Unobserved Heterogeneity in Matching Games with an Application to Venture Capital

Jeremy T. Fox
Rice University and NBER

David H. Hsu
University of Pennsylvania

Chenyu Yang
University of Michigan[*]

November 2015

Abstract

Agents in two-sided matching games vary in characteristics that are unobservable in typical data on matching markets. We investigate the identification of the distribution of unobserved characteristics using data on who matches with whom. In full generality, we consider many-to-many matching and matching with trades. The distribution of match-specific unobservables cannot be fully recovered without information on unmatched agents, but the distribution of a combination of unobservables, which we call unobserved complementarities, can be identified. Using data on unmatched agents restores identification. We estimate the contribution of observables and unobservable complementarities to match production in venture capital investments in biotechnology and medical firms.

1

1 Introduction

Matching games model the sorting of agents to each other. Men sort to women in marriage based on characteristics such as income, schooling, personality and physical appearance, with more desirable men typically matching to more desirable women. Upstream firms sort to downstream firms based on the product qualities and capacities of each of the firms. This paper is partially motivated by such applications in industrial organization and entrepreneurial finance, where downstream firms pay upstream firms money, and thus it is reasonable to work with transferable utility matching games (Koopmans and Beckmann, 1957; Gale, 1960; Shapley and Shubik, 1972; Becker, 1973). In particular, we explore an empirical application in corporate finance and management, where the upstream firms are venture capitalists and the downstream firms are entrepreneurial biotech and medical firms.

There has been recent interest in the structural estimation of (both transferable utility and non-transferable utility) matching games.[1] The papers we cite are unified in estimating some aspect of the preferences of agents in a matching game from data on who matches with whom as well as the observed characteristics of agents or of matches. The sorting patterns in the data combined with assumptions about equilibrium inform the researcher about the structural primitives in the market, namely some function that transforms an agent's own characteristics and its potential partner's characteristics into some notion of utility or output. These papers are related to, but are not special cases of, papers estimating discrete, non-cooperative (Nash) games, like the entry literature in industrial organization and the discrete outcomes peer effects literature.[2] Matching games typically use the cooperative solution concept of pairwise stability.[3]

The empirical literature cited previously structurally estimates how various structural or equilibrium objects, such as payoffs or preferences, are functions of the characteristics of agents *observed in the data*. For example, Choo and Siow (2006) study the marriage market in the United States and estimate how the equilibrium payoffs of men for women vary by the ages of the man and the woman. Sørensen (2007) studies the matching of venture capitalists to entrepreneurs as a function of observed venture capitalist experience. Fox (2010a) studies matching between automotive assemblers (downstream firms) and car parts suppliers (upstream firms) and asks how observed specialization measures in the portfolios of car parts sourced or supplied contribute to agent profit functions.

---

[1]See, among others: Dagsvik (2000); Boyd et al. (2013); Choo and Siow (2006); Sørensen (2007); Fox (2010a); Gordon and Knight (2009); Chen (2013); Ho (2009); Park (2013); Yang et al. (2009); Logan et al. (2008); Levine (2009); Baccara et al. (2012); Siow (2009); Galichon and Salanie (2012); Chiappori et al. (2015); Crawford and Yurokoglu (2012); Weese (forthcoming); Christakis et al. (2010); Echenique et al. (2013); Uetake and Watanabe (2012); Menzel (2015); Agarwal (2015); Agarwal and Diamond (2013); Akkus et al. (forthcoming).

[2]See, among others: Berry (1992); Bresnahan and Reiss (1991); Mazzeo (2002); Tamer (2003); Bajari et al. (2010); Seim (2006); Brock and Durlauf (2007); de Paula and Tang (2012).

[3]Transferable utility matching games (particularly those with "contracts" or "trades" that specify endogenous product attributes) are equivalent to models of hedonic equilibrium (Brown and Rosen, 1982; Ekeland et al., 2004; Heckman et al., 2010; Chiappori et al., 2010). Unlike the empirical literature on hedonic equilibrium, the estimation approaches in most matching papers do not rely on data on equilibrium prices or transfers. Compared to the current work, the hedonic papers do not allow for unobserved characteristics.

The above papers all use data on a relatively limited set of agent characteristics. In Choo and Siow, personality and physical attractiveness are not measured, even though those characteristics are likely important in determining the equilibrium pattern of marriages. Similarly, in Fox each firm's product quality is not directly measured and is only indirectly inferred. In Sørensen, the unobserved ability of each venture capitalist and the business prospects of each entrepreneurial firm are not measured. If matching based on observed characteristics is found to be important, it is a reasonable conjecture that matching based on *unobserved characteristics* is also important. Our empirical work on biotech and medical venture capital investments complements the earlier work by Sørensen on venture capital; we estimate distributions of functions of match-specific unobservables. Ackerberg and Botticini (2002) provide empirical evidence that farmers and landlords sort on unobservables such as risk aversion and monitoring ability, without formally estimating a matching game or the distribution of these unobservables.

Our discussion of the empirical applications cited above suggests that unobserved characteristics are potentially important. As the consistency of estimation procedures for matching games depends on assumptions on the unobservables, empirical conclusions might be more robust if the estimated matching games allow richly specified distributions of unobserved agent heterogeneity. This paper investigates what data on the sorting patterns between agents can tell us about the distributions of unobserved agent characteristics relevant for sorting. In particular, we study the nonparametric identification of distributions of unobserved agent heterogeneity in two-sided matching games. With the distribution of unobservables, the researcher can explain sorting and construct counterfactual predictions about market assignments. This paper allows for this empirically relevant heterogeneity in partner preferences using data on only observed matches (who matches with whom), not data from, say, an online dating site on rejected profiles (Hitsch et al., 2010) or on equilibrium transfers, such as wages in a labor market (Eeckhout and Kircher, 2011). Transfers are often confidential data in firm contracts (Fox, 2010a) and are rarely observed in marriage data (Becker, 1973).

In the following specific sense, this paper on identification is ahead of the empirical matching literature because, when this paper was first written, no empirical papers had parametrically estimated distributions of unobserved agent characteristics, or of match characteristics without assuming independence across matches, in matching games.[4] Thus, this paper seeks to introduce a new topic for economic investigation, rather than to simply loosen parametric restrictions in an existing empirical literature. This paper contributes to the literature on the nonparametric (allowing infinite dimensional objects) identification of transferable utility matching games (Fox, 2010b; Graham, 2011). Our paper is distinguished because of its focus on identifying distributions of unobservables, rather than mostly deterministic functions of observables. Our focus in identification on using data on many markets with finite numbers of agents in each (transferable utility) market follows Fox (2010b).

---

[4] Agarwal (2015) and Agarwal and Diamond (2013) discuss the identification and estimation of distributions of agent characteristics in non-transferable utility matching games with restrictions on preferences in order to ensure a unique stable match.

In addition to our study of identifying distributions of unobservables, there are many modeling differences between our paper and the literature on transferable utility matching games following the approach of Choo and Siow (2006), including Galichon and Salanie (2012), Chiappori, Salanié and Weiss (2015), Graham (2011) and Fox (2010a). We use data on many markets with finite numbers of players and different realizations of observables and unobservables in each market; the Choo and Siow approach has been applied to one large market with a continuum of agents. We require at least one continuous, observable characteristic per match or per agent; the Choo and Siow literature allows only a finite number of observable characteristic values. The production functions corresponding to these finite unobservables are usually recoverable without further functional form assumptions; we require a particular match or agent characteristic to enter match production additively separably. Unobservables in the Choo and Siow literature are typically i.i.d. shocks for the finite observable types rather than unobserved agent characteristics or unobserved preferences on observed, ordered characteristics, such as random coefficients. This is one interpretation of the "separability" assumption of Chiappori et al..

We first consider a baseline model, which is stripped down to focus on the key problem of identifying distributions of heterogeneity from sorting data. In our baseline transferable utility matching game, the primitive that governs sorting is the matrix that collects the production values for each potential match in a matching market. The production level of each match is additively separable in observable and unobservable terms. The observable term is a match-specific characteristic. The unknown primitive is therefore the distribution (representing randomness across markets) of the matrix that collects the unobservable terms in the production of each match in a market. We call this distribution the distribution of match-specific unobservables. Match-specific unobservables nest many special cases, such as agent-specific unobservables.

We first show that that the distribution of match-specific unobservables is not identified in a one-to-one matching game with data on who matches with whom but without data on unmatched or single agents. We provide two main theoretical results and many extensions. Our first main theoretical result states that the distribution of a change of variables of the unobservables, the distribution of what we call **unobserved complementarities**, is identified. We precisely define unobserved complementarities below. Our identification proof works by tracing the joint (across possible matches in a market) cumulative distribution function of these unobserved complementarities using the match-specific observables. We also show that knowledge of the distribution of unobserved complementarities is sufficient for computing assignment probabilities. Our second main theoretical result says that the distribution of the primitively specified, match-specific unobservables is actually identified when unmatched agents are observed in the data.

Our main theoretical results can be intuitively understood by reference to a classic result in Becker (1973). He studies sorting in two-sided, transferable utility matching games where agents have scalar characteristics (types). He shows that high-type agents match to high-type agents if the types of agents are complements in the production of matches. Many production functions for match output

4

exhibit complementarities. Say in Becker's model male and female types are x_m and x_w, respectively. A production function with horizontal preferences, such as $-(x_m - x_w)^2$, and one with vertical preferences, such as $2x_m x_w$, can both have the same cross-partial derivative, here 2. Becker's result that complementarities alone drive sorting means that data on sorting cannot tell these two production functions apart. In our more general class of matching games, we cannot identify the distribution of match-specific unobservables. However, we can identify the distribution of our notion of unobserved complementarities. These results are analogous to Becker's results, for a more general class of matching games.

Our second main theoretical result uses data on unmatched agents. In a matching game, agents can unilaterally decide to be single or not. If all other agents are single and hence available to match, the fact that one particular agent is single can only be explained by the production of all matches involving that agent being less than the production from being single. This type of direct comparison between the production of being single and the production of being matched is analogous to the way identification proceeds in discrete Nash games, where the payoff of a player's observed (in the data) strategy must be higher than strategies not chosen, given the strategies of rivals. Thus, the availability of data on unmatched agents introduces an element of individual rationality that maps directly into the data and is therefore useful for identification of the primitive distribution of match-specific characteristics.

Many empirical researchers might be tempted to specify a parametric distribution of match-specific unobservables. Our theoretical results together suggest that estimating a matching model with a parametric distribution of match-specific unobservables will not necessarily lead to credible estimates without using data on unmatched agents, as a more general nonparametrically specified distribution is not identified. Also, we present an example of a multivariate normal distribution of match-specific characteristics whose parameters are not parametrically identified. One could instead impose a parametric distribution for unobserved complementarities, as we do in our empirical work on biotech and medical venture capital.

We examine several extensions to the baseline model that add more empirical realism. Our baseline model imposes additive separability between unobservables and observables in the production of a match. We examine an extension where additional observed characteristics enter match production and these characteristics may, for example, have random coefficients on them, reflecting the *random preferences* of agents for partner characteristics. For example, observationally identical men are often observed to marry observationally distinct women. One important hypothesis is that these men have heterogeneous preferences for the observable characteristics of women. In a model with random preferences, we identify the distribution of unobserved complementarities conditional on the characteristics of agents and matches other than the match-specific characteristics used in the baseline model. Identifying a distribution of unobservables conditional on observables follows identification work using special regressors in the multinomial choice literature (Lewbel, 2000; Matzkin, 2007; Berry

and Haile, 2010).

In another extension, we identify fixed-across-markets but heterogeneous-within-a-market coefficients on the match-specific characteristics used in the baseline model. This relaxes the assumption that the match-specific characteristics enter the production of each match in the same manner. Another extension considers models where key observables vary at the agent and not the match level and enter match production multiplicatively. We can identify the distribution of unobserved complementarities if agent unobservables enter match production symmetrically to agent observables.

Our results on one-to-one, two-sided matching games extend naturally to many-to-many matching (Crawford and Knoer, 1981; Sotomayor, 1992, 1999). Our application to venture capital uses the many-to-one special case. We discuss another extension to the model of matching with trades in Hatfield et al. (2013), who significantly generalize Kelso and Crawford (1982). In matching with the trades, the same agent can make so-called trades both as a buyer and a seller and can have complicated preferences over the set of trades. An individual trade generalizes a match in that a trade can list other specifications, such as the number of startup board seats given to a venture capitalist. The matching with trades model has many special cases and is the most general model we provide identification results for.

We use our theoretical results to motivate an empirical investigation into matching between biotech and medical entrepreneurs and venture capitalists. Venture capital is a key way entrepreneurial innovation is funded. We use detailed data on the observed matches between entrepreneurial startups and venture capitalist firms over a ten year period. We collect information on the geographic locations of both startups and venture capitalists, on the patent stocks of startups, and on the past experience of venture capitalists in various biotech and medical sectors. Despite these observed characteristics being as detailed as any data set on venture capital that could realistically be collected by academic researchers, we find that the distribution of unobserved complementarities suggests that unobserved characteristics play a large role in match production.

2   Baseline Identification Results

We mainly analyze a two-sided, one-to-one matching game with transferable utility (Koopmans and Beckmann, 1957; Gale, 1960; Shapley and Shubik, 1972; Becker, 1973; Roth and Sotomayor, 1990, Chapter 8). This section imposes that all agents must be matched in order to focus purely on the identification coming from agent sorting and not from the individual rationality decision to be single. We also use a simple space of explanatory variables. We change these assumptions in later sections.

6

2.1 Baseline Model

We use the terms "agents" and "firms" interchangeably. In a one-to-one matching game, an upstream firm u matches with a downstream firm d. In biotech and medical venture capital, upstream firms are venture capitalists and downstream firms are entrepreneurs. Upstream firm u and downstream firm d can form a match $\langle u, d \rangle$. The monetary transfer from d to u is denoted as $t_{u,d}$; we will not require data on the transfers. The **production** or **total profit** from a match $\langle u, d \rangle$ is

$$z_{u,d} + e_{u,d}, \tag{1}$$

where $z_{u,d}$ is a scalar match-specific characteristic observed in the data and $e_{u,d}$ is a scalar match-specific characteristic unobserved in the data, but observable to all firms in the matching game. In our empirical work on venture capital, one match-specific characteristic $z_{u,d}$ is the distance between the headquarters of firms u and d.[5] The match-specific, unobserved characteristic $e_{u,d}$ generalizes special cases such as $e_{u,d} = e_u \cdot e_d$, where e_u and e_d are unobserved upstream and downstream firm characteristics, respectively. We allow a match-specific coefficient on each $z_{u,d}$ and, separately, use only agent-specific explanatory variables below.

We can more primitively model production for a match $\langle u, d \rangle$ as the sum of the profit of u and the profit of d, where the possibly negative transfer $t_{u,d}$ between d and u enters additively separably into both individual profits and therefore cancels in their sum.[6] However, only production levels matter for the matches that form, and we will not attempt to identify upstream firm profits separately from downstream firm profits, except in the extension to matching with trades in Section 7.2.

There are N firms on each side of the market. N can also represent the set $\{1, \ldots, N\}$. In this section, there can be no unmatched firms. The matrix

$$\begin{pmatrix} z_{1,1} + e_{1,1} & \cdots & z_{1,N} + e_{1,N} \\ \vdots & \ddots & \vdots \\ z_{N,1} + e_{N,1} & \cdots & z_{N,N} + e_{N,N} \end{pmatrix}$$

describes the production of all matches in a market, where the rows are upstream firms and the

---

[5]Distance $z_{u,d}$ is always positive and likely enters match production with a negative sign; we can always construct a new regressor $\tilde{z}_{u,d} = - \left(z_{u,d} - E\left[z_{u,d} \right] \right)$ that enters with a positive sign and has mean zero.

