characteristics. These unobserved characteristics are important for rationalizing the data. Previous approaches have typically focussed on the identification of observable components of utility, often under parametric assumptions on the distribution of unobserved characteristics.[7] For instance, Chiappori et al. (2012) study a single index model like ours and obtain identification of the marginal rates of substitution

$$\frac{\partial h(x)/\partial x_1}{\partial h(x)/\partial x_2} \text{ and } \frac{\partial g(z)/\partial z_1}{\partial g(z)/\partial z_2}.$$

These quantities can be used to determine the trade-offs between observables, such as the trade-off between a worker's education and experience. Some economic questions, however, may require an analysis of unobservables. For example, one may be interested in knowing how much of a worker's human capital is explained by experience and/or education. This exercise may require decomposing the variance of human capital into observable and unobservable components. Similarly, questions about compensating differentials in labor markets require valuing on-the-job amenities or training, some components of which may not be observed.

While several objects of interest can be measured through these marginal rates of substitution between observed characteristics, many economic questions require a deeper understanding of how agents' preferences respond to interventions in matching markets. For example, one may be interested in the effect of a subsidy on college tuition on matches that occur in equilibrium. To predict the counterfactual matches, one needs to measure the effect of this subsidy on the relative desirability of various colleges to students. Changes in the relative desirability of colleges depend on the monetary value students place on unobserved college characteristics. Therefore, an important objective in this paper is to understand when the distributions of ε and η are

[7] For instance, Galichon and Salanie (2012) generalize the model by Choo and Siow (2006) and show that $\bar{\Phi}_{x_i z_j}$ is identified for a separable surplus function of the form $\Phi_{ij} = \bar{\Phi}_{x_i z_j} + \varepsilon_i(z_j) + \eta_j(x_i)$ with known distributions of $\varepsilon_i(z_j)$ and $\eta_j(x_i)$. These models therefore allow for unobserved preferences for observed characteristics, but do not allow for unobserved characteristics. Menzel (2015) studies an NTU model with a light restriction on tail behavior of the unobservables to identify the sum of the match surpluses accruing to each side due to observables.

identified, which in turn implies identification of the probabilities

$$\mathbb{P}\left(h\left(x_1\right)+\varepsilon_1>h\left(x_2\right)+\varepsilon_2|x_1,x_2\right) \text{ and } \mathbb{P}\left(g\left(z_1\right)+\eta_1>g\left(z_2\right)+\eta_2|z_1,z_2\right). \tag{5}$$

These choice probabilities are also fundamental in the analysis of counterfactual changes in market structure, market composition, and other policy-relevant counterfactuals.

It is important to note that the latent indices we analyze, u and v, are ordinal measures of the desirability of agents in the market. Identification of the total surplus function in the transferable utility case, $\Phi\left(u,v\right)$, or a cardinal measure of utilities in the non-transferable utility case will require additional assumptions. For example, one may simply interpret the latent index as a utility measure in the NTU case or assume a particular structure for the surplus function in the TU case if this is desirable for the empirical application being considered. We avoid these assumptions for simplicity and to retain generality with respect to these choices. In applications where one of the observed traits presents a natural measure of value, our indices can be interpreted in units of this metric for value.

3 Identification

This section starts by showing that data from a single matching market are sufficient for identifying certain features of preferences. Specifically, one can identify the indices $h\left(x\right)$ and $g\left(z\right)$. We then show that data from one-to-one matches is unable to identify the distribution of the latent indices if there are unobserved characteristics on both sides of the market. Next, we show that data from many-to-one matching restores full identification of the distribution of preferences. Finally, we illustrate these results using simulations.

3.1 Sorting Patterns, Indifference Curves and a Sign-restriction

We now study what can be learned from the joint distribution μ_{XZ} of observed firm and worker traits. This is the marginal of μ on the observables, and it summarizes all information available in data from one-to-one matching. It allows the assessment of the sorting of worker observable traits to firm observables. We therefore refer to features of this distribution as "sorting patterns." As our first result shows, these features of the data allow us to identify the indices $h(x)$ and $g(z)$ up to monotonic transformations.

Lemma 1 *Under Assumption 1, the level sets of the functions $h(\cdot)$ and $g(\cdot)$ are identified from data on a one-to-one match, i.e. μ_{XZ} is observed.*

Proof. See Appendix A.1. ■

The result states that we can determine whether or not two worker types x and x' are equally desirable from the sorting patterns observed in a one-to-one matching market (hence, also if many-to-one matches are observed). Intuitively, 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 positive assortative match, under the additive structure of equations (1) and (2) and the 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 distributions of firm observables, they cannot be identical in observable quality. This result is similar to those in Chiappori et al. (2012) that show that differentiability of $h(\cdot)$ and $g(\cdot)$ implies identification of marginal rates of substitution, which are pinned down by indifference curves.

While the level sets of $h(\cdot)$ and $g(\cdot)$ are known, we cannot yet determine $h(\cdot)$ and $g(\cdot)$ even up to positive monotone transformations. In particular, we cannot tell whether a 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 a **sign restriction** is sufficient for identifying $h(\cdot)$ and $g(\cdot)$ up to positive monotone transformations.

Assumption 2 *(i) The functions $h(x)$ and $g(z)$ are strictly increasing in their first arguments*

(ii) Further, for each $x_{-1} = (x_2, \ldots, x_{k_x})$ and $z_{-1} = (z_2, \ldots, z_{k_z})$, $h(x_1, x_{-1})$ and $g(z_1, z_{-1})$ have full support on $\mathbb{R}$

Part (i) imposes a sign restriction that requires that the latent index is strictly increasing in at least one observable characteristic. It is often natural to impose this restriction in matching markets. For example, it is reasonable to argue that the desirability of workers is increasing in education, holding all else fixed. Given such an assumption, our next result shows that $h(\cdot)$ and $g(\cdot)$ can be determined up to positive monotone transformations. Part (ii) makes a large support assumption that allows ordering all the level sets of $h(x)$.

Proposition 1 *If Assumptions 1 - 2 are satisfied then, $h(\cdot)$ and $g(\cdot)$ are identified up to positive monotone transformations.*

Proof. Identification of h and g up to a positive monotone transformation follows immediately from Lemma 1 and Assumption 2. ■

The sign restriction allows us to order the level sets of h and g.

3.2 Limitations of Sorting Patterns

As mentioned earlier, typical datasets do not contain all relevant characteristics of agents on both sides of the market. The dispersion around a central positive assortative trend is a manifestation of these unobservables. Remark 1 reflects the importance of unobservables as workers with characteristic (x, ε) are matched with firms with characteristics (z, η) if

$$h(x) = F_V^{-1} \circ F_U (g(z) + \eta) - \varepsilon. \tag{6}$$

This expression indicates that there are two sources of unobservables that result in imperfect assortativity, namely η and ε. Without these unobservables, a researcher would observe perfect positive assortativity along the estimated indices $h(x)$ and $g(z)$.

A question remains about whether we can learn about the distribution of both these unobservables with data on one-to-one matches, which only contains information in F_{XZ}. The following stylized example shows that the answer is negative. A wide range of parameters could be consistent with the data, even a highly parametric case.

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 $N(0,\sigma_\eta^2)$ respectively. The distributions of U and V are therefore $N(0,1+\sigma_\eta^2)$ 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 $U|Z \sim N(Z,\sigma_\eta^2)$, and that $F_V^{-1} \circ F_U = \left[\frac{1+\sigma_\varepsilon^2}{1+\sigma_\eta^2}\right]^{1/2}$. Therefore, 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 $\eta \sim N(0,\sigma_\eta^2)$, independently of X and Z. Hence, $X|Z = z$ is distributed as

$$N\left(\frac{z}{\kappa^{1/2}}, 1 - \frac{1}{\kappa}\right), \tag{7}$$

where $\kappa = (1+\sigma_\varepsilon^2)(1+\sigma_\eta^2)$.

The distribution in equation (7) is identical for all pairs $(\sigma_\varepsilon, \sigma_\eta)$ with $(1+\sigma_\eta^2)(1+\sigma_\varepsilon^2) = \kappa$. Thus, the family of matching models with $(1+\sigma_\eta^2)(1+\sigma_\varepsilon^2) = \kappa$ are observationally equivalent when only data from one-to-one matches is available.

The example above shows that data on one-to-one matches cannot be used to identify the distribution of the two latent indices in the presence of unobservables on both sides of the market. This highlights a central limitation of data from a market with one-to-one matching such as the marriage market.[8] Section 3.4 illustrates this limitation using a simulated objective function.

The failure of identification can be understood by considering the case in which $\varepsilon \equiv 0$. Equation (6) reduces to

$$h(x) = F_V^{-1} \circ F_U (g(z) + \eta).$$

This expression shows that when $\varepsilon \equiv 0$, the model is mathematically identical to the well-studied transformation model (Ekeland et al., 2004; Chiappori and Komunjer, 2008). Appendix

[8] This observation suggests one reason why Chiappori et al. (2012) do not estimate the distribution of the latent index in their paper on the marriage market.

C.1 uses results from Chiappori and Komunjer (2008) to formally derive conditions under which any distribution F_{XZ} can be rationalized.

These results imply that a model with unobservables on both sides is under-identified. This non-identification is despite imposing additional regularity conditions. Hence, empirical strategies to estimate the distribution of latent preferences using information in sorting patterns may be suspect. Logan et al. (2008) and Boyd et al. (2013) employ empirical strategies that only use sorting patterns to estimate preferences for models that include preference heterogeneity. Our non-identification result suggests that point estimates from this approach, including for models more general than the single index model, may be sensitive to parametric assumptions. Such non-identification is 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 preferred to any worker with trait x' if $h(x) > h(x')$, even if this is not the case.

3.3 Identification from Many-to-One Matches

We now show that data from many-to-one matching markets can be used to identify the model. Consider a dataset in which there are a large number of firms, and each firm has two workers hired at the same position. Therefore, we may arbitrarily label the slots occupied by each worker as slots 1 and 2, independently of the firm and worker characteristics. The data are summarized by 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.

To see why multiple matches per partner can be useful for identification, note that the observed worker/firm characteristics present noisy measures of the true quality of the partners matched with each other. Remark 1 implies that the following two equalities when workers with characteristics (x_1, ε_1) and (x_2, ε_2) are matched with a firm with characteristics (z, η):

$$\begin{aligned} h(x_1) &= F_V^{-1} \circ F_U (g(z) + \eta) - \varepsilon_1 \\ h(x_2) &= F_V^{-1} \circ F_U (g(z) + \eta) - \varepsilon_2. \end{aligned}$$

Agarwal (2015) uses this insight and discusses it 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 programs 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 medical schools of varying quality. Our result below formally shows the usefulness of data from many-to-one matching. We therefore recommend the use of this information when available.

Theorem 3 *Under Assumptions 1 - 2, the functions $h(\cdot)$, $g(\cdot)$ and the densities f_η and f_ε are identified when data from two-to-one matching is observed, i.e. $F_{X_1,X_2,Z}$ is observed.*

Proof. See Appendix A.2. ■

The proof proceeds by interpreting our model in terms of a nonlinear measurement error model and employing techniques in Hu and Schennach (2008) to prove identification. To understand the analogy, note that the distribution observables of matched partners depends only on the latent index. Positive assortative matching implies that all partners have the same quantile of the latent index. Therefore, to write the joint distributions of the observables given a quantile q, we need to consider the conditional densities of the observables X_1, X_2 and Z given q. For expositional simplicity, assume that these densities exist. Therefore, the joint distribution $f_{X_1,X_2,Z,q}$ can be factored as follows:

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

where $f_q(q) = 1$ for $q \in [0,1]$ and 0 otherwise because quantiles are uniformly distributed, $f_{X_1|q}(x_1|q)$ is the conditional density at x_1 given that $h(x_1) + \varepsilon = F_V^{-1}(q)$, and $f_{X_2|q}(x_2|q)$ and $f_{Z|q}(z|q)$ are defined analogously. Integrating this quantity with respect to q yields the observable quantity

$$f_{X_1,X_2,Z}(x_1, x_2, z) = \int_0^1 f_{X_1|q}(x_1|q)\ f_{X_2|q}(x_2|q)\ f_{Z|q}(z|q)\, dq.$$

Intuitively, this simplification arises from the latent index assumption and positive assortative matching on v and u. Mathematically, this equation is identical to the nonlinear measurement error model of Hu and Schennach (2008), with the latent variable q governing the distribution of the observables.[9] This formulation clarifies the intuition that the observable characteristics of matched partners are noisy signals of the underlying latent index, and it allows us to identify the distributions of X and Z conditional on the quantile q. We then identify the model using the scale and location normalizations on h, g, f_ε, f_η, and Assumption 1.

While these results are derived in the specific context of a latent index model with no preference heterogeneity, they highlight the fact that data from many-to-one matches has additional empirical information that cannot be obtained from one-to-one matches. This insight has enabled and guided the empirical analyses of more flexible models of the medical match (Agarwal, 2015) and the market for oil drilling rights (Vissing, 2016). An extension of our analogy of a matching model to a measurement error model has also been used to prove identification results for and study a labor market model with data on worker productivity (Jiang, 2016).

3.4 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 they show 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. To mimic realistic empirical applications, our simulations have firms with varying capacity instead of the fixed number of workers per firm.

We simulate a dataset using known parameters and then compare objective functions of various minimum distance estimators. Specifically, we compare an objective function that exclusively uses moments based on sorting patterns to another that also uses information from

[9] A technical difference with Hu and Schennach (2008) is that we replace Assumptions 1 and 5 in their paper with implications of Assumptions 1. See appendix for details.

many-to-one matching. We model the latent indices as

$$\begin{aligned} v_i &= x_i\alpha + \varepsilon_i \\ u_j &= z_j\beta + \eta_j, \end{aligned}$$

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 $\mu : \{1, \ldots, N\} \to \{1, \ldots, J\}$ is computed for $\alpha = 1$ and $\beta = 1$. Using the same dataset 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

$$\hat{\psi}_{ov} = \frac{1}{N}\sum_i x_i z_{\mu(i)} \tag{8}$$

$$\hat{\psi}^S_{ov}(\theta) = \frac{1}{S}\sum_s \frac{1}{N}\sum_i x_i z_{\mu_s^\theta(i)} \tag{9}$$

and

$$\hat{\psi}_w = \frac{1}{N}\sum_i \left(x_i - \frac{1}{|\mu^{-1}(\mu(i))|} \sum_{i' \in \mu^{-1}(\mu(i))} x_{i'} \right)^2 \tag{10}$$

$$\hat{\psi}^S_w(\theta) = \frac{1}{S}\sum_s \frac{1}{N}\sum_i \left(x_i - \frac{1}{\left|\left(\mu_s^\theta\right)^{-1}\left(\mu_s^\theta(i)\right)\right|} \sum_{i' \in \left(\mu_s^\theta\right)^{-1}\left(\mu_s^\theta(i)\right)} x_{i'} \right)^2. \tag{11}$$

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 $\hat{\psi}^S_w(\theta)$ captures 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 $\hat{Q}^S(\theta) = \left\| \hat{\psi} - \hat{\psi}^S(\theta) \right\|_W$, 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 uses information on 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 $\hat{\psi}^S_w(\theta)$. The vertical contours indicate that the moment is able to clearly distinguish values of α because the moment $\hat{\psi}^S_w(\theta)$ is strictly decreasing in α. However, the shape of the objective function indicates that this moment cannot distinguish different values of β.

Finally, the plot of an objective function that penalizes deviations from both $\hat{\psi}_w$ and $\hat{\psi}_{ov}$ (Figure 1(c)) shows 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

This section develops an estimator for the latent index model considered above. We then study the limit properties of this estimator and derive conditions under which the estimator is consistent and asymptotically normal. As in the identification analysis, we consider a dataset from a single large matching market. This choice is motivated by the fact that researchers typically have data on a single (or few) matching markets with many participants.[10] This includes applications in labor markets, marriage markets, and education markets. The analysis of asymptotic properties in a single large market is technically challenging because the characteristics of any individual's match partner depends on the composition of the entire market. To our knowledge, consistency or asymptotic theory has not been previously established for parametric models, even with a single latent index.[11]

There are several technical insights that allow us to solve this problem. First, we use the property that we observe a positive assortative match along a single latent index. This allows us to re-write the dependence of the matches in terms of the latent indices. While restrictive on the nature of primitives, our model allows for a large parametric class of models and both transferable and non-transferable utility. Second, the problem can be decomposed into separately analyzing two distinct pieces. The first problem is to show limit theorems for the observed moments of the data as the market size increases. Separately, we must show a uniform limit theorem for the map from structural parameters to these moments. Third, we find that analyzing this map by first ignoring the behavior in the tails of the latent indices and then showing that the tails are negligible is the most tractable approach. Finally, to ensure that tails are negligible, we adapt a chaining argument from the empirical process literature, using a concentration of measure inequality to replace tail bounds for i.i.d. data that do not apply in our setting.

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

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

[10]In cases where many matching markets are observed, it may not always be appropriate to assume that the underlying preference parameters are the same across all markets.

[11]Even proving consistency is non-trivial. For example, Dupuy and Galichon (2014) show that the canonical correlation estimator suggested by Becker (1973) is inconsistent. A previously circulated version of this paper (Agarwal and Diamond, 2014) shows consistency of the estimator studied here under weaker conditions on the primitives.

where $g : \zeta \times \Theta \to \mathbb{R}$ and $h : \chi \times \Theta \to \mathbb{R}$ are known functions that are Lipschitz continuous in θ for each x and z with constants $g_{LC}(z)$ and $h_{LC}(x)$ respectively. We assume that the densities f_ε and f_η are known, and ε and η are independent of x and z respectively.

We adopt a parametric approach for several reasons. First, our identification argument does not directly suggest a non-parametric estimator. Second, our focus is on solving issues that arise from the dependent data nature of the problem. Relaxing the parametric assumption would further complicate the analysis. Finally, computational burden in empirical applications have often prevented extremely flexible functional forms from being implemented. Similar parametric assumptions are common in the discrete choice literature where one typically assumes a normal or an extreme value type I distribution for the unobservable ε.

We assume that the data contains a sample of J firms, each with $\bar{c}$ slots, 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 i.i.d. from the measure $\mu_{X,\varepsilon}$ and the characteristics of the firm are sampled i.i.d. from $\mu_{Z,\eta}$. For simplicity of analysis and notation, we set $\bar{c} = 2$.

4.1 A Minimum Distance Estimator

We propose an estimator based on a minimum distance criterion function. Specifically, let $\Psi(x_1, x_2, z) \in \mathbb{R}^{K_\Psi}$ be a bounded vector-valued moment function, i.e. $\|\Psi\|_\infty < \infty$, where x_1 and x_2 are the observed characteristics of two workers and z is the observed characteristics of the firm. We assume that Ψ is symmetric in x_1 and x_2 because the data do not make a distinction between two workers hired at the same firm (for the same position). The data consist of matches between $N = 2J$ workers and J firms. Therefore, we observe $N/2$ triples $\{(x_{2j-1}, x_{2j}, z_j)\}_{j=1}^{N/2}$, which can be used to construct empirical moments of the form

$$\psi_N = \frac{1}{N/2} \sum_{j=1}^{N/2} \Psi(x_{2j-1}, x_{2j}, z_j). \tag{12}$$

The moments discussed in equations (8) and (10) are given by particular choices for Ψ.

We now describe the value of the moment as a function of θ. 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 rewrite the sampling distribution via Bayes' rule as sampling N and J draws from the population distributions of v_i and u_i respectively, and then sampling $x_i|v_i$ 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 their respective distributions. This rewriting uses the feature that the final matches depend on the latent indices rather than directly on observable and unobservable traits. Further, conditional on the latent indices, the observable traits of two workers matched to the same firm or different firms are independent. Therefore, given the utilities $v_1,\ v_2$, and u, at parameter vector θ and any two measures m_X and m_Z for the observable traits, the value of the moment is:

$$\tilde{\psi}\left[m_X, m_Z\right](v_1, v_2, u; \theta) = \int \Psi\left(X_1, X_2, Z\right) f_{X|v_1;\theta}\left(X_1\right) f_{X|v_2;\theta}\left(X_2\right) f_{Z|u;\theta}\left(Z\right) dX_1 dX_2 dZ,$$

where $f_{X|v_1;\theta}\left(X\right)$ and $f_{Z|u;\theta}\left(Z\right)$ are the conditional densities (w.r.t. m_X and m_Z respectively) of the observable traits at θ given latent indices v and u, and m_X and m_Z. These distributions govern the observed traits of the workers and firms at any given quality.

In the limiting large market match, firms with the q-th quantile of firm quality are matched with workers on the q-th quantile of the worker quality distribution. Hence, the expected value of the moment of the q-th quantile match is given by $\tilde{\psi}$ evaluated at

$$(v_1, v_2, u) = \left(F^{-1}_{V;\theta,m_X}\left(q\right), F^{-1}_{V;\theta,m_X}\left(q\right), F^{-1}_{U;\theta,m_Z}\left(q\right)\right),$$

where $F_{V;\theta,m_X}\left(v\right)$ and $F_{U;\theta,m_Z}\left(u\right)$ are respectively the cumulative distributions of the worker and firm qualities (given θ, m_X and m_Z). This quantity must be integrated to obtain the moment as a function of the parameter θ:

$$\psi\left[m_X, m_Z\right](\theta) = \int_0^1 \tilde{\psi}\left[m_X, m_Z\right]\left(F^{-1}_{V;\theta,m_X}\left(q\right), F^{-1}_{V;\theta,m_X}\left(q\right), F^{-1}_{U;\theta,m_Z}\left(q\right); \theta\right) dq, \tag{13}$$

where $F_{V;\theta,m_X}\left(v\right) = \int_{-\infty}^{v} F_\varepsilon\left(v - h\left(X;\theta\right)\right) dm_X,\ F_{U;\theta,m_Z}\left(u\right) = \int_{-\infty}^{u} F_\eta\left(u - g\left(Z;\theta\right)\right) dm_Z$. This expression can be evaluated at any pair of measures m_X and m_Z governing the distribution of observed traits. Of particular interest are the quantities $\psi\left[\mu_X, \mu_Z\right](\theta)$ and $\psi\left[\mu_{X_N}, \mu_{Z_N}\right](\theta)$, which correspond to the values at the population and empirical measures of observables traits respectively. In this notation, the population analog of ψ_N in equation (12) is therefore $\psi\left[\mu_X, \mu_Z\right](\theta)$ evaluated at θ_0. For simplicity of notation, when referencing the moment function at populations measures μ_X and μ_Z, we will write $\psi(\theta) = \psi\left[\mu_X, \mu_Z\right](\theta)$. Similarly, when referencing their empirical analog μ_{X_N} and μ_{Z_N}, we will write $\psi_N(\theta) = \psi\left[\mu_{X_N}, \mu_{Z_N}\right](\theta)$.[12]

We now define our minimum distance estimator:

$$\widehat{\theta}_N = \arg\min_{\theta\in\Theta} \left\|\psi_N - \psi_N(\theta)\right\|_W, \tag{14}$$

where ψ_N are the moments computed from the sample as given in equation (12), $\psi_N(\theta)$ are computed from the observed sample of firms and workers as a function of θ, $\left\|\psi_N - \psi_N(\theta)\right\|_W = \left[\left(\psi_N - \psi_N(\theta)\right)' W \left(\psi_N - \psi_N(\theta)\right)\right]^{1/2}$ and W is a positive definite symmetric weight matrix. This minimum distance estimator finds the value of θ that best predicts the features of the data summarized by the moment function. For example, one can specify Ψ to summarize the overall sorting patterns and the many-to-one match moments used previously to illustrate the importance of using this information.

The next section presents conditions under which the estimator above is consistent and asymptotically normal.