[6]If the profit of u at some market outcome is $\pi_{u,d}^u + t_{u,d}$ and the profit of d is $\pi_{u,d}^d - t_{u,d}$, then the production of the match $\langle u, d \rangle$ is equal to $\pi_{u,d}^u + \pi_{u,d}^d = z_{u,d} + e_{u,d}$. We will not attempt to learn the distributions of the unobservable portions of $\pi_{u,d}^u$ and $\pi_{u,d}^d$ separately, except in the extension to matching with trades in Section 7.2 (Fox, 2010b).

columns are downstream firms. Let

$$E = \begin{pmatrix} e_{1,1} & \cdots & e_{1,N} \\ \vdots & \ddots & \vdots \\ e_{N,1} & \cdots & e_{N,N} \end{pmatrix}, Z = \begin{pmatrix} z_{1,1} & \cdots & z_{1,N} \\ \vdots & \ddots & \vdots \\ z_{N,1} & \cdots & z_{N,N} \end{pmatrix}$$

be the matrices of unobservables and observables, respectively, in a market.[7]

A feasible one-to-one **assignment** A is a set of matches $A = \{\langle u_1, d_1 \rangle, \ldots, \langle u_N, d_N \rangle\}$, where for this section each firm is matched exactly once. There are $N!$ feasible assignments. An **outcome** is a list of matches and transfers between matched agents:

$$\{\langle u_1, d_1, t_{u_1,d_1} \rangle, \ldots, \langle u_N, d_N, t_{u_N,d_N} \rangle\}.$$

An outcome is **pairwise stable** if it is robust to deviations by pairs of two firms, as defined in references such as Roth and Sotomayor (1990, Chapter 8).[8] An assignment A is called **pairwise stable** if there exists an underlying outcome (including transfers) that is pairwise stable.

The literature cited previously proves that the existence of a pairwise stable assignment is guaranteed and that an assignment A is pairwise stable if and only if it maximizes the **sum of production**

$$s(A; E, Z) = \sum_{\langle u,d \rangle \in A} (z_{u,d} + e_{u,d}).$$

If $z_{u,d}$ or $e_{u,d}$ have continuous support, $s(A; E, Z)$ has a unique maximizer with probability 1 and therefore the pairwise stable assignment is unique with probability 1. The **sum of the unobserved production** of assignment A relative to the particular assignment $A_1 = \{\langle 1, 1 \rangle, \ldots, \langle N, N \rangle\}$ is

$$\tilde{s}(A; E) = \sum_{\langle u,d \rangle \in A} e_{u,d} - \sum_{\langle u,d \rangle \in A_1} e_{u,d}. \tag{2}$$

A market is defined to be the pair (E, Z); agents in a market can match and agents in different markets cannot. A researcher observes the assignment A and the match-specific characteristics Z for many markets. In other words, in each matching market the researcher observes who matches with whom A and the characteristics Z of the realized and potential matches. This allows the identification of $\Pr(A \mid Z)$, the probability of assignment A being the pairwise stable assignment given the market-level match characteristics Z. Researchers do not observe transfers, which are often part of confidential contracts.

---

[7]Because the scalar $z_{u,d}$ is an element of the matrix Z, we do not use upper and lower case letters (or other notation) to distinguish random variables and their realizations. Whether we refer to a random variable or its realization should be clear from context.

[8]We omit standard definitions here that can be easily found in the literature.

Z is independent of the unobservable matrix E. We assume that Z has full and product support, meaning that any $Z \in \mathbb{R}^{N^2}$ is observed.[9] Each match characteristic $z_{u,d}$ enters production additively, the sign and coefficient on each $z_{u,d}$ in production is common across matches (normalized to be 1), each $z_{u,d}$ has large support, and Z is independent of E.[10] Similar large support explanatory variables have been used to prove point identification in the binary and multinomial choice literature (Manski, 1988; Ichimura and Thompson, 1998; Lewbel, 1998, 2000; Matzkin, 2007; Gautier and Kitamura, 2013; Berry and Haile, 2010; Fox and Gandhi, forthcoming). In this literature, failure to have large support often results in set rather than point identification of the distribution of heterogeneity.[11] We discuss the failure of the support condition in our matching context below. In this paper, we use large support match characteristics in part to focus on reasons specific to matching games for the failure of point identification.[12]

2.2 Data Generating Process

The unknown primitive whose identification we first explore is the CDF $G(E)$, which reflects how the match unobservables vary across matching markets. We do not restrict the support of E and we do not assume independence across the $e_{u,d}$'s within matching markets. Hence, we allow for many special cases, such as the case $e_{u,d} = e_u \cdot e_d$ mentioned earlier.

The probability of assignment A occurring given the match characteristics Z is

$$\Pr(A \mid Z; G) = \int_E 1\left[A \text{ pairwise stable assignment} \mid Z, E\right] \tilde{d}G(E), \tag{3}$$

where $1\left[A \text{ pairwise stable assignment} \mid Z, E\right]$ is equal to 1 when A is a pairwise stable assignment for the market (E, Z). The symbol $\tilde{d}$ in $\tilde{d}G$ stands in for the differential symbol "d" from calculus, to

---

[9]The example of the match-characteristic distance may not vary independently over all of $\mathbb{R}^{N^2}$ because distance is computed using the agent-specific characteristics latitude and longitude. In our empirical work, we use a second match-specific characteristic that conceptually can vary independently in N^2 dimensions: the past experience of a venture capitalist with investments in the four-digit sector of a startup. Experience in a sector varies independently in N^2 dimensions if all startups are in different four-digit sectors.

[10]We could in principle address the statistical dependence of E and Z with instrumental variables. We do not explore this. We should mention that the $e_{u,d}$ and $z_{u,d}$ for the *realized* matches in the pairwise stable assignment A will likely be statistically dependent because of the conditioning on the dependent variable A, part of the outcome to the game.

[11]Consider a binary choice model of buying a can of soda (or not) where the large support regressor is the (negative) price of the soda, which varies across the dataset. If we assume that consumers' willingnesses to pay for the can of soda are bounded by \$0 and \$10, we can point identify the distribution of the willingness to pay for soda if observable prices range between \$0 and \$10. If prices range only between \$0 and \$5, we can identify the fraction of consumers with values above \$5 by seeing the fraction who purchase at \$5. We cannot identify the fraction with values above \$6, or any value greater than \$5. If we do not restrict the support of the willingness to pay, we need prices to vary across all of $\mathbb{R}$ (including negative prices if consumers may have negative willingnesses to pay) for point identification of the distribution of the willingness to pay for soda.

[12]Our use of large support and the use of large support in most of the literature on binary and multinomial choice does not constitute identification at infinity as used in certain proofs to study Nash games by, for example, Tamer (2003). Identification at infinity in a Nash game uses only extreme values of regressors for all but one player to, in effect, turn a multi-player game into a single-player decision problem. We use large explanatory variable values only to identify the tails of distributions of heterogeneity.

distinguish this use from our notation d for a downstream firm.

The distribution G is said to be **identified** whenever, for $G^1 \neq G^2$, $\Pr\left(A \mid Z; G^1\right) \neq \Pr\left(A \mid Z; G^2\right)$ for some pair (A, Z). G^1 and G^2 give a different probability for at least one assignment A given Z. If G has continuous and full support so that all probabilities $\Pr\left(A \mid Z; G\right)$ are nonzero (for every (A, Z), $s\left(A; E, Z\right)$ will be maximized by a range of E) and continuous in the elements of Z, the existence of one such pair (A, Z) implies that a set of Z with positive measure satisfies $\Pr\left(A \mid Z; G^1\right) \neq \Pr\left(A \mid Z; G^2\right)$.

All of our positive identification results will be constructive, in that we can trace a distribution such as $G\left(E\right)$ using variation in an object such as Z. Also, our identification arguments can be used to prove the consistency of a nonparametric mixtures estimator for a distribution G of heterogeneous unobservables E, as Fox, Kim and Yang (2015) show for a particular, computationally simple mixtures estimator.[13] Other mixtures estimators can be used, including maximum simulated likelihood, the EM algorithm, NPMLE, and MCMC.[14] In the empirical work on venture capital, we use the simulated method of moments in a parametric model, because of the large numbers of firms in our matching markets (McFadden, 1989; Pakes and Pollard, 1989).

2.3 Non-Identification of the Distribution of Match-Specific Characteristics

As maximizing $s\left(A; E, Z\right)$ determines the assignment seen in the data, the ordering of $s\left(A; E, Z\right)$ across assignments A as a function of E and Z is a key input to identification. We can add a constant to the production of all matches involving the same upstream firm and the ordering of the production $s\left(A; E, Z\right)$ of all assignments will remain the same. This non-identification result is unsurprising: the differential production of matches and hence assignments governs the identity of the pairwise stable assignment in any market.

We will show another non-identification result. Consider the two realizations of matrices of unobservables

$$E_1 = \begin{pmatrix} e_{1,1} & e_{1,2} & \cdots & e_{1,N} \\ e_{2,1} & e_{2,2} & \cdots & e_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ e_{N,1} & e_{N,2} & \cdots & e_{N,N} \end{pmatrix}, \; E_2 = \begin{pmatrix} e_{1,1} & e_{1,2} + 1 & \cdots & e_{1,N} \\ e_{2,1} - 1 & e_{2,2} + 1 - 1 & \cdots & e_{2,N} - 1 \\ \vdots & \vdots & \ddots & \vdots \\ e_{N,1} & e_{N,2} + 1 & \cdots & e_{N,N} \end{pmatrix}.$$

---

[13]The proof of consistency in Fox et al. (2015) for one estimator requires the heterogeneous unobservable (such as E) to have compact support, which is not required here for identification. A second estimator in Fox et al. allows the support of E to be $\mathbb{R}^{\dim(E)}$.

[14]For large markets, these estimators all have computational problems arising from the combinatorics underlying the set of matching game assignments. Fox (2010a) uses a maximum score estimator to avoid these computational problems, but does not estimate a distribution of unobservables. Our identification arguments do not address computational issues. Likewise, random variables such as E are of large dimension and nonparametrically estimating a CDF such as $G\left(E\right)$ will result in a data curse of dimensionality.

It is easy to verify that $s(A; E_1, Z) = s(A; E_2, Z)$ for all A, Z, which means that the pairwise stable assignment A is the same for E_1 and E_2, for any Z. Therefore it is not possible to separately identify the relative frequencies of E_1 and E_2 in the data generating process; the support of the random matrix E is too flexible.

We summarize the two counterexamples in the following non-identification proposition.

Proposition 1. *The distribution $G(E)$ of market-level unobserved match characteristics is **not** identified in a matching game where all agents must be matched.*

Consider a simple case focusing on two upstream firms and two downstream firms. If we see the matches $\langle u_1, d_1 \rangle$ and $\langle u_2, d_2 \rangle$ in the data, we cannot know whether this assignment forms because $\langle u_1, d_1 \rangle$ has high production, $\langle u_2, d_2 \rangle$ has high production, $\langle u_1, d_2 \rangle$ has low production, or $\langle u_2, d_1 \rangle$ has low production. The non-identification result implies that parametric estimation of $G(E)$ under these assumptions may not be well founded, in that the generalization removing the parametric restrictions is not identified.

2.4 Unobserved Assignment Production

The pairwise stable assignment A maximizes the function $s(A; E, Z) = \sum_{\langle u,d \rangle \in A} (z_{u,d} + e_{u,d})$. This looks like a single agent, the social planner, maximizing a utility function. Rough intuition from the multinomial choice literature, cited earlier, suggests that the distribution $H\left(\tilde{S}\right)$ of

$$\tilde{S} = (\tilde{s}(A_2; E), \ldots, \tilde{s}(A_{N!}; E)) = \left(\sum_{\langle u,d \rangle \in A_2} e_{u,d} - \sum_{\langle u,d \rangle \in A_1} e_{u,d}, \ldots, \sum_{\langle u,d \rangle \in A_{N!}} e_{u,d} - \sum_{\langle u,d \rangle \in A_1} e_{u,d} \right)$$

might be identified, where the long vector $\tilde{S}$ collects the unobserved production of $N! - 1$ assignments relative to the reference assignment $A_1 = \{\langle 1, 1 \rangle, \ldots, \langle N, N \rangle\}$. Directly citing the multinomial choice literature requires a vector of $N! - 1$ assignment-specific observables with support $\mathbb{R}^{N!-1}$, where a hypothetical assignment-specific observable would enter only $s(A; E, Z)$ for a particular A. Assignment-specific observables do not exist in our matching game. However, the distribution of $H\left(\tilde{S}\right)$ is identified using only the variation in match-specific characteristics Z assumed earlier.

Lemma 1. *The distribution $H\left(\tilde{S}\right)$ of unobserved production for all assignments is identified.*

The proof, in the appendix, shows that large and product support on Z allows us to trace $H\left(\tilde{S}\right)$. The identification argument is therefore constructive. Failure of large and product support results in partial identification of $H\left(\tilde{S}\right)$.

11

Example 1. For a running example, consider the case $N = 3$. The matrix of match characteristics is

$$E = \begin{pmatrix} e_{1,1} & e_{1,2} & e_{1,3} \\ e_{2,1} & e_{2,2} & e_{2,3} \\ e_{3,1} & e_{3,2} & e_{3,3} \end{pmatrix}.$$

There are six possible assignments,

$$\begin{aligned}
A_1 &= \{\langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle\} \\
A_2 &= \{\langle 1,2\rangle, \langle 2,1\rangle, \langle 3,3\rangle\} \\
A_3 &= \{\langle 1,3\rangle, \langle 2,2\rangle, \langle 3,1\rangle\} \\
A_4 &= \{\langle 1,2\rangle, \langle 2,3\rangle, \langle 3,1\rangle\} \\
A_5 &= \{\langle 1,1\rangle, \langle 2,3\rangle, \langle 3,2\rangle\} \\
A_6 &= \{\langle 1,3\rangle, \langle 2,1\rangle, \langle 3,2\rangle\}
\end{aligned} \tag{4}$$

and

$$\tilde{S} = \begin{pmatrix} \tilde{s}(A_2; E) \\ \tilde{s}(A_3; E) \\ \tilde{s}(A_4; E) \\ \tilde{s}(A_5; E) \\ \tilde{s}(A_6; E) \end{pmatrix} = \begin{pmatrix} e_{1,2} + e_{2,1} + e_{3,3} - (e_{1,1} + e_{2,2} + e_{3,3}) \\ e_{1,3} + e_{2,2} + e_{3,1} - (e_{1,1} + e_{2,2} + e_{3,3}) \\ e_{1,2} + e_{2,3} + e_{3,1} - (e_{1,1} + e_{2,2} + e_{3,3}) \\ e_{1,1} + e_{2,3} + e_{3,2} - (e_{1,1} + e_{2,2} + e_{3,3}) \\ e_{1,3} + e_{2,1} + e_{3,2} - (e_{1,1} + e_{2,2} + e_{3,3}) \end{pmatrix}. \tag{5}$$

Lemma 1 states that the distribution $H\left(\tilde{S}\right)$ is identified using variation in

$$Z = \begin{pmatrix} z_{1,1} & z_{1,2} & z_{1,3} \\ z_{2,1} & z_{2,2} & z_{2,3} \\ z_{3,1} & z_{3,2} & z_{3,3} \end{pmatrix}.$$

$\triangle$

2.5 Unobserved Complementarities

The random vector $\tilde{S}$ has $N! - 1$ elements. Estimating a joint distribution of $N! - 1$ elements is not practical in typical datasets. We now introduce the concept of unobserved complementarities as an intuitive, lower-dimensional random variable whose distribution is point identified if and only if $H\left(\tilde{S}\right)$ is point identified.

As described in the introduction, Becker (1973) shows that complementarities govern sorting when there is one characteristic (schooling) per agent. Likewise, references such as Fox (2010b) and Graham (2011) prove that complementarities in *observed* agent or match characteristics are identified using data on matches. Likewise, while it is not possible to identify the distribution of unobserved match

characteristics, we will show that the distribution of *unobserved* complementarities can be identified.

Definition. The **unobserved complementarity** between matches $\langle u_1, d_1 \rangle$ and $\langle u_2, d_2 \rangle$ is

$$c_{u_1,d_1,u_2,d_2} = e_{u_1,d_1} + e_{u_2,d_2} - (e_{u_1,d_2} + e_{u_2,d_1}). \tag{6}$$

The unobserved complementarities capture the change in the unobserved production (unobserved profits) when two matched pairs $\langle u_1, d_1 \rangle$ and $\langle u_2, d_2 \rangle$ exchange partners and the matches $\langle u_1, d_2 \rangle$ and $\langle u_2, d_1 \rangle$ arise.

Fixing a realization of the unobserved match characteristics E, one can calculate the market-level array (of four dimensions) comprising all unobserved complementarities

$$C = (c_{u_1,d_1,u_2,d_2} \mid u_1, u_2, d_1, d_2 \in N). \tag{7}$$

We only consider values C formed from valid values of E.