4.2 Limit Properties

In this section, we outline a fairly standard set of convergence conditions on $\psi_N - \psi_N(\theta)$ and show that they imply limit properties for the estimator in equation (14). We will verify these conditions under large market asymptotics. These results are presented in the subsequent

[12] $\psi_N(\theta)$ can be approximated by first drawing ε and η to simulate $F_{N,V;\theta} = F_{V;\theta,m_{X_N}}$ and $F_{N,U;\theta} = F_{U;\theta,m_{Z_N}}$, and then using the expression in equation (13). One can also 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 μ_{X_N} and μ_{Z_N} and simulated values of v_i and u_j.

sections. We follow this organization to highlight the main ideas in the proof and clarify the contribution. We separate the conditions needed for consistency, which are weaker than those necessary for asymptotic normality of our estimator.

We require the following properties for the moment function at the population distribution of observable and unobservable traits.

Assumption 3 *(i) For any $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|\psi(\theta) - \psi(\theta_0)\|_W < \delta \Rightarrow \|\theta - \theta_0\| < \varepsilon$.*

(ii) $\psi(\theta)$ is continuously differentiable at θ_0 with an invertible Jacobian, $\psi'(\theta_0)$.

Part (i) assumes that the distance in the population $\|\psi(\theta) - \psi(\theta_0)\|_W$ is zero only if $\theta = \theta_0$. It implies that $\psi(\theta)$ identifies the parameter θ_0. Further, it requires that parameter values outside a neighborhood of the true value cannot yield a distance arbitrarily close to 0.[13] This assumption, along with the convergence condition below, will guarantee consistency of our estimator. Part (ii) is used to prove that the estimator is asymptotically normal. The commonly made assumption that the Jacobian at the limit is invertible allows us to use Taylor approximations.

We will derive limiting properties of the estimator by showing conditions under which the following properties are satisfied:

Condition 1 *(i) $(\psi_N - \psi(\theta_0)) - (\psi_N(\theta) - \psi(\theta))$ converges in probability to 0, uniformly in θ.*

(ii) a. $\sqrt{N}(\psi_N - \psi_N(\theta_0))$ converges in distribution to $N(0, \Sigma)$

b. for every sequence $\{b_N\}$ of positive numbers that converges to 0,

$$\sqrt{N} \sup_{\|\theta - \theta_0\| \leq b_N} \|(\psi_N(\theta) - \psi(\theta)) - (\psi_N(\theta_0) - \psi(\theta_0))\|_\infty = o_p(1).$$

The first conditions would follow from a uniform law of large numbers. The second condition would follow from a central limit theorem and stochastic equicontinuity. These results are not

[13] A sufficient condition for this requirement is that $\theta \in \Theta$ is compact, $\psi(\theta)$ is continuous and $\psi(\theta) = \psi(\theta_0) \Rightarrow \theta = \theta_0$.

obvious a priori because the matches depend on the composition of the entire market. The following sections prove these results under large market asymptotics. Along with Assumption 3, these conditions imply consistency and asymptotic normality of our estimator:

Theorem 4 *Suppose that the parameter space Θ is compact θ_0 lies in the interior of Θ.*

(i) If Assumption 3(i) and Condition 1(i) are satisfied, then $\widehat{\theta}_N$ converges in probability to θ_0.

(ii) If Assumption 3 and Condition 1 are satisfied, then

$$\sqrt{N}\left(\widehat{\theta}_N - \theta_0\right) \to N\left(0, \Omega\right),$$

$$\Omega = \left(\psi'\left(\theta_0\right)' C' C \psi'\left(\theta_0\right)\right)^{-1} \psi'\left(\theta_0\right) C \Sigma C' \psi'\left(\theta_0\right)' \left(\psi'\left(\theta_0\right)' C' C \psi'\left(\theta_0\right)\right)^{-1}$$

and $W = C'C$.

Proof. Part (i) follows from the arguments in Newey and McFadden (1994), Theorem 2.1. We use Theorem 3.3 in Pakes and Pollard (1989) to show part (ii). Let $G_N(\theta)$ (in the notation of Pakes and Pollard (1989)) by given by $(\psi_N - \psi_N(\theta))' C'$. Assumption 3(ii) and the definition of the estimator imply requirements (i), (ii) and (v) of Theorem 3.3 in Pakes and Pollard (1989). Requirement (iii) of Theorem 3.3 in Pakes and Pollard (1989) follows from Condition 1(ii)b. Requirements (iv) in Theorem 3.3 of Pakes and Pollard (1989) follow from Condition 1(ii)a. ■

This theorem shows that Assumption 3 and Condition 1 imply consistency and asymptotic normality in our setting. Therefore, the main difficulty in obtaining limit properties of our estimator is verifying Condition 1. This is not straightforward for two reasons. First, the triples (x_{2j-1}, x_{2j}, z_j) in the expression for our sample moments ψ_N in equation (12) are not sampled independently. This dependence occurs because their distribution is determined by the observed and unobserved characteristics of the entire sample. Second, equation (13) shows that $\psi_N(\theta) = \psi\left[\mu_{X_N}, \mu_{Z_N}\right](\theta)$ is also a function of the entire sample of observed characteristics.

To prove the required properties, we split the argument into two conceptually separate pieces. The first piece studies the distribution of sample moments ψ_N, and the second studies properties of the sample moment function $\psi_N(\theta)$. There are two reasons why this distinction helps analyze their limit distributions. First, the observed moments, ψ_N, are a function of both the sampled observed and unobserved characteristics because the realized assortative match depends on the latent indices of all agents in the market. On the other hand, $\psi_N(\theta)$, is a function only of observed traits because equation (13) shows that it is an integral with respect to the (known) distribution of unobservables. Second, ψ_N depends only on θ_0, while $\psi_N(\theta)$ is a stochastic process that must be studied uniformly in θ. The first reason complicates the analysis of the distribution of ψ_N, while the second reason complicates the analysis of $\psi_N(\theta)$.[14]

Before proceeding, we formally show that is it sufficient to treat ψ_N and $\psi[m_X, m_Z](\theta)$ as scalars.

Proposition 2 *(i) Suppose that for each $k \in \{1, \ldots, K_\Psi\}$, the k-th component of $(\psi_N - \psi(\theta_0)) - (\psi_N(\theta) - \psi(\theta))$ converges in probability to 0, uniformly in θ, then $(\psi_N - \psi(\theta_0)) - (\psi_N(\theta) - \psi(\theta))$ converges in probability to 0, uniformly in θ.*

(ii) Suppose that for any $a \in \mathbb{R}^{K_\Psi}$, $\sqrt{N}(\psi_N - \psi_N(\theta_0)) \cdot a$ converges in distribution to $N(0, a'\Sigma a)$, and for every sequence $\{b_N\}$ of positive numbers that converges to 0,

$$\sqrt{N} \sup_{\|\theta - \theta_0\| \leq b_N} \left|\left[(\psi_N(\theta) - \psi(\theta)) - (\psi_N(\theta_0) - \psi(\theta_0))\right] \cdot a\right| = o_p(1),$$

then, Condition 1(ii) is satisfied.

Proof. Part (i) follows from the definition of convergence in probability. To verify part (ii), note that Condition 1(ii) a. follows from the Cramer-Wold theorem. Condition 1(ii) b. follows from the fact that

$$\begin{aligned} & \sqrt{N} \sup_{\|\theta - \theta_0\| \leq b_N} \left\|(\psi_N(\theta) - \psi(\theta)) - (\psi_N(\theta_0) - \psi(\theta_0))\right\|_\infty \\ = & \max_{a \in \left\{e_1, \ldots, e_{K_\Psi}\right\}} \sqrt{N} \sup_{\|\theta - \theta_0\| \leq b_N} \left|\left[(\psi_N(\theta) - \psi(\theta)) - (\psi_N(\theta_0) - \psi(\theta_0))\right] \cdot a\right|, \end{aligned}$$

[14] An additional complication for analyzing the limit distribution of $\sqrt{N}(\psi_N - \psi_N(\theta))$ is that our convergence results must be joint with the empirical processes on X and Z.

where $\{e_1, \ldots, e_{K_\Psi}\}$ are the standard basis vectors of $\mathbb{R}^{K_\Psi}$. ■

The following subsections derive regularity properties under which condition Condition 1 is satisfied, assuming that Ψ is a scalar-valued function. We first analyze the limiting properties of ψ_N, and then we analyze the properties of the function $\psi_N(\theta)$.

4.3 Convergence of the Data Generating Process

The first challenge is to study the large sample properties of the sample moments, ψ_N in equation (12). The primary technical difficulty arises from the dependence of the observed matches (X_1, X_2, Z) on the observable (and unobservable) characteristics of all agents in the market. We make progress by re-writing the sampling process as one in which the utilities u and v are drawn first. This allows us to condition on the matches on latent indices in the data. The observed characteristics of the matched agents are then sampled conditional on these draws of the latent indices. This sampling process, although identical to drawing the characteristics directly from $\mu_{X,\varepsilon}$ and $\mu_{Z,\eta}$, allows for a more tractable approach to proving limit properties of the moments. The proof technique is based on using the triangular array structure implied by this process: the individual components of the triple (X_1, X_2, Z) are independent conditional on the indices drawn.

Specifically, our approach for obtaining large sample properties of ψ_N is based on the following observations. The observed characteristics X_1, X_2, and Z are a sample from $\mu_{X|v_1}$, $\mu_{X|v_2}$, and $\mu_{Z|u}$, where v_1, v_2,and u are the latent indices for these agents. The expected value of $\Psi(X_1, X_2, Z)$ given the latent indices is therefore $\tilde{\psi}[\mu_X, \mu_Z](v_1, v_2, u; \theta_0)$. Equation (13) shows that $\psi[\mu_X, \mu_Z](\theta_0)$ is the integral of $\tilde{\psi}[\mu_X, \mu_Z](v_1, v_2, u; \theta_0)$ over the population values of matched latent indices. This allows us to show that, ψ_N, which is the sample average of $\Psi(X_1, X_2, Z)$ over the matches in the data, approaches the population quantity $\psi[\mu_X, \mu_Z](\theta_0)$.

Below, we present assumptions under which we will prove our result.

Assumption 4 *(i) a.* $\tilde{\psi}[\mu_X, \mu_Z](v_1, v_2, u; \theta_0)$ *is Lipschitz continuous in* v_1, v_2 *and* u

b. The random variables ε *and* η *have continuous density with full support* *(ii) a. The derivative of $\tilde{\psi}[\mu_X, \mu_Z]\left(F_{V;\theta}^{-1}(q_1), F_{V;\theta}^{-1}(q_2), F_{U;\theta}^{-1}(q_3); \theta\right)$ with respect to $q = (q_1, q_2, q_3)$ is bounded uniformly in q, θ*

b. The random variables ε and η have continuous density with full support on $\mathbb{R}$

c. The conditional distributions of X (respectively Z) given any v (respectively u) are not degenerate

Part (i) presents conditions under which we will show that ψ_N converges to $\psi[\mu_X, \mu_Z](\theta_0)$ in probability. Part (i) a. requires Lipschitz continuity of $\tilde{\psi}[\mu_X, \mu_Z]$. This regularity condition implies that the conditional expectation of Ψ is smooth with respect to the latent indices. A more primitive condition is presented in Appendix E.1, which shows that the condition follows from bounds on the densities of X, ε and Z, η and their first derivative.[15] This regularity condition on the expectation of Ψ given the latent indices allows us to approximate the value of $\tilde{\psi}$ at the sampled latent indices for each of the matches. Part (ii) b. is a weak regularity condition on the distribution of the unobservables.

Part (ii) presents stronger assumptions, which we will use to derive the asymptotic distribution of $\sqrt{N}(\psi_N - \psi[\mu_X, \mu_Z](\theta_0))$. Part (ii) a. is analogous to (i) a., but places stronger restrictions on the sensitivity of $\tilde{\psi}$ with respect to the quality of the match. The stronger assumption ensures that $\tilde{\psi}$ is not extremely sensitive to tail behavior. Parts (ii) b. and c. are weak regularity conditions.

Our first result shows that the empirical analog in equation (12) converges at the true parameter θ_0 to ψ.