There are N^4 values c_{u_1,d_1,u_2,d_2} in C given any realization E.[15] However, all unobserved complementarities can be formed from a smaller set of other unobserved complementarities by addition and subtraction. Let

$$b_{u,d} = c_{1,1,u,d} = e_{1,1} + e_{u,d} - (e_{1,d} + e_{u,1}) \tag{8}$$

be an unobserved complementarity fixing the identities of the upstream firm u_1 and the downstream firm d_1 to both be 1. Let the matrix B be

$$B = \begin{pmatrix} b_{2,2} & \cdots & b_{2,N} \\ \vdots & \ddots & \vdots \\ b_{N,2} & \cdots & b_{N,N} \end{pmatrix},$$

which contains all unique values of $b_{u,d}$ for a market. B is a matrix of $(N-1)^2$ elements. The following lemma shows we can restrict attention to B instead of C and hence focus on identifying the joint distribution $F(B)$ of the heterogeneous matrix B.

Lemma 2.

1. *Every element of C is a linear combination of elements of B. The specific linear combination does not depend on the realizations of C or B.*

2. *For any CDF $F(B)$, there exists $G(E)$ generating $F(B)$ by the appropriate change of variables in (8).*

---

[15]There are $\begin{pmatrix} N \\ 2 \end{pmatrix}^2 = \frac{N^4}{4} - \frac{N^3}{2} + \frac{N^2}{4}$ unobserved complementarities in C with a unique pair of a pair of upstream firms and a pair of downstream firms. This is on the order of N^4.

3. If E is a exchangeable random matrix in upstream agent indices and also exchangeable in downstream agent indices, then so is B.

By the first part of the lemma, we can focus on identifying the distribution of the $(N-1)^2$ elements in B instead of all N^4 elements in C. By the second statement in the lemma, we can restrict attention to identifying $F(B)$ without restrictions on the support of B or the dependence between the elements of B, as any $F(B)$ is compatible with some distribution $G(E)$ of the primitive matrix of match-specific unobservables E. Further, the third statement in the lemma shows that in the typical empirical context where the distribution of primitive unobservables is exchangeable in agent indices, the distribution of unobserved complementarities is also exchangeable in agent indices. The proof in the appendix has a formal definition of exchangeability in agent indices. We now present examples of some of the claims in the lemma.

Example. 1 ($N = 3$) There are $3! = 6$ assignments. There are 81 unobserved complementarities c_{u_1,d_1,u_2,d_2} in C.[16] There are 4 unobserved complementarities in B:

$$
B = \begin{pmatrix} b_{2,2} & b_{2,3} \\ b_{3,2} & b_{3,3} \end{pmatrix} =
$$
$$
\begin{pmatrix} e_{1,1} + e_{2,2} - (e_{1,2} + e_{2,1}) & e_{1,1} + e_{2,3} - (e_{1,3} + e_{2,1}) \\ e_{1,1} + e_{3,2} - (e_{1,2} + e_{3,1}) & e_{1,1} + e_{3,3} - (e_{1,3} + e_{3,1}) \end{pmatrix}. \tag{9}
$$

The first part of Lemma 2 claims that the 12 elements in C can be constructed from the 4 elements in B. For one example,

$$
c_{2,2,3,3} = e_{2,2} + e_{3,3} - (e_{2,3} + e_{3,2}) = b_{2,2} - b_{2,3} - b_{3,2} + b_{3,3}.
$$

$\triangle$

Example 2. Let the distribution $G(E)$ be exchangeable in agent indices for upstream and downstream firms separately. Also let $G(E)$ be multivariate normal with zero means. Under exchangeability, zero means and the multivariate normality of E, the variance matrix of the distribution G is

---

[16]For $N = 3$, the $\begin{pmatrix} N \\ 2 \end{pmatrix}^2$ from the previous footnote is 9.

parameterized by four unique parameters as

$$\begin{aligned}
\text{Cov}\left(e_{u_1,d_1}, e_{u_2,d_2}\right) &= \psi_1, \text{ if } u_1 \neq u_2, d_1 \neq d_2 \\
\text{Cov}\left(e_{u_1,d_1}, e_{u_2,d_1}\right) &= \psi_2, \text{ if } u_1 \neq u_2 \\
\text{Cov}\left(e_{u_1,d_1}, e_{u_1,d_2}\right) &= \psi_3, \text{ if } d_1 \neq d_2 \\
\text{Var}\left(e_{u_1,d_1}\right) &= \psi^2.
\end{aligned}$$

One can use the properties of linear changes of variables for multivariate normal distributions to algebraically derive the distribution $F(B)$ of unobserved complementarities. $F(B)$ is itself exchangeable in agent indices (as Lemma 2.3 states) and is multivariate normal with a variance matrix with diagonal and off-diagonal terms

$$\begin{aligned}
\text{Cov}\left(b_{u_1,d_1}, b_{u_2,d_2}\right) &= \frac{1}{4}\nu^2, \text{ if } u_1 \neq u_2, d_1 \neq d_2 \\
\text{Cov}\left(b_{u_1,d_1}, b_{u_2,d_1}\right) &= \frac{1}{2}\nu^2, \text{ if } u_1 \neq u_2 \\
\text{Cov}\left(b_{u_1,d_1}, b_{u_1,d_2}\right) &= \frac{1}{2}\nu^2, \text{ if } d_1 \neq d_2 \\
\text{Var}\left(b_{u_1,d_1}\right) &= \nu^2,
\end{aligned}$$

where the new parameter $\nu^2 = 4\left(\psi^2 + \psi_1 - \psi_2 - \psi_3\right)$. This example shows the reduction of information from considering unobserved complementarities instead of unobserved match characteristics. In this example, $G(E)$ is parameterized by four parameters while the induced $F(B)$ has only one unknown parameter. An estimator for the parameters ψ_1, ψ_2, ψ_3 and ψ^2 will be inconsistent as only a linear combination of those four parameters is identified. $\triangle$

2.6 Identification of Unobserved Complementarities

We have shown that $H\left(\tilde{S}\right)$ is identified, where recall $\tilde{S} = \left(\tilde{s}\left(A_2, E\right), \ldots, \tilde{s}\left(A_{N!}, E\right)\right)$. We now show that identification of $H\left(\tilde{S}\right)$ gives the identification of $F(B)$, the distribution of unobserved complementarities.

Let

$$\tilde{r}\left(A; B\right) = \sum_{\langle u,d\rangle \in A} b_{u,d} - \sum_{\langle u,d\rangle \in A_1} b_{u,d}, \tag{10}$$

where for notational compactness we define $b_{u,1} = b_{1,d} = 0$ for all u and d. The term $\tilde{r}\left(A; B\right)$ gives the sum of the unobserved complementarities in B corresponding to the indices of the matches in A minus the same sum for $A_1 = \{\langle 1,1\rangle, \ldots, \langle N, N\rangle\}$.

One of the main results of the paper is that the distribution $F(B)$ of unobserved complementarities is identified.

Theorem 1.

1. $\tilde{s}(A; E) = \tilde{r}(A; B)$ *for any A and where B is formed from E.*

2. $\tilde{r}(A; B_1) = \tilde{r}(A; B_2)$ *for all A if and only if $B_1 = B_2$.*

3. *Therefore, the distribution $F(B)$ is identified because the distribution of $H\left(\tilde{S}\right)$ is identified.*

The proof is in the appendix. The first part of the theorem states that the sum of unobserved match production for an assignment can be computed using the elements of B. Therefore, knowledge of B can be used to compute pairwise stable assignments, for example for counterfactual analysis. Likewise, knowledge of $F(B)$ lets one calculate assignment probabilities $\Pr(A \mid Z; F)$. The second part of the theorem states that there is a one-to-one mapping between the sums of unobserved assignment production for assignments and values of B. Therefore, as the distribution $H\left(\tilde{S}\right)$ of the sums of unobserved match production for assignments is identified, so is the distribution $F(B)$ of unobserved match complementarities.

Example. 1 $(N = 3)$ By definition,

$$
\begin{pmatrix} \tilde{r}(A_2; B) \\ \tilde{r}(A_3; B) \\ \tilde{r}(A_4; B) \\ \tilde{r}(A_5; B) \\ \tilde{r}(A_6; B) \end{pmatrix} = \begin{pmatrix} b_{1,2} + b_{2,1} + b_{3,3} - (b_{1,1} + b_{2,2} + b_{3,3}) \\ b_{1,3} + b_{2,2} + b_{3,1} - (b_{1,1} + b_{2,2} + b_{3,3}) \\ b_{1,2} + b_{2,3} + b_{3,1} - (b_{1,1} + b_{2,2} + b_{3,3}) \\ b_{1,1} + b_{2,3} + b_{3,2} - (b_{1,1} + b_{2,2} + b_{3,3}) \\ b_{1,3} + b_{2,1} + b_{3,2} - (b_{1,1} + b_{2,2} + b_{3,3}) \end{pmatrix} = \begin{pmatrix} b_{3,3} - (b_{2,2} + b_{3,3}) \\ b_{2,2} - (b_{2,2} + b_{3,3}) \\ b_{2,3} - (b_{2,2} + b_{3,3}) \\ b_{2,3} + b_{3,2} - (b_{2,2} + b_{3,3}) \\ b_{3,2} - (b_{2,2} + b_{3,3}) \end{pmatrix} = \begin{pmatrix} -b_{2,2} \\ -b_{3,3} \\ b_{2,3} - (b_{2,2} + b_{3,3}) \\ b_{2,3} + b_{3,2} - (b_{2,2} + b_{3,3}) \\ b_{3,2} - (b_{2,2} + b_{3,3}) \end{pmatrix},
$$

(11)

where the second equality uses $b_{u,1} = b_{1,d} = 0$ for all u and d. Then using (9) for each of the four $b_{u,d}$'s and (5) for each of the five $\tilde{s}(A; E)$'s allows one to algebraically verify Theorem 1.1 for $N = 3$. The interesting direction of Theorem 1.2 for $N = 3$ states that $B_1 = B_2$ whenever $\tilde{r}(A; B_1) = \tilde{r}(A; B_2)$ for all A. This direction can be verified because $\tilde{r}(A_2; B)$ through $\tilde{r}(A_5; B)$ can be easily solved for the four elements of B. The less interesting direction of Theorem 1.2 always holds by the definition of $\tilde{r}(A; B)$ to be a function of B. Given that we previously showed that $H\left(\tilde{S}\right)$ is identified, $F(B)$ is also identified. △

2.7 Overidentification

The distribution of unobserved match characteristics $G(E)$ is not identified. Despite the model primitive $G(E)$ not being identified, the distributions $H\left(\tilde{S}\right)$ and $F(B)$ are not only identified, they are overidentified. The proof of Lemma 1 works by setting $H\left(\tilde{S}^\star\right) = \Pr(A_1 \mid Z^\star)$, where $\tilde{S}^\star$ is the point of evaluation of the CDF H, A_1 is the diagonal assignment $\{\langle 1, 1 \rangle, \dots, \langle N, N \rangle\}$ and $Z^\star$ is a specific value of Z chosen based on the value $\tilde{S}^\star$. One can identify the entire model if one only observes, in each market, whether assignment A_1 occurs or not. The assignment A_1 is just one of $N!$

16

assignments. Given the adding up constraint that the sum of probabilities of assignments is always 1, there are $N! - 2$ other probabilities $\Pr(A \mid Z)$ for each Z available to overidentify the model.

The necessity of using only one assignment probability in a proof of identification is analogous to the identification arguments for the single-agent multinomial choice model in the frameworks of Thompson (1989) and Lewbel (2000). In such multinomial choice models, only the probability of a single choice is necessary for identification. Given that choice probabilities sum to one, all but two choices provide overidentifying restrictions. Overidentification in the semiparametric multinomial choice model has not been formally exploited to form an operational testing procedure in finite samples. Given that the simpler multinomial choice model should be explored before matching models, we leave the formal exploitation of overidentification to further research.[17]

3 Generalizations of the Baseline Model

We consider two generalizations of the identification result for the one-to-one matching game where all agents are matched.

3.1 Other Observed Variables X and Random Preferences

In addition to the large support, match-specific characteristics Z, researchers often observe other match-specific and agent-specific characteristics, which we collect in the random variable X, which we think of as a long vector. We also include in X the number of agents on each side, N, to allow the size of the market to vary across the sample. An example of a production function augmented by the elements of X is

$$(x_u \cdot x_d)' \beta_{u,d,1} + x'_{u,d}\beta_{u,d,2} + \mu_{u,d} + z_{u,d}, \tag{12}$$

where x_u is a vector of upstream firm characteristics, x_d is a vector of downstream firm characteristics, $x_u \cdot x_d$ is a vector of all interactions between upstream and downstream characteristics, $x_{u,d}$ is a vector of match-specific characteristics, $\mu_{u,d}$ is a random intercept capturing unobserved characteristics of both u and d, and $\beta_{u,d,1}$ and $\beta_{u,d,2}$ are random coefficient vectors specific to the match. The two random coefficient vectors can be the sum of the **random preferences** of upstream and downstream firms for own and partner characteristics. In a marriage setting, we allow men to have heterogeneous preferences over the observed characteristics of women, which is one explanation for why observationally identical men marry observationally distinct women.

---

[17]Another source of overidentification arises if the researcher imposes that $G(E)$ and hence, by Lemma 2.3, $F(B)$ are exchangeable in agent indices. Exchangeability in agent indices is a restriction of the class of allowable F's but F is identified without assuming such a restriction.

In this example,

$$X = \left(N, (x_u)_{u \in N}, (x_d)_{d \in N}, (x_{u,d})_{u,d \in N} \right).$$

Now we define

$$e_{u,d} = (x_u \cdot x_d)' \beta_{u,d,1} + x'_{u,d} \beta_{u,d,2} + \mu_{u,d}$$

and, as before notationally,

$$c_{u_1,d_1,u_2,d_2} = e_{u_1,d_1} + e_{u_2,d_2} - (e_{u_1,d_2} + e_{u_2,d_1}).$$

Using the same notation as before, we define the array of unobserved complementarities as (7). This definition of C, and similarly of B, now depends on the realizations of X. Our previous identification argument in Theorem 1 does not use X. Therefore we can condition on a realization of X to identify the conditional-on-X distribution of unobserved complementarities $F(B \mid X)$. We of course require variation in Z as before, but now Z must have full support conditional on each realization of X. We do not require that C, B and E are independent of X, but all unobservables must still be independent of Z conditional on X.

Corollary 1. *The distribution $F(B \mid X)$ of market-level unobserved complementarities conditional on X is identified.*

Our identification of distributions of heterogeneity conditional on X follows arguments in the multinomial choice literature (Lewbel, 2000; Matzkin, 2007; Berry and Haile, 2010). This is a standard object of identification in the cited literature

We could further attempt to unpack the identified $F(B \mid X)$ into the distribution of individual random coefficients and additive unobservables, such as the vectors $\beta_{u,d,1}$ and $\beta_{u,d,2}$ and the unobserved complementarities induced only by the scalar $\mu_{u,d}$ in the example production function (12). We would need to assume full independence between the primitive unobservables and the elements of X. Using (12), we can think of the definition of $b_{u,d}$, (8), as defining a system of $(N-1)^2$ seemingly unrelated equations, relating $b_{u,d}$ to the elements of X, the random coefficients and the additive unobservables. Masten (2015) studies in part seemingly unrelated regressions with random coefficients and shows that the marginal distribution of each random coefficient or additive unobservable is identified but the joint distribution of the random coefficients and additive unobservables entering all equations is sometimes not identified. One intuition is that the number of elements of X must weakly exceed the number of random coefficients and additive unobservables. Once $F(B \mid X)$ is identified and the problem of unpacking F into the joint distribution of random coefficients and additive unobservables is placed in the framework of Masten (2015), the remaining identification issues are less specific to matching and so are not considered further here.

3.2 Heterogeneous Coefficients on Match Characteristics

We now define the production to a match $\langle u, d \rangle$ to be

$$e_{u,d} + \gamma_{u,d} \cdot z_{u,d}, \tag{13}$$

where $\gamma_{u,d} \neq 0$ is a match-specific coefficient. The coefficients $\gamma_{u,d}$ vary across matches within each matching market but not across markets. Therefore, the $\gamma_{u,d}$ are fixed parameters to be identified and not random coefficients. Fixing coefficients across markets but not within markets makes sense in a context where firm indices like u and d have a consistent meaning across markets. For example, the same set of upstream and downstream firms may participate in multiple matching markets, as in Fox (2010a), where each market is a separate automotive component category.[18] As we need the $z_{u,d}$'s to identify $F(B \mid X)$, we rule out the case where any $\gamma_{u,d} = 0$.

We apply a scale normalization on production by setting $\gamma_{1,1} = \pm 1$. Because of transferable utility, we can identify the relative scale of each match's production. We use the matrix $\Gamma = (\gamma_{u,d})_{u,d \in N}$. It is first important to note that parts 1 and 2 of Theorem 1 are only about the random variable unobservables B and E and so do not involve whether $z_{u,d}$ has a parameter $\gamma_{u,d}$ on it or not. So those statements in Theorem 1 still hold in this more general setting. Next we state that the analog to the identification claim in the third part of Theorem 1 holds in the setting with fixed parameters $\gamma_{u,d}$ on $z_{u,d}$.

Theorem 2. *The distribution $F(B \mid X)$ and the fixed matrix of parameters $\Gamma = (\gamma_{u,d})_{u,d \in N}$ are identified.*

The proof is in the appendix.[19]

4 Data on Unmatched Agents

Up until this point, we have considered matching games where all agents have to be matched. We infer $F(B \mid X)$ from sorting patterns in the data. This approach is reasonable when only data on observed matches are available. For example, it may be unreasonable to assume that data on all potential

---

[18]In a marriage setting with different individuals in each market, we could assume that $\gamma_{u,d}$ is the same for all matches where the men are all in the same demographic class (such as college graduates) and the women are all in the same demographic class (such as high-school graduates). This suggested use of demographic classes is partially reminiscent of Chiappori et al. (2015), who use data over time on the US marriage market to estimate a different variance of the type I extreme value (logit) utility errors in a Choo and Siow (2006) style model for each male demographic class and for each female demographic class. Our suggested approach in this footnote lets $\gamma_{u,d}$ vary by the intersection of male and female demographic classes, instead of agents. For example, there could be a $\gamma_{\text{age 40 men,age 30 women}}$ specific to the listed demographic groups.

[19]We could also study the production function (13) when each $\gamma_{u,d}$ is a random coefficient such that the random matrix $\Gamma = (\gamma_{u,d})_{u,d \in N}$ has some joint distribution $J(\Gamma)$ that describes how Γ varies across markets. An identification at infinity proof technique, where all but two $z_{u,d}$'s are set to $-\infty$, identifies the marginal distribution of each $\gamma_{u,d}$ but not the joint distribution $J(\Gamma)$.

entrants to a matching market exist. In some situations, however, researchers can also observe the identities of unmatched agents. Data are available, for example, on potential merger partners in some industry that do not end up undertaking mergers or on single people in a marriage market.[20] When data on unmatched agents do exist, we can go beyond unobserved complementarities B and identify the distribution of match-specific unobservables E.

Here, X can contain separate numbers of downstream firms N^d and upstream firms N^u. Let

$$
E = \left(\begin{array}{ccc} e_{1,1} & \cdots & e_{1,N^d} \\ \vdots & \ddots & \vdots \\ e_{N^u,1} & \cdots & e_{N^u,N^d} \end{array} \right).
$$

Use $\langle u, 0 \rangle$ and $\langle 0, d \rangle$ to denote an upstream firm and a downstream firm that are not matched. An assignment A can be $\{\langle u_1, 0 \rangle, \langle u_2, d_2 \rangle, \langle 0, d_2 \rangle\}$, allowing single firms. We do not require match-specific characteristics $z_{u,0}$ and $z_{0,d}$ for unmatched firms; they can be included in X if present.

The data generating process is still (3). One difference is that a pairwise stable assignment needs to satisfy individual rationality: each non-singleton realized match has production greater than 0.

Theorem 3. *The distribution $G(E \mid X)$ of market-level unobservables is identified with data on unmatched agents.*

The proof shows that the distribution $G(E \mid X)$ for some X can be traced using the probability that all agents are unmatched, conditional on Z. The individual rationality decision to be single identifies $G(E \mid X)$ while the sorting of matched firms to other matched firms identifies only $F(B \mid X)$. Using an individual rationality condition is more similar to the utility maximization assumptions used in the identification of single-agent discrete choice models and discrete Nash games (Lewbel, 2000; Matzkin, 2007; Berry and Haile, 2010; Berry and Tamer, 2006). An agent can unilaterally decide to become unmatched.

5 Agent-Specific Characteristics

Return to the case with only unmatched agents in the data. Match specific z's with full support are not always available in datasets. For example, in our venture capital work say we observed only patents for startups and the total experience in past deals for venture capitalists (in reality, we do observe match characteristics). Also say, for example, that the induced match specific characteristic $z_{u,d} = z_u \cdot z_d$, where z_u is the upstream firm characteristic experience and z_d is the downstream firm characteristic patents. Say the long vector of agent-specific characteristics $\left((z_u)_{u \in N}, (z_d)_{d \in N} \right)$ has

---

[20]For example, Uetake and Watanabe (2012) study mergers between rural banks, where each county is a separate matching market.

support in $\mathbb{R}^{2N}$. Even this product support on agent characteristics will not allow the matrix Z with induced match-specific characteristics $z_{u,d} = z_u \cdot z_d$ to have support on say $\mathbb{R}^{N^2}$.

This is a problem for identification in the baseline model because the matrix of match-specific unobservables E has N^2 elements and the matrix of unobserved complementarities B has $(N-1)^2$ elements. There needs to be some symmetry between the distributions of observables and unobservables in the model. More prosaically, we cannot prove Lemma 1 when Z has limited support, such as when its constituents $\left((z_u)_{u \in N}, (z_d)_{d \in N}\right)$ vary only in $\mathbb{R}^{2N}$ while the match-specific unobservables E vary over $\mathbb{R}^{N^2}$.

The solution is to restrict attention to a production function where, in the primitive model, unobserved characteristics enter symmetrically to observed characteristics. If observed characteristics vary at the agent level, then unobserved characteristics should also be restricted to only vary at the agent level. Taking the example of $z_{u,d} = z_u \cdot z_d$ above, consider the production function

$$e_u \cdot e_d + z_u \cdot z_d, \tag{14}$$

where e_u and e_d are unobserved agent-specific characteristics.[21] Unobserved agent characteristics enter symmetrically to the observed agent characteristics in the production function. Say the unobserved agent characteristics $\left((e_u)_{u \in N}, (e_d)_{d \in N}\right)$ take on support on $\mathbb{R}^{2N}$. Then observed agent characteristics $\left((z_u)_{u \in N}, (z_d)_{d \in N}\right)$ also should vary on $\mathbb{R}^{2N}$ in order for slight extensions to the previous identification arguments to go through. Specifically, one can alter the first lines of the proof of Lemma 1 and then that lemma and hence the remainder of the identification machinery leading up to and including Theorem 1 will apply to the agent-specific case.

Define $e_{u,d} = e_u \cdot e_d$ and define unobserved complementarities using (6). We show that the distribution of the matrix B of unobserved complementarities is identified.

Corollary 2. *The distribution $F(B \mid X)$ of unobserved complementarities B is identified in the agent-specific case.*

The short proof is in the appendix.

6 Many-to-Many Matching

Venture capitalists can make multiple investments during the same year. Likewise, startups often contract with multiple venture capitalists, although for simplicity our application considers only the lead venture capital investor in a startup. It is important to extend the previous results to many-to-many, two-sided matching.

Consider a two-sided matching game where upstream firm u can make a quota of q_u possible matches and downstream firm d can make q_d possible matches. The researcher has data on q_u and

---

[21] All assignments would be pairwise stable if the match production was instead $e_u + e_d + z_u + z_d$.

q_d and the quotas can vary across firms in the same market and across markets. The previous case of one-to-one matching is $q_u = q_d = 1$ for all firms. Leaving a quota slot unfilled gives production of zero for that slot. The number of upstream firms N^u may differ from the number of downstream firms N^d.

Let the production function for an individual match still be (1) and let the production of the matches of the single upstream firm u with the pair of downstream firms d_1 and d_2 be equal to

$$z_{u,d_1} + e_{u,d_1} + z_{u,d_2} + e_{u,d_2}.$$

This *additive separability* in the production of multiple matches involving the same firm yields the many-to-many matching model of Crawford and Knoer (1981) and Sotomayor (1992, 1999). Like in the one-to-one case, a pairwise stable assignment is proven to exist, to be efficient and to be unique with probability 1. Redefine the following objects to allow $N^u \neq N^d$:

$$E = \begin{pmatrix} e_{1,1} & \cdots & e_{1,N^d} \\ \vdots & \ddots & \vdots \\ e_{N^u,1} & \cdots & e_{N^u,N^d} \end{pmatrix}, Z = \begin{pmatrix} z_{1,1} & \cdots & z_{1,N^d} \\ \vdots & \ddots & \vdots \\ z_{N^u,1} & \cdots & z_{N^u,N^d} \end{pmatrix}, B = \begin{pmatrix} b_{2,2} & \cdots & b_{2,N^d} \\ \vdots & \ddots & \vdots \\ b_{N^u,2} & \cdots & b_{N^u,N^d} \end{pmatrix}.$$

We also extend to many-to-many matching the model in Section 3.2, where the production of a match between u and d is $\gamma_{u,d} \cdot z_{u,d} + e_{u,d}$. The matrix of homogeneous parameters is $\Gamma = (\gamma_{u,d})_{u \in N^u, d \in N^d}$.

Say first that the number of firms, quotas and production functions are such that all firms make a number of matches equal to their quotas: there are no unused quota slots in equilibrium. Leaving no unused quota is feasible if

$$\sum_{u=1}^{N^u} q_u = \sum_{d=1}^{N^d} q_d.$$

In this case, every mathematical argument for the baseline model in Section 2 and many of the subsequent models extends to many-to-many matching. In particular, the distribution of unobserved complementarities $F(B \mid X)$ is identified using the sorting patterns in the data. Likewise, if unmatched firms are in the data and so quota slots can be left unused, the same analysis as Section 4 applies.

Corollary 3. *Consider the many-to-many matching model.*

1. *The distribution of unobserved complementarities $F(B \mid X)$ and the coefficients Γ (if included) are identified from data on matches only.*

2. *The distribution of match-specific unobservables $G(E \mid X)$ is identified if unmatched firms are in the data.*

The proof is omitted as it just checks previous mathematical arguments to see that properties unique to one-to-one matching with $N^u = N^d = N$ are not used.

7 Matching with Trades

Investigating a fairly general matching model is useful because many models of empirical interest will be special cases. Consider matching with trades in the so-called trading networks model in Hatfield, Kominers, Nichifor, Ostrovsky and Westkamp (2013), which is a significant generalization of Kelso and Crawford (1982). In matching with trades, agents engage in trades ω from some finite set Ω. A trade indexes the name of the buyer and the name of the seller and can specify other aspects, such as the quality or other specifications of the goods in question. In a labor market, trades could specify benefits such as health care plans and vacation time. In venture capital, a trade could specify the number of board seats a startup gives a venture capitalist. Trades generalize our previous notion of a match. We require data on all aspects of the trade; if quality is part of a trade then the qualities for all trades in the set Ω must be measured. The price of trade ω is p_ω, although, as before, we study identification when prices are not observed in the data. Prices play the same role as transfers in the earlier matching models.

Firms are not necessarily divided into buyers and sellers ex ante; a firm can be a buyer on some trades and a seller on other trades. In a model of mergers, a firm is not ex ante either a target or acquirer; these roles arise endogenously as part of a stable outcome. Two-sided, many-to-many matching is a strict special case of trading networks where the profits of an upstream firm undertaking trades as a buyer are $-\infty$ and, likewise, the profits of a downstream making trades as a seller are $-\infty$.

7.1 Simple Matching with Trades

As there are no ex ante upstream and downstream firms, index a firm by i. Consider first the case where the production of a trade ω between buyer i and seller j is

$$z_\omega + e_\omega. \tag{15}$$

Notationally, the indices of i and j are subsumed into the trade ω. If a trade should give production of $-\infty$, we notationally remove it from Ω. This matching with trades game is a generalization of the two-sided, many-to-many matching game in Section 6. In this simple setup, trades that give positive production occur and trades that give negative production do not occur. We observe the entire set of trades Ω for each market, so the data measures whether a trade occurs or does not occur; firms that make no trades are therefore observed as well. Let the vector $Z = (z_\omega)_{w \in \Omega}$ and, similarly, $E = (e_\omega)_{\omega \in \Omega}$. Let X collect observables entering E.

Theorem 4. *Consider a trades model where the production of trade ω is $z_\omega + e_\omega$. Then $G(E \mid X)$ is identified.*

7.2 Matching with Trades Under Submodularity

We now consider matching with trades where firms have profit functions defined over portfolios of trades. Let $\Omega_i \subset \Omega$ be the set of trades where i is either a buyer or a seller. The **individual profit** of a firm i undertaking the trades $\Psi_i \subseteq \Omega_i$ at prices p_ω for $\omega \in \Omega$ is

$$u(i, \Psi_i) + \sum_{\omega \in \Psi_{i\rightarrow}} p_\omega - \sum_{\omega \in \Psi_{\rightarrow i}} p_\omega, \tag{16}$$

where the set $\Psi_{i\rightarrow}$ is the trades in Ψ_i where i is the seller and $\Psi_{\rightarrow i}$ is the trades in Ψ_i where i is the buyer. Hatfield et al. prove that a stable assignment (here a set of trades) exists and is efficient (and therefore unique with probability 1) under a condition on preferences called **substitutes**. A companion paper shows that the substitutes condition is equivalent to the indirect utility (profit) version of the direct utility (profit) in (16) being **submodular** for all vectors of prices, p_ω for $\omega \in \Omega$ (Hatfield, Kominers, Nichifor, Ostrovsky and Westkamp, 2015, Theorem 6). See the cited paper for a definition of submodularity. Submodularity of the indirect utility function is restrictive for many empirical applications. However, submodularity is only a restriction when the profit from a set of trades is not additively separable across the trades. Therefore, the underlying direct utility firm profits justifying the production-of-a-trade (15) in Section 7.1 imply that the corresponding indirect utility functions are submodular.

For all firms i and trades $\psi_i \subseteq \Omega_i$ let the **pre-transfer profit** (or **valuation**) be

$$u(i, \Psi_i) = z_{i,\psi_i} + e_{i,\psi_i},$$

where z_{i,ψ_i} is an observable specific to firm i and the set of trades ψ_i and e_{i,ψ_i} is an unobservable specific to firm i and the set of trades ψ_i. Let $Z = (z_{i,\psi_i})_{i \in N, \psi_i \subseteq \Omega_i}$ be the array of observables corresponding to pairs of firms and sets of trades and let $E = (e_{i,\psi_i})_{i \in N, \psi_i \subseteq \Omega_i}$ be a similar array for unobservables. The arrays Z and E are typically large but are always finite as the set of trades Ω is finite. For identification, the support of $-Z$ must be a weak superset of the support of E.[22] Further, the supports of E and Z should be restricted so that the corresponding indirect utility functions are submodular for all players for all realizations of E and Z. We leave to other work the question of how to enforce submodularity in empirical applications.[23] As before, observable characteristics other than

---

[22]Therefore, if the support of E is a product space, then the support of $-Z$ must contain that product space. This is in principle a very strong requirement because z_{i,ψ_i} varies by the identity of the firm i and set of trades ψ_i. If restrictions are placed on how e_{i,Ψ_i} varies across i and ψ_i than correspondingly less variation is needed in z_{i,ψ_i}.

[23]If the submodularity condition fails, a stable assignment or a competitive equilibrium as defined in Hatfield et al. (2013) may fail to exist. One practical but perhaps inelegant approach is to ignore the lack of existence. As an example of this approach, the important contribution by Ciliberto and Tamer (2009) estimates a Nash game of finite actions and restricts attention to pure strategy equilibria, even though Nash's existence theorem applies to mixed strategy equilibria. Note that Hatfield et al. use a stability definition that is stronger than *pairwise* stability. Under the models discussed earlier in the current paper, *pairwise* stability implies the stronger notion of stability, while the equivalence does not hold in Hatfield et al.. Theorem 5 uses the stability definition in Hatfield et al..

the z_{i,ψ_i} are collected in a long vector X and can enter e_{i,ψ_i}.

Say that the profit from making no trades is zero and that the researcher observes data on firms that make no trades. Then the following identification result holds.