Proposition 3 *(i) If Assumption 4(i) is satisfied, then $\psi_N - \psi(\theta_0)$ converges in probability to 0.*

(ii) If Assumption 4(ii) is satisfied, then for any μ_X- and μ_Z- Donsker classes Γ_X and Γ_Z

[15] See Assumption E.1.8 and Lemma E.1.13.

of bounded functions on X and Z respectively,

$$\begin{bmatrix} \sqrt{N}\left(\psi_N - \psi\left(\theta_0\right)\right) \\ \sqrt{N}\left(\mu_{X_N} - \mu_X\right) \\ \sqrt{N/2}\left(\mu_{Z_N} - \mu_Z\right) \end{bmatrix},$$

where $\sqrt{N}\left(\mu_{X_N} - \mu_X\right)$ and $\sqrt{N/2}\left(\mu_{Z_N} - \mu_Z\right)$ are respectively empirical processes indexed by Γ_X and Γ_Z, converges to a mean zero Gaussian process (G_Ψ, G_X, G_Z) with covariance kernel V (given in Appendix B.1).

Proof. See Appendix B.1. ■

The result derives the large sample properties of $\psi_N - \psi\left[\mu_X, \mu_Z\right](\theta_0)$ based on the Assumption 4. The proof is based on studying the large sample properties of $E\left[\psi_N | v_1, \ldots, v_N, u_1, \ldots, u_{N/2}\right]$, the expectation of ψ_N given the sample of latent utilities $v_1, \ldots, v_N$ and $u_1, \ldots, u_{N/2}$. Because the observed characteristics are drawn independently given these latent indices, we are able to characterize the large sample properties of $\psi_N - E\left[\psi_N | v_1, \ldots, v_N, u_1, \ldots, u_{N/2}\right]$. Next, we show that $E\left[\psi_N | v_1, \ldots, v_N, u_1, \ldots, u_{N/2}\right]$ approximates $\psi\left(\theta_0\right)$ by appealing to the regularity and smoothness conditions in Assumption 4. We do this by relying on smoothness of $\tilde{\psi}$ and noting that the empirical quantiles of the latent indices approximate the limit quantiles. Therefore, the key to the result is that the dependence across the observed matches is only through the latent indices, and that the matching is assortative on these indices.

4.4 Differentiability of the Moment Function

The large sample results on ψ_N require evaluating the moment function only at θ_0. To study the limit properties of the estimator defined in equation (14), we need to understand the properties of the sample moment function. In this section, we derive conditions under which this map is smooth. This will allow us to use a continuous mapping theorem and the functional delta method for our results.

The approach is based on separately analyzing the behavior of $\psi_N(\theta)$ away from the tails of the latent index distribution, then showing that the tails are negligible. This approach is convenient because deriving the asymptotic distribution of the tails is technically challenging. Specifically, we will show that the functional

$$\psi^{\delta}\left[\mu_X, \mu_Z\right](\theta) = \int_{\delta}^{1-\delta} \tilde{\psi}\left[\mu_X, \mu_Z\right]\left(F_{V;\theta}^{-1}(q), F_{V;\theta}^{-1}(q), F_{U;\theta}^{-1}(q)\right)(\theta)\, dq$$

is smooth in μ_X, μ_Z for all $\delta \in (0, 1/2)$. The integral above, when evaluated at $\delta = 0$, is equal to $\psi\left[\mu_X, \mu_Z\right](\theta)$ in equation. We require the following weak assumption on the distribution of unobservable traits:

Assumption 5 *(i) f_{ε} and f_{η} are bounded and have continuous, bounded first derivatives. Further, f_{ε} and f_{η} are bounded away from zero on any compact interval of $\mathbb{R}$.*

(ii) $h(X;\theta)$ and $g(Z;\theta)$ are uniformly μ_X- and μ_Z- integrable over all $\theta \in \Theta$

Part (i) imposes a weak regularity condition that allows us to show that the conditional distributions of X and Z given the latent indices v and u vary smoothly with θ, except at extreme quantiles of the latent index distribution. This assumption is satisfied for the most commonly used parametric forms in applied analysis. Part (ii) places a weak restriction on the tail behavior of $h(X;\theta)$ and $g(Z;\theta)$ by assuming that, uniformly across θ, with high probability, these random variables belong to a compact set.

To formally state our result on smoothness of ψ^{δ}, we need to define a metric in which to measure distances in the domain and range of ψ^{δ}. We use the Banach space of vector-valued functions of $\theta \in \Theta$ endowed with the sup-norm, denoted by L_{∞}^{Θ}, as the range. We use L_{∞}^{Γ} for the domain, which is the space of measures (m_X, m_Z) endowed with the sup-norm over the class of sets Γ. We let $\Gamma = \Gamma_X \cup \Gamma_Z$, where Γ_X is a class of sets that includes

1. $\Psi(x_1, x_2, z) f_{\varepsilon}\left(F_{V;\theta}^{-1}(q) - h(x_1;\theta)\right) f_{\varepsilon}\left(F_{V;\theta}^{-1}(q) - h(x_2;\theta)\right) f_{\eta}\left(F_{U;\theta}^{-1}(q) - g(z;\theta)\right)$ and $\Psi(x_1, x_2, z) f_{\varepsilon}'\left(F_{V;\theta}^{-1}(q) - h(x_1;\theta)\right) f_{\varepsilon}\left(F_{V;\theta}^{-1}(q) - h(x_2;\theta)\right) f_{\eta}\left(F_{U;\theta}^{-1}(q) - g(z;\theta)\right)$ indexed by (x_1, z, q, θ)[16]

[16] Since the functions considered are symmetric in x_1 and x_2, we have implicitly also included the analogous class of functions, indexed by (x_2, z, q, θ).

2. $F_\varepsilon (v - h(x;\theta))$, $f_\varepsilon (v - h(x;\theta))$ and $f'_\varepsilon (v - h(x;\theta))$ indexed by (v, θ)

3. $1\{c_1 \leq x \leq c_2\}$ indexed by c_1 and c_2[17]

and Γ_Z is a class of sets that includes

1. $\Psi(x_1, x_2, z) f_\varepsilon \left(F_{V;\theta}^{-1}(q) - h(x_1;\theta)\right) f_\varepsilon \left(F_{V;\theta}^{-1}(q) - h(x_2;\theta)\right) f_\eta \left(F_{U;\theta}^{-1}(q) - g(z;\theta)\right)$ and $\Psi(x_1, x_2, z) f_\varepsilon \left(F_{V;\theta}^{-1}(q) - h(x_1;\theta)\right) f_\varepsilon \left(F_{V;\theta}^{-1}(q) - h(x_2;\theta)\right) f'_\eta \left(F_{U;\theta}^{-1}(q) - g(z;\theta)\right)$ indexed by (x_1, x_2, q, θ)

2. $F_\eta (u - g(z;\theta))$, $f_\eta (u - g(z;\theta))$ and $f'_\eta (u - g(z;\theta))$ indexed by (u, θ)

3. $1\{c_1 \leq z \leq c_2\}$ indexed by c_1 and c_2

Therefore, we will consider smoothness of the map $\psi^\delta : L_\infty^\Gamma \to L_\infty^\Theta$. The class Γ defines a norm in which we measure distances between two pairs (m_X, m_Z) and (m'_X, m'_Z). The first two groups of sets in Γ_X and Γ_Z arise from Taylor expansions of terms in the expression for ψ^δ. The last two sets are intersections of half-spaces. To use the continuous mapping theorem and the functional delta method, we will need to ensure that the empirical measures μ_{X_N} and μ_{Z_N} converge to the population measures with distance measured in this norm. The required properties on the primitives to ensure that Γ_X and Γ_Z are respectively μ_X- and μ_Z- Donsker classes are stated formally in the Online Appendix (Proposition E.3.7).

We are now ready to state the main results in this section.

Proposition 4 *If Assumption 5 is satisfied, then for each* $\delta \in \left(0, \frac{1}{2}\right)$, $\psi^\delta : L_\infty^\Gamma \to L_\infty^\Theta$ *is Hadamard differentiable tangentially to the space of bounded uniformly continuous functions at* (μ_X, μ_Z). *The Hadamard derivative at* (μ_X, μ_Z) *in the direction* (G_X, G_Z) *is* $\nabla_{(G_X, G_Z)} \psi^\delta [\mu_X, \mu_Z]$ *(given in Appendix D.2).*

Proof. See Appendix B.2 for a sketch of the proof and Appendix D.2 for details. ■

This result formalizes the idea that the small perturbations of the measures μ_X, μ_Z result in small deviations in the value of the moments (outside the tails) as a function of θ. This is

[17] If a and b are vectors, we say that $a \leq b$ if each element of a is weakly less than each element of b.

useful because we expect the empirical distributions of X and Z to be close to μ_X and μ_Z in a large sample. Assuming that tails are negligible, the result implies that the moment function in a large sample approximates the population moment function. The next section uses this result and Proposition 3 to verify Condition 1.

4.5 Verifying Condition 1

We now put together the results in the previous sections to show that Condition 1 is satisfied. First we show part (i), which implies consistency of the estimator by Theorem 4(i). We will use a continuous mapping theorem and the following assumption for this result:

Assumption 6 (i) *Γ_X and Γ_Z are respectively μ_X- and μ_Z- Glivenko Cantelli.*

This assumption implies that the expectations of functions in Γ_X and Γ_Z evaluated at the empirical measures μ_{X_N} and μ_{Z_N} respectively converge (in probability) to the population values. Further, the Glivenko-Cantelli theorem implies that the convergence is uniform over all functions in these classes. The assumption is satisfied under weak conditions on the elements of Γ_X and Γ_Z.[18] We now formally state that Condition 1(i) is satisfied for our model and sketch the proof.

Proposition 5 (i) *If Assumptions 4(i), 5 and 6(i) are satisfied, then $\psi_N - \psi_N(\theta)$ converges in probability to $\psi - \psi(\theta)$, uniformly in θ.*

Proof. See Online Appendix D.3, part (i). ■

The result shows that the difference between the empirical distance function $\psi_N - \psi_N(\theta)$ and the population analog $\psi - \psi(\theta)$ converges to zero (in probability) as the sample increases in size. The proof is proceeds by using the triangle inequality to observe that this difference is at most $|\psi_N - \psi| + |\psi_N(\theta) - \psi(\theta)|$. Proposition 3 implies that the first term, which measures the distance between the empirical and population values of the moments, converges in probability to zero. The second term, which measures the distance of the sample moment function to

[18] Proposition E.3.7 formally states conditions on primitives under which Γ_X and Γ_Z are Donsker classes.

the population function at θ, is $\psi^0\left[\mu_{X_N},\mu_{Z_N}\right](\theta)-\psi^0\left[\mu_X,\mu_Z\right](\theta)$ by definition. To show that this term also converges in probability to zero (uniformly in θ), we approximate ψ^0 with ψ^δ. Specifically, $\psi_N(\theta)$ and $\psi(\theta)$ can be approximated by $\psi^\delta\left[\mu_{X_N},\mu_{Z_N}\right](\theta)$ and $\psi^\delta\left[\mu_X,\mu_Z\right](\theta)$, where the error is of the order of δ because Ψ is bounded. Proposition 4 and Assumption 6 imply, by the continuous mapping theorem, that $\psi^\delta\left[\mu_{X_N},\mu_{Z_N}\right](\theta)$ converges in probability to $\psi^\delta\left[\mu_X,\mu_Z\right](\theta)$ uniformly in θ. Together, these observations imply the result.

The approach to a limit theorem that verifies Condition 1(ii) is similar in spirit, but technically more demanding. Proposition 3 provides a result for the term $\sqrt{N}\left(\psi_N-\psi\right)$. Our next challenge is to prove a limit theorem for $\sqrt{N}\left(\psi_N\left(\hat{\theta}\right)-\psi\left(\hat{\theta}\right)\right)$, where $\hat{\theta}$ is our estimator. We do this by approximating $\sqrt{N}\left(\psi_N\left(\hat{\theta}\right)-\psi\left(\hat{\theta}\right)\right)$ with $\sqrt{N}\left(\psi_N^\delta(\theta_0)-\psi^\delta(\theta_0)\right)$. The functional delta method and Proposition 4 imply that asymptotic distribution of $\sqrt{N}\left(\psi^\delta(\theta_0)-\psi_N^\delta(\theta_0)\right)$ is given by $\nabla_G\psi^\delta(\theta_0)=\left(\nabla\psi^\delta\circ G\right)(\theta_0)$, where G is a mean zero Gaussian process on L_∞^Γ. The remaining term is the approximation error $\left(\nabla_G\psi^\delta-\nabla_G\psi^0\right)(\theta_0)$. Therefore, we need to ensure that the errors in approximating $\nabla_G\psi^0(\theta_0)$ with $\nabla_G\psi^\delta(\theta_0)$ and approximating $\sqrt{N}\left(\psi_N(\theta)-\psi(\theta)\right)$ in a neighborhood of θ_0 with $\sqrt{N}\left(\psi_N^\delta(\theta_0)-\psi^\delta(\theta_0)\right)$ are negligible. Ensuring that these errors do not affect the limit distribution of $\sqrt{N}\left(\left(\psi_N-\psi\right)-\left(\psi_N(\theta)-\psi(\theta)\right)\right)$ requires tighter controls of the tails than our consistency result. Specifically, the limit theorem requires us to replace Assumption 6(i) with the following stronger requirement:

Assumption 6 (ii) *a. Γ_X and Γ_Z are respectively μ_X- and μ_Z- Donsker.*

b. for every sequence $\{b_N\}$ of positive numbers that converges to 0,

$$\sqrt{N}E\sup_{\|\theta-\theta_0\|\le b_N}\left|\left(\psi_N(\theta)-\psi(\theta)\right)-\left(\psi_N^\delta(\theta)-\psi^\delta(\theta)\right)\right|$$

converges to zero as $\delta\to 0$ and $N\to\infty$

c. for fixed $\delta\in\left(0,\frac{1}{2}\right)$ and every sequence $\{b_N\}$ of positive numbers that converges to 0,

$$\sup_{\|\theta-\theta_0\|\le b_N}\left|\nabla_G\psi^\delta(\theta)-\nabla_G\psi^\delta(\theta_0)\right|$$

converges in probability to zero as $N \to \infty$

d. $\left(\nabla_G \psi^\delta - \nabla_G \psi^0\right)(\theta_0)$ *converges in probability to zero as* $\delta \to 0$.

Part a. strengthens Assumption 6(i) to allow a functional central limit theorem over the classes Γ_X and Γ_Z. Parts b. and d. are technical assumptions that ensure that tails are negligible. Part b. controls the rate at which the dependence of the moment function on the tails vanishes with the sample size. Part d. assumes that tails have a negligible contribution to the dependence of the moment function on perturbations of the data. Part c. assumes that the process $\nabla_G \psi^\delta(\theta)$ is well-behaved in a neighborhood of θ_0. For completeness, the Online Appendix derives primitive conditions under which each of these requirements are satisfied. Specifically, Theorem E.2.5 shows that smoothness conditions and bounds on the tail behavior of the primitives imply these requirements. Assumption c. is relatively straightforward to verify and is based on showing that $\nabla_G \psi^\delta(\theta)$ have sample paths continuous in θ by bounding the L^2 covering numbers of the related Gaussian process. Assumption d. follows from showing that an upper bound on the variance of $\left(\nabla_G \psi^\delta - \nabla_G \psi^0\right)(\theta_0)$ converges to 0 as $\delta \to 0$. Verifying assumption b. is the most difficult technical aspect of proving our limit theorem and requires relatively novel proof techniques.

The difficulty in verifying assumption b follows from the fact that $\sqrt{N}\left(\psi_N(\theta) - \psi(\theta)\right)$ is a nonlinear function of the empirical measures $\left(\mu_{X_N}, \mu_{Z_N}\right)$. While the functional delta method is a conceptually straightforward approach to proving a limit theorem for $\sqrt{N}\left(\psi_N^\delta(\theta) - \psi^\delta(\theta)\right)$ with $\delta \in (0,1)$, showing that the tails are negligible requires a proof by first principles. Although direct computations play a large part in this proof, the conceptual core is a modification of the method of chaining with adaptive truncation exposited by Pollard (2002), where it is used to prove Ossiander's bracketing limit theorem for empirical processes. Our proof technique follows a similar approach as Pollard (2002) by similarly approximating Θ using finite subsets of increasing size and similar truncation techniques. After a suitable truncation, the moment generating function of the increments of an empirical process can be bounded using techniques that apply to sums of independent random variables. Because the increments of

$(\psi_N(\theta) - \psi(\theta))$ have no simple expression, we use the concentration of measure inequality of Boucheron et al. (2003) in order to get the needed bound on the moment generating function. This application of an abstract concentration of measure inequalities within the broader context of a chaining argument may be a more generally useful technique for proving functional limit theorems. This approach is necessary due to the dependent data nature of our problem, which makes standard empirical process techniques for i.i.d. data inapplicable. This feature of our model may be shared with other contexts such as network formation models.

The control of tail behavior implied by these results allow us to verify Condition 1(ii). Formally, we have:

Proposition 5 (ii) *If Assumptions 4(ii), 5 and 6(ii) are satisfied, then Condition 1(ii) is satisfied.*

Proof. See Appendix D.3, part (ii). ■

As discussed earlier, the basic ideas are similar to the consistency result proved earlier, with a more technically demanding method for handling the approximation in the tails. Proposition 5 shows Condition 1 for our model. Therefore, we can use Theorem 4 to assure consistency and asymptotic normality of the minimum distance estimator.

5 Monte Carlo Evidence

This section presents Monte Carlo experiments to assess the properties of a method of simulated moments estimator. The results are presented for a simulation based estimator of the form

$$\begin{aligned} \hat{\theta}_N &= \arg\min_{\theta\in\Theta} \left\| \psi_N - \psi_{N,S}(\theta) \right\|_W & (15) \\ &= \arg\min_{\theta\in\Theta} \left[\left(\psi_N - \psi_{N,S}(\theta)\right)' W \left(\psi_N - \psi_{N,S}(\theta)\right) \right]^{1/2}, & (16) \end{aligned}$$

where ψ_N is as defined in equation (12) 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, and set $\psi_{N,S}(\theta) = \frac{1}{S}\sum_s \psi_{N,s}(\theta)$. The moments used are as defined in equations (8) and (10). We include an "overall moment" of the form in equation (8) for each component of x interacted with each component of z. A "within moment" of the form in equation (10) is included for each observed component of x.

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 two characteristics for residents and up to four characteristics for programs. The characteristics z_j of program j are distributed as

$$z_j = (z_{j1}, z_{j2}) \sim N(a, I_2),$$

where $a = (1, 2)$ and I_2 is a 2×2 identity matrix. Similarly, the characteristics of the residents, x_i, are distributed as

$$x_i = (x_{i1}, x_{i2}) \sim N(a, I_2).$$

For each model specification, we generate 500 samples indexed by b and parameter estimates $\hat{\theta}_b$.[19] The confidence intervals are generated by using a parametric bootstrap described in Online Appendix F.

The preferences are of the form,

$$\begin{aligned} v_i &= x_i\alpha + \varepsilon_i && (17) \\ u_j &= z_j\beta + \eta_j, && (18) \end{aligned}$$

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

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

squared error (RMSE), and the standard error fall with J and $\bar{c}$ 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. Also notice that estimates for α are more precise than estimates of β in both specifications and all sample size.

6 Conclusion

This paper provides results on the identification and estimation of preferences from data from a matching market with positive assortative matching on a latent index when data only on matches are observed. Our results apply to both transferable and non-transferable utility models of matching. 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. These identification results use insights from the analysis of non-linear measurement error models. Intuitively, the observable characteristics of the multiple agents with the same match partner can be seen as noisy measures of the quality of the agents in the match.

We then prove consistency and $\sqrt{N}-$asymptotic normality of an estimator for a parametric class of models. Our limit theorems are based on several insights in this model. First, we use the fact that the matches are determined by the latent indices and that the observables are conditionally independent given these indices. Second, we show that the moment function is smooth in the distribution of observables, except at the extreme quantiles of the latent index. Third, we show that approximating this function by ignoring the tails has a negligible effect on the asymptotic distribution of the estimator using a general concentration of measure inequality for dependent data. 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 also be possible to use variation in the characteristics of participants across markets for identification. This can be particularly important for the empirical study of marriage markets. Our results are also restricted to a single latent index model on each side 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 it is likely technically challenging. It may be particularly difficult to analyze both transferable and non-transferable utility models in a single framework. Finally, we have also left the exploration of computationally more tractable estimators for future research.

References

Agarwal, N. (2015). An Empirical Model of the Medical Match. *American Economic Review 105*(7), 1939–78.

Agarwal, N. (2017). Policy Analysis in Matching Markets. *American Economic Review, Papers and Proceedings, Forthcoming*.

Agarwal, N. and W. Diamond (2014). Identification and Estimation in Two-sided Matching Markets. *Cowles Foundation Discussion Paper No. 1905*.

Azevedo, E. M. and E. Budish (2013). Strategyproofness in the large. *Chicago Booth, mimeo*.

Becker, G. S. (1973). A Theory of Marriage: Part I. *Journal of Political Economy 81*(4), 813 – 46.

Boucher, V. and I. Mourifie (2012, aug). My Friend Far Far Away: Asymptotic Properties of Pairwise Stable Networks. *SSRN Electronic Journal*.

Boucheron, S., G. Lugosi, and P. Massart (2003, jul). Concentration inequalities using the entropy method. *The Annals of Probability 31*(3), 1583–1614.

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.

Chandrasekhar, A. G. and M. O. Jackson (2015, oct). A Network Formation Model Based on Subgraphs.

Chiappori, P.-A., M. Iyigun, and Y. Weiss (2009, dec). Investment in Schooling and the Marriage Market. *American Economic Review 99*(5), 1689–1713.

Chiappori, P.-A. and I. Komunjer (2008, dec). Correct Specification and Identification of Nonparametric Transformation Models. *Columbia University, mimeo*.

Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2012). Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market. *Journal of Political Economy 120*(4), 659–695.

Chiappori, P.-A. and B. Salanié (2015). The Econometrics of Matching Models. *Journal of Economic Literature, forthcoming*.

Chiappori, P.-A., B. Salanié, and Y. Weiss (2015). Partner Choice and the Marital College Premium. *CEPR Discussion Paper No. DP10403*.

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. *Contributions in Theoretical Economics 6*(1), 1–28.

Csorgo, M. and P. Revesz (1978, jul). Strong Approximations of the Quantile Process. *The Annals of Statistics 6*(4), 882–894.

Dudley, R. M. (2014). *Uniform Central Limit Theorems*. Cambridge University Press.

Dunford, N. and J. T. Schwartz (1971). *Linear Operators*. New York: Wiley.

Dupuy, A. and A. Galichon (2014). Canonical Correlation and Assortative Matching: A Remark. *IZA Discussion Papers*.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004). Identification and Estimation of Hedonic Models. *Journal of Political Economy 112*(S1).

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. *University of Michigan, mimeo*.

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

Galichon, A., S. D. Kominers, and S. Weber (2014). Costly Concessions: An Empirical Framework for Matching with Imperfectly Transferable Utility.

Galichon, A. and B. Salanie (2012). Cupid's Invisible Hand: Social Surplus and Identification in Matching Models. *Columbia University, mimeo*.

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

Graham, B. S. (2014, jul). An Econometric Model of Link Formation with Degree Heterogeneity.

Hsieh, Y.-W. (2011). Understanding Mate Preferences from Two-Sided Matching Markets: Identification, Estimation and Policy Analysis. *University of Southern California, mimeo*.

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, jan). 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, jan). Marriage, honesty, and stability. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 53–62. Society for Industrial and Applied Mathematics.

Jiang, Y. (2016). Matching Games with Unobserved Heterogeneity: A Structural Analysis of Online Labor Markets.

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

Leung, M. P. (2015, sep). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics 188*(1), 182–195.

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.

Mattner, L. (1993, dec). Some Incomplete But Boundedly Complete Location Families. *The Annals of Statistics 21*(4), 2158–2162.

Menzel, K. (2011). Robust Decisions For Incomplete Structural Models Of Social Interactions. *New York University, mimeo*.

Menzel, K. (2015). Large Matching Markets As Two-Sided Demand Systems. *Econometrica 83*(3), 897–941.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*, 2111–2245.

Niederle, M. and L. Yariv (2009). Decentralized Matching with Aligned Preferences. *Stanford University, mimeo*.