Theorem 5. *The distribution of unobservables $G\left(E \mid X\right)$ is upper bounded by an identified function $\bar{G}\left(E \mid X\right)$ in the trading networks model. $\bar{G}\left(E \mid X\right) < 1$ if, for each X and Z, there exists E with positive probability where trades occur.*

The statement means that we can identify a function $\bar{G}\left(E \mid X\right)$ such that $G\left(E \mid X\right) \leq \bar{G}\left(E \mid X\right)$ for all arrays of unobservables E and conditioning observables X. A distribution function is a probability, so the trivial bound $\bar{G}\left(E \mid X\right) = 1$ satisfies this property. However, if some assignment other than the assignment with no trades occurs with positive probability, then $\bar{G}\left(E \mid X\right)$ is a tighter bound than the trivial bound of 1.[24]

The bound $\bar{G}\left(E \mid X\right)$ in the proof Theorem 5 is actually $\Pr\left(A_0 \mid Z^\star, X\right)$ for some $Z^\star$, where A_0 is the assignment where no trades are made. The proof of Theorem 5, in the appendix, extends the argument in the proofs of Theorems 3 and 4. In the proofs of Theorems 3 and 4, $G\left(E \mid X\right)$ itself and not a bound equals $\Pr\left(A_0 \mid Z^\star, X\right)$. This is because the unobservables $e_{u,d}$ in Theorem 3 and e_ω in Theorem 4 correspond to the production of a match or trade, which is the sum of profits of the two firms for the match or trade. In Theorem 5, the unobservables e_{i,Ψ_i} correspond to the profit of an individual firm i and not the production of all firms in the trades. The theorem shows that it is possible to identify bounds on distributions of aspects of individual firm profit functions (up to scale) and not just aspects of production functions for matches, as in earlier results. The reason is that the individual profit functions are not additively separable across individual trades, leaving no role for the concept of the production of a trade.

8 Biotech and Medical Venture Capital

We estimate the roles of observed match and firm characteristics as well as the distribution of unobserved complementarities in the biotech and medical/health venture capital industries. In these industries, investment firms known as venture capitalists contribute money to entrepreneurial startups. We seek to understand the role of venture capitalists in the productive surplus of an investment; this contribution to match production is only present if the venture capitalist adds value over and above offering financing. We present separate parameter estimates for the biotech and medical industries.

In most cases, the first round of venture capital funding is secured well before an entrepreneur is ready to market its products to consumers or even to undertake a final round of testing. Therefore, the first round venture capital funding is essential to nurturing the entrepreneur during a period where the entrepreneur has low or no revenues of its own. We study only the identity of the lead venture

---

[24]The bound is likely not sharp. Indeed, it is possible $G\left(E \mid X\right)$ is point identified and we do not know the proof.

capital investor in the first round of funding. This lead venture capitalist often takes more of an active management role in the startup than other investors.

We model each life science venture capital industry as a many-to-one, two-sided matching market where each entrepreneur is funded by the lead venture capitalist and each venture capitalist can fund multiple startups. We lack data on unfunded startups and venture capitalists who make no investments. Because we focus on many-to-one matching and consider only matched firms, the appropriate nonparametric identification result is the many-to-one special case of the many-to-many result in Corollary 3.1.

Sørensen (2007) uses a structural approach to estimate a matching game between entrepreneurs and venture capitalists. Sørensen estimates a matching game where matched agents could not exchange transfers, unlike the transferable utility matching game we study. His assumption is that the unobservables are match-specific, normally distributed and independent across matches involving the same or differing firms. By assuming independence across matches involving the same firm, Sørensen's model rules out that matches for many firms tend to be unobservedly more productive than matches for many other firms. Allowing for correlation in the matches involving the same firm may be important in venture capital, as such a correlation structure allows certain firms to contribute more to match profit than other firms. These are the high type firms: highly capable venture capitalists or entrepreneurial firms with great prospects in the life sciences industries.

8.1 Data and Observed Characteristics

We start with a carefully collected dataset on, ideally, all venture capital transactions in the biotech and medical industries. The data come from ThomsonOne. We then merge the venture capital data with data from the US Patent and Trademark Office on the stock of patents held by the entrepreneurial firms at the time of the first round investment that we model. Our data showed that the number of venture capital deals increased substantially after 1996 and our patent data have missing records after 2008. Therefore, we use data on venture capital deals between 1997 to 2007, although we use earlier years of data to compute venture capitalist experience.

8.2 Matching Markets

For computational reasons to be discussed, our method of simulated moments estimator can handle only what we describe as medium sized matching markets. Therefore, our matching market definition is made to keep the number of matches medium sized: small compared to the entire biotech and medical venture capital industries but still large compared to the number of potential entrants in the entry literature in industrial organization (Bresnahan and Reiss, 1991; Berry, 1992; Ciliberto and Tamer, 2009). We define a matching market to be one of the eleven two-digit biotech sectors in Table 1 in a particular calendar year of the data. Based on our matching market definitions, in our

Table 1: Two Digit Biotech and Medical Sectors Used to Define Matching Markets

Biotech	Medical
Human	Diagnostics
Agricultural and animal	Therapeutics
Industrial	Health products
Biosensors	Health services
Research & production equipment	Pharmaceuticals
Research & other services	

model venture capitalists consider matching only with the set of entrepreneurs in the same two-digit biotech/medical sector and year that, in the data, the venture capitalist's match partner was in.[25] The assumption that venture capitalists pre-commit to a two-digit sector is strong; relaxing the restriction to two-digit sectors requires addressing computational concerns.[26]

We only consider two-digit sectors with fewer than 30 startups in a year. We estimate the model parameters separately for the biotech and medical industries. For biotech, we have 38 matching markets. The mean number of startups per matching market is 7.2 with a maximum of 27. The mean number of venture capitalists per matching market is 6.8 with a maximum of 25. The 6.8 venture capitalists per matching market is only a little lower than 7.2 startups because there are only a small number of venture capitalists making multiple matches in the same two-digit biotech sector and year. Now consider the medical industry, which has larger numbers of startups per two-digit sector and year. There are 15 matching markets with fewer than 30 startups. Among those 15 markets, the mean number of startups is 22.3 with a maximum of our chosen upper bound of 30. The mean number of venture capitalists is 21.5 with the maximum, in the estimation sample, of 30.[27]

Again, we model the lead (largest) venture capital investor in the first round of venture capital funding. Each entrepreneur appears once in the data, reflecting this first round. Each venture capitalist can occur multiple times. A venture capitalist can be engaged in multiple investments within a year and can be observed in multiple years. If a venture capitalist makes multiple first round investments as the lead investor in a given year and two-digit sector, it is treated as a single firm with a quota (the maximum number of matches it can make) equal to the number of matches that venture capitalist made in the data for that year and two-digit sector. A venture capitalist with a single match in the data has a quota of one. Our model's use of these quotas focuses on time constraints as the reason

---

[25]Our market definition differs from Sørensen (2007), who defines a matching market as a six month period in one of two US states, California and Massachusetts. We use worldwide data and impose no limits that venture capitalists and entrepreneurs look for partners only within one narrow geographic region, which corresponds with our data.

[26]Sheng (2014) is an estimator for large network games that could likely be applied to matching games in future work. The computational savings of the estimator result in set instead of point identification in the limit, which is not in the spirit of our paper's theoretical results about point identification.

[27]We use servers with up to twenty cores for estimation. Using a cluster of multiple servers could allow us to increase the number of agents per matching market some, but not tremendously so because of the curse of dimensionality with computing a pairwise stable assignment to the matching game.

a superstar venture capitalist is not the lead investor in all entrepreneurial startups. Indeed, a lead venture capital investor takes seats on the board and an active management role in its entrepreneurial firms. The venture capitalist has scarce time to do this and so carefully selects a small number of investments, in the context of a matching model where it competes with other venture capitalists for these deals.

We do not model synergies between multiple entrepreneurs matched to the same venture capitalist; the production of a set of matches involving the same venture capitalist is equal to the sum of the production of the individual matches, as in our Corollary 3 but not our Theorem 5. Therefore, we do not study venture capitalist financial strategies such as portfolio diversification. We also do not model post-matching externalities in match production caused by, say, multiple entrepreneurs competing to treat the same, narrowly defined medical condition. It is rare for entrepreneurs in the same two-digit sector and year of the initial investment to be directly competing in the sense of treating, say, the same, narrowly defined medical condition. Therefore, in our model an entrepreneur cares about the outcome of the venture capital market for matches only because it affects the entrepreneur's own final venture capital match and corresponding transfer, not because a rival's match with a top venture capitalist could create a fierce competitor for consumers.

Unlike, say, a dataset on mergers, there are no unmatched firms in our data. While presumably there are entrepreneurial firms that fail to secure a first round of venture capital funding and venture capitalists with the equivalent of free quota slots (say spare time to help manage an additional startup), our data do not cover them. In what follows, our model operates as if these unmatched entrepreneurs and venture capitalist quota slots do not exist.[28]

8.3   Observables in the Match Production Function

The production function for the output of a match involving one entrepreneur d and one venture capitalist u is

$$\pm 1 \cdot \text{Distance}_{u,d} + \beta_{\text{Sector}} \cdot \text{SectorExper}_{u,d} + \beta_{\text{ExperPatents}} \cdot \text{TotalExper}_u \cdot \log\left(\text{Patents}_d + 1\right) + e_{u,d}. \quad (17)$$

In this section, we discuss the contribution of each of the listed observable (in the data) variables; we postpone a discussion of the unobservables until the next section. Overall, we feel we have collected close to the best firm and match characteristics on a broad portion of the VC sector that academic researchers could have access to. As we will see, even these rich characteristics will leave room for unobservables.

Table 2 reports the means and standard deviations across realized matches for the key observable characteristics in the production function. Estimating the matching model requires us to compute

---

[28]The matching maximum score estimator of Fox (2010a) is robust to missing data on quotas in part because it does involve computing pairwise stable assignments as part of a nested fixed point procedure.

Name	Pre-Rescaling				Rescaled			
	Biotech		Medical		Biotech		Medical	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
log (Patent count + 1)	0.45	0.85	0.46	0.87	-0.04	0.97	0.01	1
VC overall experience	0.06	0.12	0.07	0.14	0.22	1.19	0.25	1.36
VC four-digit sector specific experience	0.19	0.26	0.13	0.18	0.04	1.05	-0.02	0.98
Distance: km/1000	1.74	3.08	2.07	3.2	0.05	1	-0.02	1
# of Patents (no logs)	2.12	7.48	2.35	11.16				
Interaction term: patents * log(patents)					0.02	1.08	0.18	1.62

Table 2: Summary Statistics for Realized Matches

these characteristics for both actual and counterfactual matches. In estimation, we rescale many of the variables as discussed separately for each variable; the table reports the variables before and after rescaling.[29]

The scale normalization of match production is in terms of the match-specific variable distance. Distance is measured as the distance on the surface of the Earth from the headquarters of the venture capitalist to the headquarters of the entrepreneur.[30] We have worldwide data so some of these distances are quite large: from Europe to Australia, say. Table 2 measures distance in thousands of kilometers. Hence, the table reports that the mean distance across realized matches is around 1700 kilometers. We allow the coefficient on distance to be either positive or negative, estimating the model once for a positive coefficient and a second time for a negative coefficient, taking the parameter estimates with the lowest objective function value.[31] Not surprisingly, we will find that the coefficient on distance is indeed negative. We rescale distance to have a mean of zero and standard deviation of 1.

Distance plays an important role in the venture capital literature. The literature has argued that geographic proximity helps investors and entrepreneurs find out about each other, thereby increasing investment likelihood (Sorenson and Stuart, 2001). Furthermore, Lerner (1995) finds that VCs are more likely to sit on boards of their portfolio companies the closer are the companies, a finding consistent with lower governance costs associated with geographically proximate investments. As a result, geographic proximity is likely to be a factor in selecting investment opportunities. Such governance considerations are likely to be especially important in empirical settings like ours in which startup assets are mainly intangible, the length of product development can be decades, and product development often costs hundreds of millions of dollars (Lerner, Shane and Tsai, 2003). Co-location

---

[29]The table reports summary statistics by realized matches but some rescalings are based on other samples, like all venture capitalists or all startups for agent-specific characteristics.

[30]We focus on headquarters rather than branch offices (for both startups and VCs) in part for data availability reasons. Institutionally, in many cases leaders from headquarters will need to be involved, particularly for the most important meetings.

[31]As discussed in Section 3.2, the assumption that distance is valued the same across all matches is more than a mere normalization. Using the fact that each venture capitalist appears in multiple markets, Theorem 2 allows us to identify a coefficient on distance specific to each venture capitalist. However, all of the parameters in (17), not just distance, have a homogeneous coefficient.

can help for governance and monitoring reasons, as well as for facilitating the provision of value-added business development services such as organizational professionalization (Hellmann and Puri, 2000).[32]

While VCs tend to specialize by factors such as the startup lifecycle stage of development, perhaps the most-mentioned aspect of specialization is industry experience (Hsu, 2004; Sørensen, 2007). Industry domain experience may be important in both assessing investment opportunities as well as intermediating startup business development services (such as connecting to executive-level managers in an industry domain or striking alliance relationships with industry incumbents) – indeed Sørensen estimates that the former factor is twice as important as the latter in explaining the likelihood that a startup goes public. Furthermore, Sorenson and Stuart (2001) find that industry domain experience can allow VCs to invest in more geographically distant portfolio firms.

Our measure of a venture capitalist's past experience in all biotech and medical sectors is equal to the number of past deals where that venture capitalist was the lead investor in the first round, which typically corresponds to knowledge in a startup domain. Experience is constructed using the complete history of our data, which starts in 1960. We do not wish for our measure of experience to trend over time, so we normalize experience each year to be between 0 and 1, with 1 being the venture capitalist with the most experience that year and 0 being the venture capitalist with the least experience. Table 2 reports that the mean level of the venture-capitalist specific variable experience across realized matches (not venture capitalists) in the estimation samples are 0.06 for biotech and 0.07 for venture capital.[33] This implies that the venture capitalist with the most experience in a given year typically has around 20 times the past deals as the mean venture capitalist. In estimation, we scale venture capitalist experience to have a mean of zero and a standard deviation of 1. This makes the standard deviation of experience similar to the standard deviation of distance.[34]

We also compute the venture capitalist's experience in the specific four-digit sector of an entrepreneur. There are 85 four-digit sectors in our biotech and medical data. Again, we normalize sector-specific experience to be between 0 and 1 for each year of data by making the venture capitalist with the most sector-specific experience have a value of 1. The mean level of four-digit, sector-specific experience across realized matches for biotech is 0.19, or about 1/5 of the past deals of the venture capitalist with the most sector-specific experience. For the medical industry, the mean sector-specific

---

[32]Our estimates are in units of distance instead of monetary units, such as dollars. Therefore, we cannot definitively say that the differences we estimate in match production correspond to large differences in monetary values. Nevertheless, the prior importance of distance in academic research on venture capital suggests that distance is economically important in monetary terms.

[33]Table 2 reports summary statistics for realized matches, so a venture capitalist's overall experience is counted twice if the venture capitalist makes two matches in a given year.

[34]Total experience is a count of past matches and so could be considered to be a function of a lagged dependent variable in another matching model where venture capitalists were allowed to be unmatched or have vacant quota slots and so the number of matches of each venture capitalist (in addition to the identity of the startup partners) was an outcome of the matching model. Past observed agent and match characteristics are statistically exogenous shifters of the outcomes of past matching games and so provide exogenous variation in past matches and hence lagged dependent variables. While this variation could lead to an approach that distinguishes venture capitalist serially correlated (over time) unobserved heterogeneity from true state dependence from experience, we do not explore that here as we assume that unobserved complementarities are statistically independent of experience.

experience is 0.13. In estimation, we rescale sector specific experience to have close to a mean of zero and a standard deviations of 1. Note that for a given venture capitalist, four-digit experience varies across entrepreneurs within the same two-digit sector. Sector experience is therefore a match-specific characteristic. We do not measure an entrepreneurial firm's experience because we observe no information on the founders of the startup.

The most important asset of an entrepreneurial startup is likely its intellectual property. Perhaps the only way for an outside researcher to directly measure this intellectual property is to look at patents. We have data on the number of patents held by the entrepreneurial firm at the time of the first round of venture capital investment. The literature has identified at least two distinct roles of patenting for startups: as a legal instrument to exclude others from using intellectual property in the product market (or to license those rights to others in the market for technology) and as signaling devices to capital providers. Patents are important in the biotechnology and medical industries for both reasons (Levin et al., 1987). Table 2 shows that the mean number of patents is just over 2 for both biotech and medical, with high standard deviations of 7.5 for biotech and 11.1 for medical. 32% of entrepreneurs have zero patents at the time of the initial round of venture capital investment. The production function uses the logarithm of the patent count plus one, rescaled to have a mean close to zero and a standard deviation close to 1. We recognize that patent counts are not a perfect measure of intellectual property; this partly motivates this paper's focus on unobservable characteristics.

As discussed in the identification sections, complementarities between matched firms drive matching. Therefore, the firm-specific but not match-specific characteristics venture capitalist experience and startup patent count would drop out of the calculation of the production maximizing assignment if included without interactions. The production function includes the interaction between overall venture capital experience and the log of the patent count plus one. There may be positive complementarities between startups' patent position and more experienced (and therefore more reputable) venture capitalists. In patenting's exclusionary role or in patenting's role of facilitating markets for technology, startups with more patents may wish to match with more experienced venture capitalists. Similarly on the signaling side, more experienced venture capitalists may value startup patents more highly (Hsu and Ziedonis, 2008). Table 2 reports sample statistics for the interaction term across actual matches (not hypothetical matches). For biotech, the interaction term has a mean of around zero and a standard deviation of 1.1. For medical, the mean is 0.2 and the standard deviation is 1.6.

8.4 Distribution of Unobserved Complementarities

In our empirical work, we will treat firm indices as having no common meaning across matching markets. For entrepreneurs, firm indices do in fact lack meaning, as each startup appears in exactly one matching market. Venture capitalists may appear in multiple matching markets, but we still treat the same venture capitalist in different matching markets as being a different firm.

The production function in (17) has match-specific unobservables. The identification argument in Corollary 3.1 uses match-specific characteristics, which in (17) are the distance between the head-quarters of the entrepreneur and the venture capitalist and the venture capitalist's experience in the entrepreneur's four-digit sector. We impose that unobservables and observables are distributed independently.

Given our data and approach of treating firm indices as irrelevant within a market, we impose that the joint distribution of unobserved complementarities is exchangeable in firm indices. Lemma 2.3 states that if $G(E)$ is unobservable in agent indices, then so is $F(B)$. This result extends naturally to many-to-one matching. Even though the number of venture capitalists can be less than the number of entrepreneurs, we treat each venture capitalist as a single firm and not a number of synthetic firms equal to the venture capitalist's quota.[35]

We operationalize our ideas about exchangeability using the following parametric structure. We assume the joint distribution F of the unobserved complementarities in B will be multivariate normal with the following properties

$$
\begin{aligned}
\mathrm{Corr}\left(b_{u_1,d_1}, b_{u_2,d_2}\right) &= \rho_1, \text{ if } u_1 \neq u_2, d_1 \neq d_2 \\
\mathrm{Corr}\left(b_{u_1,d_1}, b_{u_2,d_1}\right) &= \rho_2, \text{ if } u_1 \neq u_2 \\
\mathrm{Corr}\left(b_{u_1,d_1}, b_{u_1,d_2}\right) &= \rho_3, \text{ if } d_1 \neq d_2 \\
\mathrm{SD}\left(b_{u_1,d_1}\right) &= \sigma.
\end{aligned}
$$

There are different correlations between pairs of unobserved complementarities depending on whether the venture capitalists u_1 and u_2 are the same and whether the startups d_1 and d_2 are the same. Keep in mind that the unobserved complementarity $b_{u,d}$ always involves the upstream firm 1 and the downstream firm 1, in addition to the listed firms u and d. We will estimate ρ_1, ρ_2, ρ_3 and σ^2 in addition to the parameters in the production function (17). We collect these parameters into the vector $\theta = (\pm 1, \beta_{\mathrm{Sector}}, \beta_{\mathrm{ExperPatents}}, \rho_1, \rho_2, \rho_3, \sigma)$, where the ± 1 corresponds to the coefficient on distance.

By Lemma 2.2, there exists a distribution $G(E)$ for the unobserved match characteristics that induces our multivariate normal $F(B)$ by the transformation (8). By Example 2, the underlying distribution $G(E)$ giving our $F(B)$ is not in general the simple multivariate normal distribution in the text of the example.

---

[35] A venture capitalist firm typically has multiple employees known as venture capitalists. Our model and data focus on a VC firm, not its individual employees. One interpretation of $e_{u,d}$ might be that it includes information on the VC employee that would be assigned to the startup if the match forms.

8.5   Estimator

The many-to-one matching model is a special case of Sotomayor (1999), and therefore has a unique pairwise stable assignment with probability 1. Furthermore, this pairwise stable assignment can be computed with the linear program described in that paper. Therefore, a simulated nested fixed point estimator is appropriate, where the objective function involves an integral over the unobserved complementarities B and the corresponding integrand involves solving the linear program for each realization of B.

A likelihood exists because we have fully specified the data generating process. Let the object X_m collect all the match and firm-specific observables (including the number of firms on each side of the market and the elements that in the discussion of identification would instead be in Z) for realized and counterfactual matches for market m and let A_m be the assignment in the data for market m. The likelihood contribution for market m involves the computation of $\Pr(A_m \mid X_m; \theta)$, or

$$\Pr(A_m \mid X_m; \theta) = \int_B 1\left[A_m \text{ stable} \mid X_m, B; \theta\right] \tilde{d}F_B(B; \theta). \tag{18}$$

The indicator function $1\left[A_m \text{ stable} \mid X_m, B; \theta\right]$ is equal to 1 if A_m is computed to be the pairwise stable assignment for market m with draw B, using the linear program in Sotomayor (1999).

A computational challenge with the likelihood contribution (18) is that $\Pr(A_m \mid X_m; \theta)$ will typically be intractably close to 0. Consider the simple example of one-to-one matching without the option of being unmatched. There are $N!$ possible assignments with N upstream and N downstream firms. Typically, $\Pr(A_m \mid X_m; \theta)$ will be on the order of $1/N!$. As $N!$ can be close to the number of atoms in the universe with $N = 50$, the likelihood will involve the computation of intractably small numbers. The same issue with intractably small numbers will apply to generalized method of moments (GMM) estimators using the efficient choice of moments, which are based on the scores of the likelihood (McFadden, 1989; Hajivassiliou and McFadden, 1998).

Instead of attempting to compute $\Pr(A_m \mid X_m; \theta)$ directly, we work with a simulated moments estimator that uses moments that are easier to compute. Our chosen estimator is statistically inefficient but is tractable to compute. Let $g(A, X)$ be a function of an assignment A and agent and match characteristics X that gives some market-level output. Let $A(X, B; \theta)$ be the pairwise stable assignment of a market with observables X and unobserved complementarities B, evaluated at the parameter θ. With data on M markets, an empirical moment as a function of θ is

$$Q_{g,M}(\theta) = \frac{1}{M} \sum_{m=1}^{M} \left[\int_B g\left(A\left(X_m, B; \theta\right), X_m\right) \tilde{d}F_B(B; \theta) - g\left(A_m, X_m\right) \right]. \tag{19}$$

The moment condition is that $Q_{g,M}(\theta) = 0$. Each choice of g indexes a separate empirical moment

$Q_{g,M}(\theta)$.[36] The exact choices of $g(A,X)$ are described in Appendix B.[37] We have separate moments based on agent-specific and firm-specific moments. For example, a moment might be based on the quantiles of the match-specific characteristics for only the matches in the pairwise stable assignment. These selections for the moments $Q_{g,M}(\theta)$ provide an estimator that uses only the sorting patterns captured by the choices for g. Inside the integral over the random matrix B in each moment $Q_{g,M}(\theta)$, the portion of the integrand $g(A(X_m,B;\theta),X_m)$ is typically nonzero. This is unlike the integrand of the likelihood contribution in (18), which involves an indicator function that will be be 0 for an intractably large number of realizations of B for matching markets with the numbers of agents in our data. Intuitively, our moments work only with agent characteristics while the likelihood contribution (18) exploits agent indices fully to compute the exact probability of the assignment in the data. Our estimator is statistically inefficient but easier to compute. We use the usual optimal weighting matrix from two-step GMM.[38]

The integral in each moment $Q_{g,M}(\theta)$ is approximated on the computer using simulation over the random matrix of unobservable complementarities B (McFadden, 1989; Pakes and Pollard, 1989). The method of simulated moments estimator is consistent for θ as $M \to \infty$ for a fixed number of simulation draws for the matrix B. In practice, we use Halton sequences to sample B while reducing simulation error at the risk of introducing some small bias from the deterministic simulation draws. The number of draws of the entire matrix B is 1000 for biotech and 500 for medical. The integral that is simulated has a dimension equal to the number of elements in the matrix B, $(N_m^u - 1) \cdot (N_m^d - 1)$ for market m. In the estimation sample, our market size cap means that the maximum number of elements of B is 841.

The standard errors are adjusted for simulation error. The standard errors use the usual sandwich formulas with numerical derivatives approximating the actual derivatives, as the assignment outcome $A(X,B;\theta)$ is discrete and hence not differentiable in the simulation estimator (it is smoothed by the integral over the multivariate normal distribution for B in an estimator without numerical integration error). We calibrate the stepsize of the numerical derivatives to achieve somewhat decent confidence interval coverage in the Monte Carlo studies to be discussed now.

8.6 Monte Carlo Study

We conduct a Monte Carlo study at the reported point estimates to ensure that our chosen moments are informative in that they lead to low bias in the estimates of θ, as done, for example, in Eisenhauer,

---

[36]For readability, the above notation suppresses the reality that the dimension of the random matrix B depends on the numbers of entrepreneurial and venture capitalist firms in X_m.

[37]The outcome $A(X,B;\theta)$ of an assignment being pairwise stable is discrete and hence induces a non-differentiability in the simulated GMM objective function. We use a non-gradient based, global optimization routine known as a genetic algorithm to maximize the function. We vary optimization routine settings, such as the population size of points, in order to check that our estimates appear to be a global minimum.

[38]Simulation estimators have been used to estimate non-transferable utility matching games (Boyd et al., 2013; Sørensen, 2007; Agarwal, 2015; Agarwal and Diamond, 2013).

Heckman and Mosso (2015). We also use the Monte Carlo study to calibrate the stepsizes of the numerical derivatives used for the standard errors.

We perform one Monte Carlo study based on the biotech industry and a separate Monte Carlo study based on the medical industry. For each industry, the true parameters are taken from the actual point estimates described below in Table 4. Each replication proceeds as follows. We randomly sample 35 matching markets for biotech and 15 matching markets for medical; these are similar to the numbers of matching markets used in the real data estimation. Sampling a matching market means using the number of startups, the number of venture capitalists, the match characteristics, and the agent characteristics from the data for that market. We then also sample a matrix B of unobserved complementarities and compute the pairwise stable assignment. We then use the data on assignments and observable characteristics to estimate the parameter vector θ. We conduct 100 Monte Carlo replications for the biotech industry and, for computational reasons, 45 Monte Carlo replications for the medical industry.

Table 3 reports the two Monte Carlo studies, one for biotech and one for medical. For each parameter in θ, the table reports the bias, the root mean squared error (RMSE) and the coverage of the nominal 95% confidence intervals, adjusted for simulation error. RMSE is calculated as

$$\sqrt{\frac{1}{I}\sum_{i=1}^{I}\left(\hat{\theta}_i^{\mathrm{syn}} - \theta^{\mathrm{true}}\right)^2},$$

where $\hat{\theta}_i^{\mathrm{syn}}$ is the estimator using synthetic data for the ith out of I Monte Carlo replications. Table 3 shows that our choice of moments do lead to relatively low bias and RMSE; the absolute value of bias is always less than or equal to 0.05 for all parameter values.[39] For matching markets with much larger numbers of firms, the same number of simulation draws and the same number of markets in the data, unreported Monte Carlo studies indicate that the finite-sample bias from simulation error will be more substantial. Table 3 shows that the coverage of the nominal 95% confidence intervals are above 90% for all but one parameter for the biotech industry and two parameters for the medical industry. The RMSEs are small so we are not worried about the undercoverage on these three parameters leading to falsely rejecting the hypothesis that a parameter is zero.

8.7 Estimates

Table 4 reports the estimates and standard errors. There are four separate sets of estimates: two for the biotech industry and two for the medical industry. For each industry, we report estimates without and with the term interacting startup patents and venture capitalist total experience. The

---

[39]As discussed in the section on the point estimates, the economic magnitude of Monte Carlo's true value on the interaction term between total experience and patents is close to zero (0.02) for the biotech industry. Therefore, the bias of 0.01 is also small in economic magnitude even it is large compared to the true parameter in the Monte Carlo.

Table 3: Monte Carlo Study

	Biotech				Medical			
	True	Bias	RMSE	Cover.	True	Bias	RMSE	Cover.
Sector experience	1.28	0.04	0.15	0.98	0.31	0.03	0.10	0.80
Total experience * log(patents+1)	0.02	0.01	0.06	0.88	0.67	0.05	0.18	0.84
Standard deviation of UC	2.08	0.01	0.03	0.94	1.96	0.02	0.04	1.00
Correlation, no common firm	0.17	0.00	0.02	0.94	0.74	-0.02	0.04	0.93
Correlation, common startup	0.23	0.01	0.02	0.94	0.74	0.00	0.04	0.98
Correlation, common VC	0.94	-0.01	0.02	0.91	0.77	-0.02	0.04	0.96

Both studies uniformly sample matching markets with replacement from the corresponding real-data matching markets.

Biotech: sample size of 38 markets, 100 Monte Carlo replications.

Medical: sample size of 15 markets, 45 Monte Carlo replications

Measure	Biotech		Medical	
Distance	-1	-1	-1	-1
	(-)	(-)	(-)	(-)
Sector experience	1.32	1.28	0.22	0.31
	(0.30)	(0.20)	(0.03)	(0.06)
Total experience * log(patents+1)		0.02		0.67
		(0.02)		(0.11)
Standard deviation of	2.52	2.08	2.36	1.96
unobserved complementarities	(0.19)	(0.04)	(0.17)	(0.10)
Correlation	0.41	0.17	0.66	0.74
no common firms	(0.06)	(0.02)	(0.13)	(0.06)
Correlation	0.45	0.23	0.66	0.74
same startup	(0.06)	(0.05)	(0.12)	(0.06)
Correlation	0.96	0.94	0.79	0.77
same venture capitalist	(0.38)	(0.01)	(0.08)	(0.04)

Table 4: Venture Capital Parameter Estimates

parameters on the match-specific characteristics are not overly sensitive to the inclusion of the agent-specific characteristics, so we focus on the estimates with the interaction term. The top half of Table 4 reports the estimates of the production function parameters and the bottom half reports the estimates of the parameters of the multivariate normal distribution of unobserved complementarities.

We first consider the production function parameter estimates for the biotech industry in Table 4. Both distance and sector experience are match characteristics that have been normalized to have a standard deviation of around 1. The point estimate of 1.28 on VC experience in the four-digit sector of the startup indicates that sector experience is more important than distance, although the confidence interval for the sector experience parameter contains the absolute value of the normalized coefficient on distance, 1. The coefficient on the interaction between startup patents and venture capitalist experience is 0.02. A change in the number of patents from 1 to 2 is equivalent to a 0.18

change in the rescaled log of patents measure. At say a value of the rescaled VC total experience of 0.5, this change in patents results in a change in match production of $0.02 \cdot 0.18 \cdot 0.5 = 0.002$ units of the standard deviation of distance, an economically tiny effect. Another way to interpret the coefficient on the interaction term is to look at the standard deviation in the value of the interaction across realized matches, which is 1.1 in Table 2. An one standard deviation increase in the interaction term then results in a $0.02 \cdot 1.1 = 0.02$ change in production in units of the standard deviation of distance. Patents and total experience do not seem to play a role in match production in biotech.

Now consider the production function parameter estimates for the medical industry in Table 4. The parameter of 0.31 on sector experience, with a low standard error, is economically and statistically lower than the normalized coefficient of 1 on distance. So distance is more important than sector experience in the medical industry. The coefficient on the interaction between patents and VC total experience is 0.67, 34 times the coefficient of 0.02 in biotech. Still, the change from 1 to 2 patents at a rescaled total VC experience of 0.5 results in a change of production of 0.06 distance standard deviations, still not a large effect in economic magnitudes. Table 2 reports the standard deviation of the interaction term across realized matches is 1.6. A one standard deviation change in the interaction term then results in production increasing by $0.67 \cdot 1.6 = 1.1$ distance standard deviations. Interpreted using the standard deviation of the interaction term across realized matches, the effect of the interaction term is about the same as distance and greater than the effect of sector experience.