Pakes, A. and D. Pollard (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica 57*(5), 1027 – 57.

Pollard, D. (2002, dec). Maximal inequalities via bracketing with adaptive truncation. *Annales de l'Institut Henri Poincare (B) Probability and Statistics 38*(6), 1039–1052.

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

Shi, X. and M. Shum (2014). On the Empirical Content of the Beckerian Marriage Model. *Caltech, mimeo*.

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 (2013). Entry by Merger: Estimates from a Two-Sided Matching Model with Externalities. *Yale University, mimeo*.

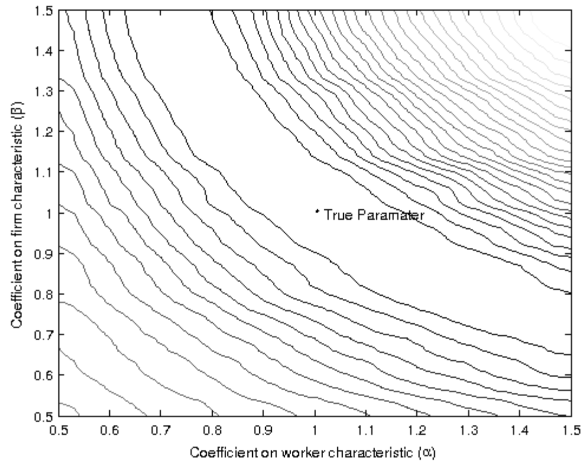
van der Vaart, A. and J. Wellner (2000). *Weak Convergence and Empirical Processes: With Applications to Statistics (Springer Series in Statistics)* (Corr. 2. p ed.). Springer.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

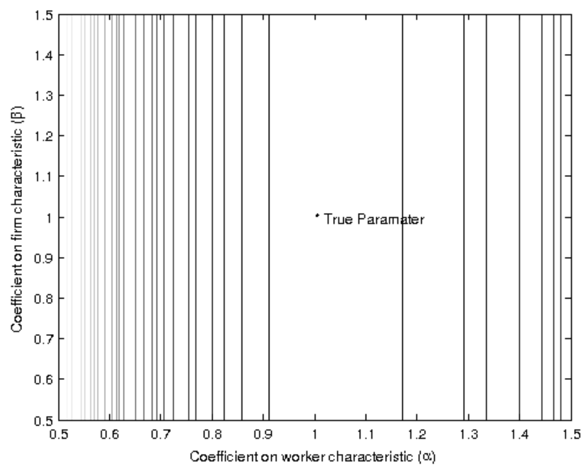
Vissing, A. (2016). One-to-Many Matching with Complementary Preferences: An Empirical Study of Natural Gas Lease Quality and Market Power. Technical report, Duke University, Durham.

Table 1: Monte Carlo Evidence: Double-Vertical Model

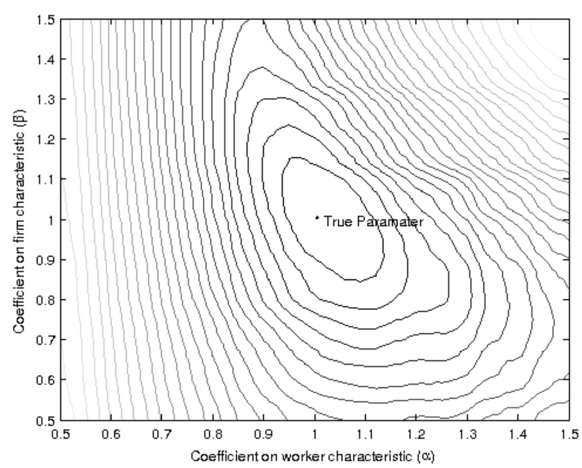
	One Characteristic (1)		Two Characteristics (2)			
	$\alpha_1\,(x_1)$	$\beta_1\,(z_1)$	$\alpha_1\,(x_1)$	$\alpha_2\,(x_2)$	$\beta_1\,(z_1)$	$\beta_2\,(z_2)$
$J = 100,\ \bar{c} = 5$						
True Par.	1	1	1	2	-1	2
Bias	0.005	0.053	0.033	0.042	-0.593	1.134
RMSE	0.093	0.239	0.270	0.412	1.063	1.938
SE	0.131	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



(a) Sorting Moments Only



(b) Within-program Moments Only



(c) All Moments

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

Appendix: Latent Indices in Assortative Matching Models[20]

William Diamond, Harvard University

Nikhil Agarwal, MIT and NBER

A Proofs: Identification

A.1 Proof of Lemma 1

We present the argument for the identification of the level sets of $h(\cdot)$ since the proof for $g(\cdot)$ is identical. The cdf of v conditional on $h(x)$ is given by $F_{V|h(x)}(v) = F_\varepsilon(v - h(x))$. Note that $F_{V|h(x)}(v)$ is increasing in v and decreasing in $h(x)$. Let $F_{q|h(x)}(q|h(x)) = F_{V|h(x)}\left(F_V^{-1}(q) - h(x)\right)$ be the cdf of the quantile of v given $h(x)$. Since F_V^{-1} is an increasing function of q, $F_{q|h(x)}(q|h(x))$ is increasing in q and decreasing in $h(x)$. As noted in Remark 1, the q-th quantile of each side matches with the q-th quantile of the other. Therefore, the density of $g(Z)$ that $h(x)$ is matched with is given by

$$\begin{aligned} f_{g(Z)|h(x)}(g|h) &= \int_0^1 f_{g(Z)|q}(g|q)\, f_{q|h(x)}(q|h)\, dq \\ &= \int_0^1 f_{q|g}(q|g)\, f_g(g)\, f_{q|h(x)}(q|h)\, dq \\ &= \int f_\eta(u-g)\, f_g(g)\, f_{q|h(x)}(F_U(u)|h)\, du, \end{aligned}$$

where $f_g(\cdot)$ is the density of $g(Z)$. The second equality uses Bayes' rule. The last equality follows from a change of variables $q = F_U(u)$ and the fact that $f_{q|g}(F_U(u)|g) = \frac{f_\eta(u-g) f_g(g)}{f_U(u)}$. Since $f_g(g) > 0$ for all g, and f_η has a non-vanishing characteristic function, $f_{g(Z)|h(x)}(\cdot|h)$ is injective in h. Since $F_{q|h(x)}(q|h)$ is decreasing in h, if $h(x') > h(x)$, then $f_{q|h(x)}(q|h(x')) \neq f_{q|h(x)}(q|h(x))$ for some q. Hence, we have that $f_{g(Z)|h(x)}(g|h(x')) \neq f_{g(Z)|h(x)}(g|h(x))$ if $h(x') \neq h(x)$. If $Z|x \sim Z|x'$ then $g(Z)|x \sim g(Z)|x'$. Therefore, it must be that the distribution of Z given x differs from the distribution of Z given x'. Therefore, the level sets of $h(\cdot)$ are identified.

A.2 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 identify $f_{x|q}(x|q)$ and $f_{z|q}(z|q)$, which are the conditional densities of x and z respectively given $h(x) + \varepsilon = F_U^{-1}(q)$ and $g(z) + \eta = F_V^{-1}(q)$, where q is the quantile of the latent index.[21] Lemma C.2.2 implies that the primitives $h(\cdot)$, $g(\cdot)$, f_η and f_ε, are identified from $f(x|q)$ and $f(z|q)$.

We begin by verifying Assumptions HS.2-HS.4. Assumption HS.2 requires $f_{z|x_1,x_2,q}(z|x_1,x_2,q) = f_{z|q}(z|q)$, and $f_{x_1|z,x_2,q}(x_1|z,x_2,q) = f_{x_1|q}(x_1|q)$. This is satisfied since the quantile of the latent index q is a sufficient

[20] Refer to the Online Appendix for Appendices C - F.

[21] 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.

statistic for the distribution of observable characteristics in any match.

Assumption HS.3 requires that $L_{x|q}$ and $L_{x_1|x_2}$ are injective, where $L_{x|q}(m) = \int_0^1 f_{x|q}(x|q)\, m(q)\, dq$ and $L_{x_1|x_2}(m) = \int_0^1 f_{x_1|x_2}(x_1|x_2)\, m(x_2)\, dx_2$. Lemmas C.2.4 and C.2.5 imply that under Assumption 1, $L_{x|q}$ and $L_{x_1|x_2}$ are injective.

Assumption HS.4 requires that for all q_1 and q_2 in $[0,1]$, the set $\left\{z : f_{z|q}(z|q_1) \neq f_{z|q}(z|q_2)\right\}$ has positive probability (under the marginal distribution of z) if $q_1 \neq q_2$. This assumption is satisfied since

$$\begin{aligned} f_{z|q}(z|q) &= \frac{f_{q|z}(q|z)\, f_z(z)}{f_q(q)} \\ &= f_{q|z}(q|z) \\ &= \frac{1}{f_U\left(F_U^{-1}(q)\right)} f_\eta\left(F_U^{-1}(q) - g(z)\right) \end{aligned}$$

is complete (Lemma C.2.3). The first equality follows from Bayes' rule, the second equality uses the fact that z and q are uniformly distributed, and the third equality transforms $u = F_U^{-1}(q)$, using the fact that $f_{u|z}(u|z) = f_\eta(u - g(z))$.

For a function $m(\cdot)$, and any z and q, define the operator $\Delta_{z;q} m(q) = f_{z|q}(z|q)\, m(q)$ as in HS. Since $f(z, x_1|x_2)$ is observed, for any real valued function m and z, we can compute

$$L_{z;x_1|x_2}(m) = \int_0^1 f(z, x_1|x_2)\, m(x_2)\, dx_2 = L_{x_1|q} \circ \Delta_{z;q} \circ L_{q|x_2}(m)$$

as shown in HS. They then use Assumption HS.1 to show that (i) $L_{x_1|x_2}^{-1}$ exists and is densely defined, and (ii) $T = L_{z,x_1|x_2} L_{x_1|x_2}^{-1}$ has a unique spectral decomposition. Lemmas C.2.5 and C.2.6 respectively show that these results follow under our assumptions (the conditions needed for Lemma C.2.6 are verified in Lemmas C.2.5 and C.2.4). Hence, the conditional densities $f_{z|q}(z|q)$ and $f_{x|q}(x|q)$ are identified up to a reindexing via a bijection $Q(\cdot)$ where $\tilde{q} = Q(q)$. That is, for every pair $\tilde{f}_{x|q}$ and $\tilde{f}_{z|q}$ satisfying our regularity conditions that can rationalize $f(z, x_1|x_2)$, the proof of Theorem 1 in HS shows that there exist bijections $Q_x : [0,1] \to [0,1]$ and $Q_z : [0,1] \to [0,1]$ such that $f_{x|Q_x(q)} = \tilde{f}_{x|q}$ and $f_{z|Q_z(q)} = \tilde{f}_{z|q}$.

This remaining under-identification issue is referred to as the ordering/indexing ambiguity issue in HS. They solve this ambiguity by using Assumption HS.5, which assumes that there is a known functional M such that $M\left[f_{x|q}(\cdot|q)\right] = q$ for all q. Since our model does not deliver such a functional, we instead solve the ordering/indexing ambiguity by using the fact that in our model, the q indexes the quantiles of the latent index and $f_{x|q}$ and f_q must therefore satisfy certain known properties. Specifically, we use Lemma C.2.7 to show directly, that Q_x and Q_z must be the identity function under the assumptions of our model. To apply Lemma C.2.7, we need to show that $f_{x|q}(x|q) = \frac{f_\varepsilon\left(F_V^{-1}(q) - h(x)\right)}{f_V\left(F_V^{-1}(q)\right)}$ and $f_{z|q}(z|q) = \frac{f_\eta\left(F_U^{-1}(q) - g(z)\right)}{f_U\left(F_U^{-1}(q)\right)}$ where q are quantiles satisfies Condition C.2.2. Since the proof is symmetric, we show this only for $f_{x|q}(x|q)$. Condition C.2.2(i) is satisfied since $f_{x|q}(x|q)$ is complete (Lemma C.2.3).