Next, we interpret the parameters of the multivariate normal distribution of unobserved complementarities. Recall the definition in (6): an unobserved complementarity is indexed by four firms and is the sum and difference of unobserved match characteristics involving four matches. For the unobserved complementarities $b_{u,d}$ from (8), two of the four firms are always fixed at upstream firm 1 and downstream firm 1, although firm indices have no meaning across markets in our analysis. The other two firms, the u and d in $b_{u,d}$, vary across the unobserved complementarities in the matrix B. Interpreting the distribution of unobserved complementarities requires subtlety; it is easier to think in terms of unobserved match characteristics. Unfortunately, the distribution of unobserved match characteristics is not identified.

Consider the standard deviations of the unobserved complementarities. For biotech, the standard deviation is 2.08, with a small standard error. The standard deviation can be interpreted to mean, in a loose sense, that unobserved complementarities are twice as important as distance, which has a coefficient of -1 and a standard deviation of 1. In the same loose sense, the point estimate is that unobserved complementarities are $2.08/1.28 = 1.6$ times as important as sector experience. For the medical industry, the standard deviation of unobserved complementarities is 1.96, meaning that, again in a loose sense, that unobserved complementarities are twice as important as distance and $2.08/0.31 = 6.3$ times as important as sector experience.

There are three correlation parameters in the multivariate normal distribution. The first correlation is $\text{Corr}(b_{u_1,d_1}, b_{u_2,d_2})$, the correlation between two unobserved complementarities when neither of

37

the upstream firms or downstream firms (other than upstream firm 1 and downstream firm 1) are the same. For biotech, this correlation is positive and low, at 0.17, with a small standard error. Presumably the positive correlation reflects the presence of upstream firm 1 and downstream firm 1 in all unobserved complementarities $b_{u,d}$. For the medical industry, the correlation is much higher, at 0.74, also with a low standard error. When upstream firm 1 and downstream firm 1 contribute to unobserved complementarities in one set of four matches, they do so in others.

The second correlation is Corr $(b_{u_1,d_1}, b_{u_2,d_1})$, the correlation between two unobserved complementarities when the startup is the same. The correlation for biotech of 0.23 has the same sign but is lower than the correlation for the medical industry of 0.74. The third correlation is Corr $(b_{u_1,d_1}, b_{u_1,d_2})$, the correlation between two unobserved complementarities when the venture capitalist is the same. For biotech, this correlation of 0.94 is quite high. For the medical industry, the correlation of 0.77 is a little lower. For biotech, the correlation between two complementarities involving upstream firm 1, downstream firm 1 and the same venture capitalist is much higher than the correlation between complementarities involving upstream firm 1, downstream firm 1 and the same startup as well as the correlation when only upstream firm 1 and downstream firm 1 are the same. This seems to mean that venture capitalists are playing more of systematic role in unobserved complementarities across a variety of sets of four matches than the more idiosyncratic role of startups. In the medical industry, the three correlations are all high and about the same. The distribution of B is close to being equicorrelated and hence exchangeable in all the elements of B, not just exchangeable in the agent indices (which is imposed by our choice of distribution).

8.8 Discussion of Unobserved Complementarities

The somewhat subtle interpretation of the estimated distribution of unobserved complementarities showcases the loss of information from using data on only matched agents, which is common in empirical work on matching. Nevertheless, the distribution of unobserved complementarities is in the same units as the contribution of observables to the production function and the two can be compared. For both the biotech and medical industries, we found that the standard deviation of unobserved complementarities was greater than the individual contributions of distance, VC sector experience, startup patents and VC total experience. However, combining the contributions of all observables in the production function, particularly sector experience and distance, suggests that the role of unobserved complementarities is roughly the same as the role of all observables.

Reporting the standard deviation of the unobserved complementarities has some analogs to reporting the standard deviation of error terms in other empirical literatures. For example, the standard deviation of wage regression residuals is often thought of as representing the dispersion in unobserved worker ability. The standard deviation of production function residuals is typically called the dispersion in total factor productivity. There are thousands of papers on understanding the often un-

measured factors that affect worker ability and firm total factor productivity. Likewise, our estimates of large standard deviations of unobserved complementarities in venture capital suggest that there is good motivation for additional academic research on the factors making venture capital investments more productive.

9 Conclusion

Matching models that have been structurally estimated to date have not allowed rich distributions of unobservables. It has been an open question whether data on who matches with whom as well as match or agent characteristics are enough to identify such distributions of unobservables. In this paper, we explore several sets of conditions that lead to identification.

Using data on only matched firms, one can identify distributions of what we call unobserved complementarities but not the underlying primitive distribution of match-specific (or agent-specific) unobservables. The distribution of complementarities is enough to compute assignment production levels and therefore counterfactual assignment probabilities. In extensions, we can include other covariates X and identify distributions of unobservable complementarities conditional on X. We show that it is possible to identify heterogeneous-within-a-market coefficients on the large support match characteristics. The results extend naturally to two-sided, many-to-many matching.

If the data contain unmatched firms, the individual rationality decision to not be unmatched helps identify the distribution of primitively specified unobserved match characteristics, not just the distribution of unobserved complementarities. We extend this result to the fairly general case of matching with trades, as in Hatfield et al. (2015).

Our empirical work studies biotech and medical venture capital. We estimate the degree to which venture capitalists change the production of matches. Among many other empirical results, we find that the standard deviation of unobserved complementarities is roughly of the same order of magnitude as the contribution to production from the observables match and agent characteristics.

A Proofs

A.1 Lemma 1

Fix a realization E^* of the primitive unobservable, E. Using the elements of E^* and the large support on Z, set $z_{u,d}^\star = -e_{u,d}^\star$. Then $s(A; E^\star, Z^\star) = \sum_{\langle u,d \rangle \in A} \left(e_{u,d}^\star + z_{u,d}^\star \right) = 0$ for all assignments A.

The definition of the joint CDF $H\left(\tilde{S}\right)$ at some vector of evaluation $\tilde{S}^\star$ formed from $E^\star$ is

$$H\left(\tilde{S}^\star\right) = \mathrm{Pr}_E\left(\tilde{s}(A; E) \leq \tilde{s}(A; E^\star), \forall A \neq A_1\right).$$

Here, each element of the vector $\tilde{S}^\star$ is $\tilde{s}(A; E^\star)$ for some $A \neq A_1$. Identification of H follows from

$$
\begin{aligned}
H\left(\tilde{S}^\star\right) &= \Pr_E\left(\tilde{s}(A; E) \leq \tilde{s}(A; E^\star), \forall A \neq A_1\right) \\
&= \Pr_E\left(s(A; E, Z^\star) - s(A_1; E, Z^\star) \leq s(A; E^\star, Z^\star) - s(A_1; E^\star, Z^\star), \forall A \neq A_1\right) \\
&= \Pr_E\left(s(A; E, Z^\star) - s(A_1; E, Z^\star) \leq 0, \forall A \neq A_1\right) \\
&= \Pr_E\left(s(A; E, Z^\star) \leq s(A_1; E, Z^\star), \forall A \neq A_1\right) \\
&= \Pr\left(A_1 \mid Z^\star\right).
\end{aligned}
$$

Here the first line is the definition of the joint CDF, the second line adds the observed production of assignments A and A_1 to both sides of the inequality, the third line uses $s(A; E^\star, Z^\star) = 0 \, \forall A$, the fourth line moves $s(A_1, E, Z^\star)$ to the right side of the inequality for each A, and the fifth lines uses the fact that assignment A_1 is pairwise stable whenever A_1 has a higher total production than all other assignments A.

A.2   Lemma 2

A.2.1   First part of Lemma 2

For the first part of the lemma, we need to show that every element in C is a linear combination of elements in B. Note that any unobserved complementarity of the form c_{1,d_1,u,d_2} is equal to the difference of two elements of B:

$$
\begin{aligned}
c_{1,d_1,u,d_2} &= e_{1,d_1} + e_{u,d_2} - (e_{1,d_2} + e_{u,d_1}) \\
&= e_{1,1} + e_{u,d_2} - (e_{1,d_2} + e_{u,1}) - (e_{1,1} + e_{u,d_1} - (e_{1,d_1} + e_{u,1})) \\
&= b_{u,d_2} - b_{u,d_1}.
\end{aligned}
$$

Next, we represent an arbitrary unobserved complementarity c_{u_1,d_1,u_2,d_2} in terms of unobserved complementarities of the form c_{1,d_1,u,d_2}.

$$
\begin{aligned}
c_{u_1,d_1,u_2,d_2} &= e_{u_1,d_1} + e_{u_2,d_2} - (e_{u_1,d_2} + e_{u_2,d_1}) \\
&= e_{1,d_1} + e_{u_2,d_2} - (e_{1,d_2} + e_{u_2,d_1}) - (e_{1,d_1} + e_{u_1,d_2} - (e_{1,d_2} + e_{u_1,d_1})) \\
&= c_{1,d_1,u_2,d_2} - c_{1,d_1,u_1,d_2}.
\end{aligned}
$$

Because we have shown that any unobserved complementarity of the form c_{1,d_1,u,d_2} is a difference of two elements in B, c_{u_1,d_1,u_2,d_2} can be written as the sums and differences of elements in B.

A.2.2 Second Part of Lemma 2

For the second part of the lemma, we are given a $F(B)$ and need to find a $G(E)$ such that G generates F by the change of variables given by the definition of the unobserved complementarities in B, (8). Note that every element b_{u_2,d_2} of B contains a unique element e_{u_2,d_2} of E. Place a distribution G on E's such that $e_{1,d} = e_{u,1} = 0 \, \forall u, d$ and the other elements of each E are such that $e_{u,d} = b_{u,d}$ for some B in the support of $F(B)$. If each E of $G(E)$ has the same frequency as the paired B in $F(B)$, the distribution $G(E)$ generates $F(B)$.

A.2.3 Third Part of Lemma 2

Let π_u be permutation of upstream firm indices and π_d a permutation of downstream firm indices. Define the random matrix E to be exchangeable in agent indices if the distribution of

$$
E_{\pi_u,\pi_d} = \begin{pmatrix} e_{\pi_u(1),\pi_d(1)} & \cdots & e_{\pi_u(1),\pi_d(N)} \\ \vdots & \ddots & \vdots \\ e_{\pi_u(N),\pi_d(1)} & \cdots & e_{\pi_u(N),\pi_d(N)} \end{pmatrix}
$$

is the same as E for all π_u and π_d. Similarly, B is exchangeable in agent indices if

$$
B_{\pi_u,\pi_d} = \begin{pmatrix} b_{\pi_u(2),\pi_d(2)} & \cdots & b_{\pi_u(2),\pi_d(N)} \\ \vdots & \ddots & \vdots \\ b_{\pi_u(N),\pi_d(2)} & \cdots & b_{\pi_u(N),\pi_d(N)} \end{pmatrix}
$$

has the same distribution as B for all π_u and π_d.

We wish to prove that B is exchangeable in agent indices if E is exchangeable in agent indices. B is formed from E by a linear transformation D, representing the formula (8) for each element of B. Dean and Verducci (1990, Condition 2, Theorem 4) provide a sufficient (and necessary) condition for a linear transformation to preserve exchangeability in *all elements* of a random vector. If we vectorize the matrices B and E, the argument in the first paragraph of the proof of Theorem 4 of Dean and Verducci can be reproduced for our definition of exchangeability in *agent indices*. We skip this step of reproducing one direction of the proof of Theorem 4 of Dean and Verducci for our different notion of exchangeability in agent indices for conciseness.

The sufficiency condition from Dean and Verducci that we need to verify is that for any permutation in agent indices of B, there exists a permutation of agent indices in E that gives B_{π_u,π_d} through the linear transformation D. This condition is satisfied for B and E. Given permutations of agent indices π_u and π_d themselves giving B_{π_u,π_d}, the same permutations of agent indices give E_{π_u,π_d}. It is clear that B_{π_u,π_d} is related to E_{π_u,π_d} though the linear transformation D by inspection of (8).

A.3 Theorem 1

A.3.1 First Part of Theorem 1

Using the definition of $\tilde{r}(A; B)$ gives

$$
\begin{aligned}
\tilde{r}(A; B) &= \sum_{\langle u,d\rangle \in A} b_{u,d} - \sum_{\langle u,d\rangle \in A_1} b_{u,d} \\
&= \sum_{\langle u,d\rangle \in A} (e_{1,1} + e_{u,d} - (e_{1,d} + e_{u,1})) - \sum_{\langle u,d\rangle \in A_1} (e_{1,1} + e_{u,d} - (e_{1,d} + e_{u,1})) \\
&= \sum_{\langle u,d\rangle \in A} e_{u,d} - \sum_{\langle u,d\rangle \in A_1} e_{u,d} - \sum_{\langle u,d\rangle \in A} (e_{1,d} + e_{u,1}) + \sum_{\langle u,d\rangle \in A_1} (e_{1,d} + e_{u,1}) \\
&= \sum_{\langle u,d\rangle \in A} e_{u,d} - \sum_{\langle u,d\rangle \in A_1} e_{u,d} \\
&= \tilde{s}(A; E),
\end{aligned}
$$

where the fourth inequality uses the fact that each firm is matched the same number of times (in one-to-one matching, exactly once) in both the assignments A and A_1 and the last inequality is just the definition of $\tilde{s}(A; E)$ in (2).

A.3.2 Second Part of Theorem 1

If $B_1 = B_2$, then $\tilde{r}(A; B_1) = \tilde{r}(A; B_2)$ simply because (10) is a definition of a function of B. For the other direction, assume $\tilde{r}(A; B_1) = \tilde{r}(A; B_2)$ for all A. Focus on a particular scalar unobserved complementarity $b_{u,d}$ in B. We will show that $b_{u,d}$ can be written as $\tilde{s}(A_2, E) - \tilde{s}(A_3, E)$ for particular assignments A_2 and A_3. As the first part of the theorem is that $\tilde{s}(A; E) = \tilde{r}(A; B)$ for any A and where B is formed from E, this implies that $b_{u,d}$ is the same in B_1 and B_2. Because $b_{u,d}$ was arbitrary, $B_1 = B_2$.

Let A_2 be an assignment that contains the matches $\langle u, d\rangle$ and $\langle 1, 1\rangle$. Let A_3 be the same assignment as A_2 except that A_3 includes the matches $\langle 1, d\rangle$ and $\langle u, 1\rangle$ and does not include $\langle u, d\rangle$ and $\langle 1, 1\rangle$. Then

$$
\tilde{s}(A_2, E) - \tilde{s}(A_3, E) = e_{1,1} + e_{u,d} - (e_{u,1} + e_{1,d}) = b_{u,d}.
$$

By the above argument and because the match $\langle u, d\rangle$ was arbitrary, $B_1 = B_2$.

A.3.3 Third Part of Theorem 1

$H\left(\tilde{S}\right)$ is identified from Lemma 1. The first part of Theorem 1 shows that the change of variables from $\tilde{S}$ to the vector of all $\tilde{r}$ is one-to-one. The second part of Theorem 1 shows that the change of variables from the vector of all $\tilde{r}$ to the matrix of unobserved complementarities B is one-to-one.

Therefore, $F(B)$ is identified.

A.4   Proof of Theorem 2

If Γ is identified, then following a slightly modified version of the proof of Lemma 1 and the same argument as the proof of Theorem 1.3 demonstrate that F is also identified. So consider identifying Γ. Recall that the scale normalization is that $\gamma_{1,1} = \pm 1$. We can easily identify the sign of $\gamma_{1,1}$. Consider some assignment A that includes match $\langle 1, 1 \rangle$. Then we can compare Z_1 and Z_2 that differ only in the value of $z_{1,1}$: $z_{1,1}^1 > z_{1,1}^2$. If $\Pr(A \mid Z_1) > \Pr(A \mid Z_2)$, we conclude that $\gamma_{1,1} = +1$ and if $\Pr(A \mid Z_1) < \Pr(A \mid Z_2)$ we conclude that $\gamma_{1,1} = -1$. The main text rules out the case where any $\gamma_{u,d} = 0$. In what follows, we focus on the case where the sign of every $\gamma_{u,d}$ is identified and, in particular, $\gamma_{1,1} = +1$. The case of $\gamma_{1,1} = -1$ is symmetric.