To verify condition C.2.2(ii), we compute $\frac{\partial f_{x|q}(x|q)}{\partial q}$. Note that $f_{x|q}(x|q) = f_{q|x}(q|x) f_x(x)/f_q(q) = f_{q|x}(q|x)$ by Bayes' rule and the (normalized) marginal distributions of x and q. Therefore,

$$\begin{aligned} \frac{\partial f_{x|q}(x|q)}{\partial q} = \frac{\partial f_{q|x}(q|x)}{\partial q} &= \frac{\partial}{\partial q} \frac{f_\varepsilon\left(F_V^{-1}(q) - h(x)\right)}{f_V\left(F_V^{-1}(q)\right)} \\ &= \frac{f_V(F_V^{-1}(q)) f_\varepsilon'(F_V^{-1}(q) - h(x)) - f_\varepsilon(F_V^{-1}(q) - h(x)) f_V'(F_V^{-1}(q))}{f_V(F_V^{-1}(q))^3}. \end{aligned}$$

Therefore, Condition C.2.2(ii) follows from Assumption 1 since each of the terms is finite and $f_V(v) > 0$ since

ε and $h(X)$ have full support on $\mathbb{R}$.

We can verify Condition C.2.2(iii) by showing that for each $q \in (0,1)$, there exists x such that $\frac{d}{dq} f_{q|x}(q|x) \neq 0$. Towards a contradiction, for a given $q \in (0,1)$, assume that $\frac{d}{dq} f_{q|x}(q|x) = 0$ for all x. As shown above, $\frac{d}{dq} f_{q|x}(q|x) = \frac{d}{dq} \frac{f_\varepsilon(F_V^{-1}(q) - h(x))}{f_V(F_V^{-1}(q))}$. Since $f_V(v) > 0$, $\frac{d}{dq} f_{q|x}(q|x) = 0$ for all x if and only if $\frac{d}{dv} \frac{f_\varepsilon(v - h(x))}{f_V(v)}$ evaluated at $v = F_V^{-1}(q)$ if zero for all x, It must therefore be that

$$\frac{d}{dv} \frac{f_\varepsilon(v - h(x))}{f_V(v)} = \frac{f_V(v) f'_\varepsilon(v - h(x)) - f_\varepsilon(v - h(x)) f'_V(v)}{f_V(v)^2}$$

is zero for all x for each $v \in (-\infty, \infty)$. Since $f_V(v) > 0$, it must be that $f_V(v) f'_\varepsilon(v - h(x)) = f_\varepsilon(v - h(x)) f'_V(v)$ for all x. Since $h(x)$ has full support on $\mathbb{R}$, this implies that $f'_\varepsilon(\varepsilon) = K_1 f_\varepsilon(\varepsilon)$ for all $\varepsilon \in (\infty, \infty)$. Hence, $f_\varepsilon(\varepsilon) = K_2 \exp(K_1 \varepsilon)$ for a constants K_1 and K_2. Note that f_ε is a density with full support, which is a contradiction with this functional form.

Condition C.2.2(iv) is definitional for the particular model considered since q indexes quantiles. Condition C.2.2(v) follows from Lemma C.2.4 under Assumption 1. Conditions C.2.2(vi) is also definitional in our case since $f_{x|q}$ are conditional densities and q indexes quantiles. We have thus verified Condition C.2.2 for $f_{x|q}$. An identical argument follows for $f_{z|q}$. Therefore, by Lemma C.2.7, Q_x and Q_z are the identity function. Hence, we have identified $f_{x|q}$ and $f_{z|q}$.

B Proofs: Estimation

B.1 Proof of Proposition 3

We first rewrite

$$\psi_N - \psi = \left(\psi_N - E\left(\psi_N | \mu_{V_N}, \mu_{U_N}\right)\right) + \left(E\left(\psi_N | \mu_{V_N}, \mu_{U_N}\right) - \psi\right).$$

Proof of Part (i): Lemma D.1.9(i) shows that if Assumption 4(i) is satisfied, $E\left(\psi_N | \mu_{V_N}, \mu_{U_N}\right) - \psi$ converges in probability to 0 as $N \to \infty$. This result is proved by rewriting

$$\begin{aligned} E\left(\psi_N | \mu_{V_N}, \mu_{U_N}\right) &= \frac{1}{N/2} \sum_{k=1}^{N/2} \tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{2k-1}{N}\right), F_{V_N}^{-1}\left(\frac{2k}{N}\right), F_{U_N}^{-1}\left(\frac{k}{N/2}\right)\right) \\ &= \frac{1}{N} \sum_{i=1}^{N} \tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{U_N}^{-1}\left(\frac{i}{N}\right)\right) + R, \end{aligned} \tag{B.1.19}$$

where F_{V_N} and F_{U_N} are the cdfs representing the empirical measures μ_{V_N} and μ_{U_N} respectively, and R is a remainder term. We then show that R and

$$\frac{1}{N} \sum_{i=1}^{N} \tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{V_N}^{-1}\left(\frac{i}{N}\right), F_{U_N}^{-1}\left(\frac{i}{N}\right)\right) - \psi$$

converge in probability to zero. Lemma D.1.10(i) shows that $\psi_N - E\left(\psi_N | \mu_{V_N}, \mu_{U_N}\right)$ converges in probability to zero by bounding its variance by $\frac{1}{J} 4 \|\Psi\|_\infty^2$. Since $\psi_N - \psi = \left(\psi_N - E\left(\psi_N | \mu_{V_N}, \mu_{U_N}\right)\right) + \left(E\left(\psi_N | \mu_{V_N}, \mu_{U_N}\right) - \psi\right)$ is the sum of two terms that converge in probability to 0, the result follows directly from Slutsky's theorem.

Proof of Part (ii): Lemma D.1.9(ii) shows if Assumption 4(ii) is satisfied, then for any bounded μ_X-

Donsker class Γ_X and for any bounded μ_Z-Donsker class Γ_Z,

$$\left[\sqrt{N}\left(E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)-\psi\right),\sqrt{N}\left(\mu_{X_N}-\mu_X\right)\left(\gamma_X\right),\sqrt{N/2}\left(\mu_{Z_N}-\mu_Z\right)\left(\gamma_Z\right)\right]$$

indexed by $\gamma_X\in\Gamma_X$ and $\gamma_Z\in\Gamma_Z$ is asymptotically equivalent to

$$\begin{bmatrix} \sqrt{N}\int_0^1 \nabla\tilde{\psi}_q\left(q,q,q\right)\cdot\begin{bmatrix} \left(\mu_{(X,\varepsilon)_N}-\mu_{X,\varepsilon}\right)\left(1\left\{h\left(x;\theta_0\right)+\varepsilon\leq F_V^{-1}\left(q_X\right)\right\}\right) \\ \left(\mu_{(X,\varepsilon)_N}-\mu_{X,\varepsilon}\right)\left(1\left\{h\left(x;\theta_0\right)+\varepsilon\leq F_V^{-1}\left(q_X\right)\right\}\right) \\ \left(\mu_{(Z,\eta)_N}-\mu_{Z,\eta}\right)\left(1\left\{g\left(z;\theta_0\right)+\eta\leq F_U^{-1}\left(q_Z\right)\right\}\right)\end{bmatrix}dq \\ \sqrt{N}\left(\mu_{X_N}-\mu_X\right)\left(\gamma_X\right) \\ \sqrt{N/2}\left(\mu_{Z_N}-\mu_Z\right)\left(\gamma_Z\right)\end{bmatrix}, \qquad \text{(B.1.20)}$$

which converges weakly to a mean-zero Gaussian process with a covariance kernel V'. This covariance kernel is derived by using equation (B.1.19) to show that $\sqrt{N}R$ converges in probability to zero, and then analyzing

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{\psi}\left(F_{V_N}^{-1}\left(\frac{i}{N}\right),F_{V_N}^{-1}\left(\frac{i}{N}\right),F_{U_N}^{-1}\left(\frac{i}{N}\right)\right)-\psi\right)$$

using Taylor approximations. Since $\left\|\nabla\tilde{\psi}_q\right\|_\infty<\infty$, the expression in (B.1.20) is a sum of $\mu_{X,\varepsilon}$- and $\mu_{Z,\eta}$-Donsker classes because we have added a finite number of sums of i.i.d. random variables to Γ_X and Γ_Z. Let $\Gamma_{X,\varepsilon}$ and $\Gamma_{Z,\eta}$ be the index sets for this empirical process. Lemma D.1.11 shows that if Assumption 4(ii) is satisfied, then for any bounded $\mu_{X,\varepsilon}$-Donsker class $\Gamma_{X,\varepsilon}$ and for any bounded $\mu_{Z,\eta}$-Donsker class $\Gamma_{Z,\eta}$,

$$\left[\sqrt{N}\left(\psi_N-E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)\right),\sqrt{N}\left(\mu_{(X,\varepsilon)_N}-\mu_{X,\varepsilon}\right)\left(\gamma_{X,\varepsilon}\right),\sqrt{N/2}\left(\mu_{(Z,\eta)_N}-\mu_{Z,\eta}\right)\left(\gamma_{Z,\eta}\right)\right]$$

indexed by $\gamma_{X,\varepsilon}\in\Gamma_{X,\varepsilon}$ and $\gamma_{Z,\eta}\in\Gamma_{Z,\eta}$ converges weakly to a mean-zero Gaussian process with a covariance kernel V''. To prove this result, we first compute the joint moment generating function for particular elements, $\gamma_{X,\varepsilon}$ and $\gamma_{Z,\eta}$, to show that it approaches the moment generating function of a mean-zero normal random variable, and derive the covariance $V''\left(\gamma_{X,\varepsilon},\gamma_{Z,\varepsilon}\right)$. We then verify equicontinuity of the process to show weak convergence.

Therefore, applying this result to the process indexed by $\Gamma_{X,\varepsilon}$ and $\Gamma_{Z,\eta}$, we have that the process

$$\begin{bmatrix} \sqrt{N}\left(E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)-\psi\right)+\sqrt{N}\left(\psi_N-E\left(\psi_N|\mu_{V_N},\mu_{U_N}\right)\right) \\ \sqrt{N}\left(\mu_{X_N}-\mu_X\right)\left(\gamma_X\right) \\ \sqrt{N/2}\left(\mu_{Z_N}-\mu_Z\right)\left(\gamma_Z\right)\end{bmatrix}$$

indexed by $\gamma_X\in\Gamma_X$ and $\gamma_Z\in\Gamma_Z$ converges weakly to a mean zero Gaussian process with covariance kernel V.

We now compute V. Note that $V\left(\gamma_\Psi,\gamma_Z\right)=V'\left(\gamma_\Psi,\gamma_Z\right)+\sqrt{2}V''\left(\gamma_\Psi,\gamma_Z\right)$ and $V\left(\gamma_\Psi,\gamma_X\right)=V'\left(\gamma_\Psi,\gamma_X\right)+2V''\left(\gamma_\Psi,\gamma_X\right)$ since covariance is bilinear. $V\left(\gamma_\Psi,\gamma_\Psi\right)=V'\left(\gamma_\Psi,\gamma_\Psi\right)+2V''\left(\gamma_\Psi,\gamma_\Psi\right)$ since $Cov\left(X-E\left[X|\mathcal{I}\right],E\left[X|\mathcal{I}\right]-E\left[X\right]\right)=0$ for any sigma-field $\mathcal{I}$ by the law of iterated expectations. Finally, by definition, $V\left(\gamma_X,\gamma_Z\right)=0$, $V\left(\gamma_X,\gamma_X'\right)=V'\left(\gamma_X,\gamma_X'\right)$ and $V\left(\gamma_Z,\gamma_Z'\right)=V'\left(\gamma_Z,\gamma_Z'\right)$. The remaining elements are $V\left(\gamma_\Psi,\gamma_X\right)=V'\left(\gamma_\Psi,\gamma_X\right)+2V''\left(\gamma_\Psi,\gamma_X\right)$, $V\left(\gamma_\Psi,\gamma_\Psi\right)=V'\left(\gamma_\Psi,\gamma_\Psi\right)+2V''\left(\gamma_\Psi,\gamma_\Psi\right)$ and $V\left(\gamma_\Psi,\gamma_Z\right)=V'\left(\gamma_\Psi,\gamma_Z\right)+\sqrt{2}V''\left(\gamma_\Psi,\gamma_Z\right)$, where V' and V'' are as defined in Lemmata D.1.9 and D.1.10 respectively.

B.2 Proof Sketch for Proposition 4

Consider a sequence of measures $\left(\mu_{X_N}, \mu_{Z_N}\right)$ and scalars $h_N \to 0$ such that $\frac{1}{h_N}\left(\mu_{X_N} - \mu_X, \mu_{Z_N} - \mu_Z\right)$ converges to $G = (G_X, G_Z)$ uniformly in L_∞^Γ, where G is bounded and uniformly continuous. The Hadamard derivative is the limit of

$$
\begin{aligned}
& \frac{1}{h_N}\left[\tilde{\psi}^\delta\left[\mu_X, \mu_Z\right](\theta) - \tilde{\psi}^\delta\left[\mu_{X_N}, \mu_{Z_N}\right](\theta)\right] && \text{(B.2.21)}\\
= \quad & \frac{1}{h_N}\left[\int_\delta^{1-\delta} \frac{\int \Psi\left(x_1, x_2, z\right) \phi_\varepsilon\left(q, x_1; \theta\right) \phi_\varepsilon\left(q, x_2; \theta\right) \phi_\eta(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z}{\int \phi_\varepsilon\left(q, x_1; \theta\right) \phi_\varepsilon\left(q, x_2; \theta\right) \phi_\eta(q, z; \theta)\, d\mu_{X_1} d\mu_{X_2} d\mu_Z} dq\right. && \text{(B.2.22)}\\
& \left.- \int_\delta^{1-\delta} \frac{\int \Psi\left(x_1, x_2, z\right) \phi_{\varepsilon,N}\left(q, x_1; \theta\right) \phi_{\varepsilon,N}\left(q, x_2; \theta\right) \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}}{\int \phi_{\varepsilon,N}\left(q, x_1; \theta\right) \phi_{\varepsilon,N}\left(q, x_2; \theta\right) \phi_{\eta,N}(q, z; \theta)\, d\mu_{X_{N,1}} d\mu_{X_{N,2}} d\mu_{Z_N}} dq\right]
\end{aligned}
$$

where

$$
\begin{aligned}
\phi_\eta(q, z; \theta) &= f_\eta\left(F_{u;\theta,\mu_Z}^{-1}(q) - g(z; \theta)\right) & \phi_\varepsilon(q, x; \theta) &= f_\varepsilon\left(F_{v;\theta,\mu_X}^{-1}(q) - h(x; \theta)\right)\\
\phi_{\eta,N}(q, z; \theta) &= f_\eta\left(F_{N,U;\theta,\mu_{Z_N}}^{-1}(q) - g(z; \theta)\right) & \phi_{\varepsilon,N}(q, x; \theta) &= f_\varepsilon\left(F_{N,V;\theta,\mu_{X_N}}^{-1}(q) - h(x; \theta)\right).
\end{aligned}
$$

in terms of G_X and G_Z. The detailed calculations are presented in Appendix D.2. Here, we illustrate the basic ideas of the argument and the components of the derivative by computing the limit of the following simplified expression:

$$
\frac{1}{h_N}\left[\int_\delta^{1-\delta} \frac{\int \Psi(x) \phi_\varepsilon(q, x; \theta)\, d\mu_X}{\int \phi_\varepsilon(q, x; \theta)\, d\mu_X} dq - \int_\delta^{1-\delta} \frac{\int \Psi(x) \phi_{\varepsilon,N}(q, x; \theta)\, d\mu_{X_N}}{\int \phi_{\varepsilon,N}(q, x; \theta)\, d\mu_{X_N}} dq\right].
$$

We first rewrite the difference

$$
\begin{aligned}
& \int_\delta^{1-\delta} \frac{\int \Psi(x) \phi_\varepsilon(q, x; \theta)\, d\mu_X}{\int \phi_\varepsilon(q, x; \theta)\, d\mu_X} dq - \int_\delta^{1-\delta} \frac{\int \Psi(x) \phi_{\varepsilon,N}(q, x; \theta)\, d\mu_{X_N}}{\int \phi_{\varepsilon,N}(q, x; \theta)\, d\mu_{X_N}} dq\\
= \quad & \int_\delta^{1-\delta} \frac{\int \Psi(x) \phi_\varepsilon(q, x; \theta)\left(d\mu_X - d\mu_{X_N}\right)}{\int \phi_\varepsilon(q, x; \theta)\, d\mu_X} dq + \int_\delta^{1-\delta} \frac{\int \Psi(x) \phi_\varepsilon(q, x; \theta)\, d\mu_{X_N}}{\int \phi_\varepsilon(q, x; \theta)\, d\mu_X} dq\\
& - \int_\delta^{1-\delta} \frac{\int \Psi(x) \phi_{\varepsilon,N}(q, x; \theta)\, d\mu_{X_N}}{\int \phi_\varepsilon(q, x; \theta)\, d\mu_X} \frac{\int \phi_\varepsilon(q, x; \theta)\, d\mu_X}{\int \phi_{\varepsilon,N}(q, x; \theta)\, d\mu_{X_N}} dq\\
= \quad & \int_\delta^{1-\delta} \frac{\int \Psi(x) \phi_\varepsilon(q, x; \theta)\left(d\mu_X - d\mu_{X_N}\right)}{\int \phi_\varepsilon(q, x; \theta)\, d\mu_X} dq + \int_\delta^{1-\delta} \frac{\int \Psi(x)\left(\phi_\varepsilon(q, x; \theta) - \phi_{\varepsilon,N}(q, x; \theta)\right) d\mu_{X_N}}{\int \phi_\varepsilon(q, x; \theta)\, d\mu_X} dq\\
& + \int_\delta^{1-\delta} \frac{\int \Psi(x) \phi_{\varepsilon,N}(q, x; \theta)\, d\mu_{X_N}}{\int \phi_\varepsilon(q, x; \theta)\, d\mu_X} \times \left(1 - \frac{\int \phi_\varepsilon(q, x; \theta)\, d\mu_X}{\int \phi_{\varepsilon,N}(q, x; \theta)\, d\mu_{X_N}}\right) dq\\
= \quad & \int_\delta^{1-\delta} T_1(q) + T_2(q) + T_3(q)\, dq
\end{aligned}
$$

To obtain the limit of $\frac{1}{h_N}\int_\delta^{1-\delta} T_1(q)\, dq$, note that $\frac{1}{h_N}\left(\mu_{X_N} - \mu_X\right)$ converges uniformly to $G_X \in L_\infty^{\Gamma_X}$. Therefore,

$$
\frac{1}{h_N}\int_\delta^{1-\delta} T_1(q)\, dq = \frac{1}{h_N}\int_\delta^{1-\delta} \frac{\int \Psi(x) \phi_\varepsilon(q, x; \theta)\left(d\mu_X - d\mu_{X_N}\right)}{\int \phi_\varepsilon(q, x; \theta)\, d\mu_X} dq \to \int_\delta^{1-\delta} \frac{\int \Psi(x) \phi_\varepsilon(q, x; \theta)\, dG_X}{\int \phi_\varepsilon(q, x; \theta)\, d\mu_X} dq.
$$

To obtain the limit of $\frac{1}{h_N}T_2(q)$, note that

$$\begin{aligned}
\frac{1}{h_N}T_2(q) &= \frac{\int \Psi(x)\frac{1}{h_N}\left(\phi_\varepsilon(q,x;\theta)-\phi_{\varepsilon,N}(q,x;\theta)\right)d\mu_{X_N}}{\int \phi_\varepsilon(q,x;\theta)\,d\mu_X} \\
&= \frac{1}{h_N}\left(F_{V;\theta}^{-1}(q)-F_{N,V;\theta}^{-1}(q)\right)\frac{\int \Psi(x)\,f_\varepsilon'\left(F_{V;\theta}^{-1}(q)-h(x;\theta)\right)d\mu_{X_N}}{\int \phi_\varepsilon(q,x;\theta)\,d\mu_X}+o(1) \\
&= \frac{1}{h_N}\left(F_{V;\theta}^{-1}(q)-F_{N,V;\theta}^{-1}(q)\right)\frac{\int \Psi(x)\,f_\varepsilon'\left(F_{V;\theta}^{-1}(q)-h(x;\theta)\right)d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\,d\mu_X}+o(1)
\end{aligned}$$

where the second equality follows from a Taylor expansion and the dominated convergence theorem (since f_ε' is bounded), and the last equality follows from the fact that $d\mu_{X_N}-d\mu_X\to 0$ and uniform bounds over $q\in(\delta,1-\delta)$ on the remaining terms. We then show that

$$\frac{1}{h_N}\left(F_{V;\theta}^{-1}(q)-F_{N,V;\theta}^{-1}(q)\right)\to\frac{1}{f_{V;\theta}\left(F_{V;\theta}^{-1}(q)\right)}\int G_X\left(1\left\{h(x;\theta)+\varepsilon\le F_{V;\theta}^{-1}(q)\right\}\right)dF_\varepsilon=G_V^q(\theta)$$

uniformly in $q\in(\delta,1-\delta)$ to obtain the limit

$$\frac{1}{h_N}\int_\delta^{1-\delta}T_2(q)\,dq\to\int_\delta^{1-\delta}G_V^q(\theta)\frac{\int \Psi(x)\,f_\varepsilon'\left(F_{V;\theta}^{-1}(q)-h(x;\theta)\right)d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\,d\mu_X}dq.$$

Finally, we rewrite

$$\begin{aligned}
T_3(q) &= \frac{\int \Psi(x)\,\phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}{\int \phi_\varepsilon(q,x;\theta)\,d\mu_X}\times\left(1-\frac{\int \phi_\varepsilon(q,x;\theta)\,d\mu_X}{\int \phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}\right) \\
&= \frac{\int \Psi(x)\,\phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}{\int \phi_{\varepsilon,N}(q,x;\theta)\,d\mu_X}\times\left(\frac{\int \phi_{\varepsilon,N}(q,x;\theta)\,d\mu_X-\int \phi_\varepsilon(q,x;\theta)\,d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\,d\mu_{X_N}}\right) \\
&= \frac{\int \Psi(x)\,\phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}{\int \phi_{\varepsilon,N}(q,x;\theta)\,d\mu_X}\times\left(-\tilde{T}_1(q)-\tilde{T}_2(q)\right) \\
&= \frac{\int \Psi(x)\,\phi_\varepsilon(q,x;\theta)\,d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\,d\mu_X}\times\left(-\tilde{T}_1(q)-\tilde{T}_2(q)\right) \\
&\quad+\left(\frac{\int \Psi(x)\,\phi_{\varepsilon,N}(q,x;\theta)\,d\mu_{X_N}}{\int \phi_{\varepsilon,N}(q,x;\theta)\,d\mu_X}-\frac{\int \Psi(x)\,\phi_\varepsilon(q,x;\theta)\,d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\,d\mu_X}\right)\times\left(-\tilde{T}_1(q)-\tilde{T}_2(q)\right)
\end{aligned}$$

where $\tilde{T}_1(q)=T_1(q)$ and $\tilde{T}_2(q)=T_2(q)$ evaluated at $\Psi(x)=1$. Since $\frac{1}{h_N}\left(-\tilde{T}_1(q)-\tilde{T}_2(q)\right)$ is finite, the second term is negligible. Hence,

$$\frac{1}{h_N}T_3(q)\to-\frac{\int \Psi(x)\,\phi_\varepsilon(q,x;\theta)\,d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\,d\mu_X}\left(\frac{\int \phi_\varepsilon(q,x;\theta)\,dG_X}{\int \phi_\varepsilon(q,x;\theta)\,d\mu_X}+G_V^q(\theta)\frac{\int f_\varepsilon'\left(F_{V;\theta}^{-1}(q)-h(x;\theta)\right)d\mu_X}{\int \phi_\varepsilon(q,x;\theta)\,d\mu_X}\right).$$

The limit of $\frac{1}{h_N}\int_\delta^{1-\delta}T_1(q)+T_2(q)+T_3(q)\,dq$ given by the expressions above yields the Hadamard derivative of interest. Online Appendix D.2 uses a dominated convergence argument to ensure that $T_1(q)+T_2(q)+T_3(q)$ converges uniformly in q.