We now show how to identify the arbitrary parameter $\gamma_{\tilde{u}, \tilde{d}}$. Consider assignments $A^1 = \left\{ \langle 1, 1 \rangle, \left\langle \tilde{u}, \tilde{d} \right\rangle, \ldots \right\}$ and $A^2 = \left\{ \left\langle 1, \tilde{d} \right\rangle, \langle \tilde{u}, 1 \rangle, \ldots \right\}$ that are identical except for the explicitly listed matches. In a proof shortcut borrowing an idea from identification at infinity, let the matches not in $A^1 \cup A^2$ correspond to $z_{u,d}$'s where $\gamma_{u,d} z_{u,d} = -\infty$, so we consider only Z's where the total production of any assignment other than A^1 and A^2 is $-\infty$ and hence $\Pr\left(A^1 \mid Z\right) + \Pr\left(A^2 \mid Z\right) = 1$. Set $z_{1,\tilde{d}} = 0$ and $z_{\tilde{u},1} = 0$. Then A^1 occurs whenever $z_{1,1} + e_{1,1} + \gamma_{\tilde{u},\tilde{d}} z_{\tilde{u},\tilde{d}} + e_{\tilde{u},\tilde{d}} \geq e_{1,\tilde{d}} + e_{\tilde{u},1}$, or by (8), $z_{1,1} + \gamma_{\tilde{u},\tilde{d}} z_{\tilde{u},\tilde{d}} + b_{\tilde{u},\tilde{d}} \geq 0$. This decision rule is equivalent to a decision rule in a single agent binary choice model. As $b_{\tilde{u},\tilde{d}}$ is fully independent from $z_{1,1}$ and $z_{\tilde{u},\tilde{d}}$, we can apply the results on binary choice from Manski (1988) under full independence and identify $\gamma_{\tilde{u},\tilde{d}}$.

A.5   Proof of Theorem 3

Condition on X. Let A_0 denote the assignment where no agents are matched, then $\tilde{s}(A_0; E) = 0$ for all E. Let $E^\star$ be an arbitrary realization of the matrix of match-specific unobservables. Let $Z^\star = \left(z_{u,d}^\star \right)_{u,d \in N}$ be such that $z_{u,d}^\star = -e_{u,d}^\star$. Then $s(A; Z^\star, E^\star) = 0$ for all A and $S(A_0; Z^\star, E) = 0$ for any E. Thus for all A and all $E \leq E^\star$ elementwise, $S(A; Z, E) \leq 0 = S(A_0; Z^\star, E)$. Further, if any element of E is greater than the corresponding element of $E^\star$, assignment A_0 will not maximize $s(A, Z^\star, E)$ and so A_0 will not be pairwise stable. Therefore $G(E^\star) = \Pr(E \leq E^* \text{ elementwise} \mid E^\star) = \Pr(A_0 \mid Z^\star)$.

A.6   Proof of Corollary 2

Condition all arguments on X.

We first argue that the equivalent of Lemma 1 holds. Fix a realization $\left((e_u^\star)_{u \in N}, (e_d^\star)_{d \in N} \right)$ of the agent-specific unobservables. Using the elements of $\left((e_u^\star)_{u \in N}, (e_d^\star)_{d \in N} \right)$ and the large support on $\left((z_u)_{u \in N}, (z_d)_{d \in N} \right)$, set $z_u^\star = -e_u^\star$ and $z_d^\star = e_d^\star$, the latter without a negative sign as the multiplication

of two negatives is positive. Then $s\left(A;E^\star,Z^\star\right)=\sum_{\langle u,d\rangle\in A}\left(e_u^\star\cdot e_d^\star+z_u^\star\cdot z_d^\star\right)=0$ for all assignments A. The rest of the proof is then identical to the corresponding portion of the proof of Lemma 1, for the match-specific case.

The first two parts of Theorem 1 do not refer to Z at all and do not impose any restrictions on the E matrix. So they automatically apply to the less general case where $e_{u,d}=e_u\cdot e_d$. The main identification result, the third part of Theorem 1, then follows for the agent-specific case from the slight modification to the proof of Lemma 1 above and the first two parts of Theorem 1.

A.7 Proof of Theorem 4

Condition on X. Let A_0 denote the assignment (of trades) where no trades are made, then the sum of unobservables for this assignment is 0 for all E and Z. Let $E^\star$ be an arbitrary realization of E, the vector of trade-specific unobservables. Let $Z^\star$ be such that $z_\omega^\star=-e_\omega^\star$ for all $\omega\in\Omega$. Define

$$s\left(A;E,Z\right)=\sum_{\omega\in A}\left(e_\omega+z_\omega\right)$$

to be the total production from an assignment. Then $s\left(A;Z^\star,E^\star\right)=0$ for all A and $S\left(A_0;Z^\star,E\right)=0$ for any E. Therefore for all A and all $E\le E^\star$ elementwise, $S\left(A;Z,E\right)\le 0=S\left(A_0;Z^\star,E\right)$. Therefore, $G\left(E^\star\right)=\Pr\left(E\le E^*\text{ elementwise}\mid E^*\right)=\Pr\left(A_0\mid Z^\star\right)$.

A.8 Proof of Theorem 5

Condition on X. Let A_0 denote the assignment (of trades) where no trades are made, then the sum of unobservables for this assignment is 0 for all E and Z. Let $E^\star$ be an arbitrary realization of the array of match-specific unobservables. Let $Z^\star=\left(z_{i,\psi_i}^\star\right)_{i\in N,\psi_i\subseteq\Omega_i}$ be such that $z_{i,\psi_i}^\star=-e_{i,\psi_i}^\star\ \forall\,i\in N,\psi_i\subseteq\Omega_i$. Define

$$s\left(A,E,Z\right)=\sum_{i\in N}\left(e_{i,\psi_i^A}+z_{i,\psi_i^A}\right),$$

where ψ_i^A are the trades of i in A, to be the total profit or production from an assignment. Then $s\left(A;Z^\star,E^\star\right)=0$ for all A and $S\left(A_0;Z^\star,E\right)=0$ for any E. Therefore for all A and all $E\le E^\star$ elementwise, $S\left(A;Z,E\right)\le 0=S\left(A_0;Z^\star,E\right)$. Therefore assignment A_0 will occur whenever $E\le E^\star$ elementwise. Can assignment A_0 occur for E not less than or equal (elementwise) to $E^\star$? For such an E, there is at least one (i,ψ_i) such that $e_{i,\psi_i}>e_{i,\psi}^\star$. In this case, the valuation $u\left(i,\Psi_i\right)$ of i for ψ_i at $Z^\star$, $e_{i,\psi_i}+z_{i,\psi}^\star=e_{i,\psi_i}-e_{i,\psi_i}^\star$ is positive. However, it could still be that at this E and $Z^\star$ a vector of prices for trades cannot be formed so that an assignment of trades other than A_0 is pairwise stable. So A_0 can occur at realizations of unobservables E not less than or equal (elementwise) to $E^\star$. Therefore, $G\left(E^\star\right)=\Pr\left(E\le E^*\text{ elementwise}\mid E^*\right)\le\Pr\left(A_0\mid Z^\star\right)$. Define $\bar{G}\left(E\mid X\right)=\Pr\left(A_0\mid Z^\star\right)$.

B   Inputs to Moments

We describe the functions g that determine the empirical moments in (19).

The market m subscript is omitted to reduce notation; firm characteristics, numbers of upstream and downstream firms and matching outcomes are all market specific. Let the upstream firm characteristics be $\boldsymbol{x}_u = (x_u(1), \ldots, x_u(K))$ and downstream $\boldsymbol{x}_d = (x_d(1), \ldots, x_d(K))$. K is the number of firm specific characteristics for any firm. In principle K could vary for upstream and downstream firms although it does not in our venture capital application. Let the match specific characteristics be $\boldsymbol{x}_{u,d} = (x_{u,d}(1), \ldots, x_{u,d}(K^{u,d}))$, where $K^{u,d}$ is the number of match specific characteristics. Let an assignment A induce a function of downstream firm indices: $A(d) = u$ if u is matched with d. The function is well defined because our venture capital application is to many-to-one matching and we do not have data on unmatched firms.

1. We use the following functions g to construct moments if firm specific characteristics are included in the match production function.

    (a) For firm characteristic pair k, consider the vector

$$L^1(k) = \left(x_{u=A(1)}x_{d=1}, \ldots, x_{u=A(N_D)}x_{d=N_D}\right).$$

   The pth quantile of $L^1(k)$ is $q(L(k), p)$, where p is a number between 0 and 1. The sample mean and variance of $L^1(k)$ are $\mu^1(k)$ and $\text{Var}^1(k)$. Define the functions

$$g^1_{k,p} = \frac{q(L^1(k), p)}{1 + \max(|L^1(k)|)}, \quad g^2_{k,p} = \frac{q\left(\left(L^1(k) - \mu^1(k)\right)^2, p\right)}{1 + \text{Var}^1(k)}.$$

    (b) Consider the vector

$$L^2(k_1, k_2) = \left(x_{u=A(1)}(k_1) x_{d=1}(k_1) \sum_{u \neq A(1)} x_u(k_2) x_{d=1}(k_2), \right.$$
$$\left. \ldots, x_{u=A(N^d)}(k_1) x_{d=N^d}(k_1) \sum_{u \neq A(N^d)} x_u(k_2) x_{d=N^d}(k_2) \right).$$

   Let the corresponding sample mean and variance be $\mu^2(k_1, k_2)$ and $\text{Var}^2(k_1, k_2)$. Define the functions

$$g^3_{k_1,k_2,p} = \frac{q(L^2(k_1, k_2), p)}{1 + \max(|L^2(k_1, k_2)|)}, \quad g^4_{k_1,k_2,p} = \frac{q\left(\left(L^2(k_1, k_2) - \mu^2(k_1, k_2)\right)^2, p\right)}{1 + \text{Var}^2(k_1, k_2)}.$$

(c) Let the sample mean of $\left(\sum_{u \neq A(1)} x_u(k_2) x_{d=1}(k_2), \ldots, \sum_{u \neq A(N_D)} x_u(k_2) x_{d=N_D}(k_2)\right)$ be $\mu^3(k_2)$. Define the moment

$$g^5_{k_1, k_2} = \frac{\mu^2(k_1, k_2) - \mu^1(k_1) \mu^3(k_2)}{1 + Var^1(k_1)}.$$

(d) We next define functions of market-specific regression coefficients. Note that the assignment A can also be interpreted as an N^u by N^d binary matrix. Reshape the matrix A into a column vector A^c while preserving the columnwise ordering (like the MATLAB reshape command). Also construct the following matrices

$$X^u(k) = \begin{pmatrix} \overbrace{\begin{matrix} x_{u=1}(k), & \ldots, & x_{u=1}(k) \\ \vdots & & \vdots \\ x_{u=N_U}(k), & \ldots, & x_{u=N_U}(k) \end{matrix}}^{N^d} \end{pmatrix}, X^d(k) = \begin{pmatrix} N^u \left\{ \begin{matrix} x_{d=1}(k), & \ldots, & x_{d=N_D}(k) \\ \vdots & & \vdots \\ x_{d=1}(k), & \ldots, & x_{d=N_D}(k) \end{matrix} \right. \end{pmatrix},$$

$$X^{u \cdot d}(k) = \begin{pmatrix} x_{u=1}(k) x_{d=1}(k), & \ldots, & x_{u=1}(k) x_{d=N_D}(k) \\ \vdots & & \vdots \\ x_{u=N_U}(k) x_{d=1}(k), & \ldots, & x_{u=N_U}(k) x_{d=N_D}(k) \end{pmatrix}.$$

Next similarly reshape the matrices to the column vectors $X^{u,c}(k), X^{d,c}(k), X^{u \cdot d,c}(k)$. Define

$$X^c = \left[\mathbf{1}, \left(X^{u,c}(1), \ldots, X^{u,c}(K)\right), \left(X^{d,c}(1), \ldots, X^{d,c}(K)\right), \left(X^{u \cdot d,c}(1), \ldots, X^{u \cdot d,c}(K)\right)\right].$$

We use the regression coefficients

$$g^6 = \left(\text{transpose}\left(X^c\right) X^c\right)^{-1} \text{transpose}\left(X^c\right) A^c$$

as our last set of functions involving the firm specific characteristics.

2. We use the following functions to construct moments if match specific characteristics are available.

(a) Previously we defined the functions g^1 through g^5; in those definitions agent characteristics always enter as the product $x_u \cdot x_d$. Replacing products $x_u \cdot x_d$ with true match characteristics $x_{u,d}$, we can define g^7 through g^{11} like we did for g^1 through g^5.

46

(b) We also use the additional set of functions

$$g^{12}_{k_1,k_2,p} = \frac{q\left(\left|L^2\left(k_1,k_2\right) - \mu^1\left(k_1\right)\mu^3\left(k_2\right)\right|, p\right)}{1 + \text{Var}^1\left(k_1\right)},$$

where the means and variances are analogous to those in the section above, replacing $x_u \cdot x_d$ with $x_{u,d}$.

(c) To define an analog to g^6, we use the following "regressors". Define the matrices $X^{u,d}\left(k\right)$ with u,d elements $x_{u,d}\left(k\right)$, $\tilde{X}^u\left(k\right)$, where u,d elements $\sum_{u' \neq u} x_{u',d}\left(k\right)$, and $\tilde{X}^d\left(k\right)$ with u,d elements $\sum_{d' \neq d} x_{u,d'}\left(k\right)$. Reshape the matrices into column vectors, and define

$$\tilde{X}^c = \left[\mathbf{1}, \left(X^{u,d,c}\left(1\right),\ldots,X^{u,d,c}\left(S\right)\right), \left(\tilde{X}^{u,c}\left(1\right),\ldots,\tilde{X}^{u,c}\left(K^{u,d}\right)\right), \left(\tilde{X}^{d,c}\left(1\right),\ldots,\tilde{X}^c_D\left(K^{u,d}\right)\right)\right].$$

The functions corresponding with g^6 are defined as

$$g^{13} = \left(\text{transpose}\left(\tilde{X}^c\right)\tilde{X}^c\right)^{-1}\text{transpose}\left(\tilde{X}^c\right)A^c.$$

3. When both firm and match specific characteristics are present, we interact the firm and match characteristics.

$$
\begin{aligned}
L^u\left(k\right) &= \left(x_{u=A(1)}\left(k\right),\ldots,x_{u=A(N_d)}\left(k\right)\right) \\
L^d\left(k\right) &= \left(x_{d=1}\left(k\right),\ldots,x_{d=N_d}\left(k\right)\right) \\
L^{u,d}\left(k\right) &= \left(x_{u=A(1),d=1}\left(k\right),\ldots,x_{u=A(N_d),d=N_d}\left(k\right)\right),
\end{aligned}
$$

and with corresponding means and variances $\mu^u\left(k\right), \mu^d\left(k\right), \mu^{u,d}\left(k\right)$ and $\text{Var}^u\left(k\right), \text{Var}^d\left(k\right), \text{Var}^{u,d}\left(k\right)$. Define the functions

$$g^{14}_{k_1,k_2,p} = \frac{q\left(L^u\left(k_1\right) * L^{u,d}\left(k_2\right), p\right)}{1 + \max\left(\left|L^u\left(k_1\right)\right|\right)}, g^{15}_{k_1,k_2,p} = \frac{q\left(\left(L^u\left(k_1\right) * L^{u,d}\left(k_2\right) - \mu^u\left(k_1\right)\mu^{u,d}\left(k_2\right)\right)^2, p\right)}{1 + \text{Var}^u\left(k_1\right)}$$

and

$$g^{16}_{k_1,k_2,p} = \frac{q\left(L^d\left(k_1\right) * L^{u,d}\left(k_2\right), p\right)}{1 + \max\left(\left|L^d\left(k_1\right)\right|\right)}, g^{17}_{k_1,k_2,p} = \frac{q\left(\left(L^d\left(k_1\right) * L^{u,d}\left(k_2\right) - \mu^d\left(k_1\right)\mu^{u,d}\left(k_2\right)\right)^2, p\right)}{1 + \text{Var}^d\left(k_1\right)},$$

where the $*$ operator for two vectors represents element-wise multiplication. Construct the

following matrices:

$$\tilde{\tilde{X}}^c = \left[\mathbf{1}, \left(X^{u,c}\left(1\right), \dots, X^{u,c}\left(K\right) \right), \left(X^{d,c}\left(1\right), \dots, X^{d,c}\left(K\right) \right), \left(X^{u \cdot d,c}\left(1\right), \dots, X^{u \cdot d,c}\left(K\right) \right) \right.$$
$$\left. \left(X^{u,d,c}\left(1\right), \dots, X^{u,d,c}\left(K^{u,d}\right) \right) \right].$$

The regression-based functions are

$$g^{18} = \left(\text{transpose}\left(\tilde{\tilde{X}}^c \right) \tilde{\tilde{X}}^c \right)^{-1} \text{transpose}\left(\tilde{\tilde{X}}^c \right) A^c.$$

The numbers of firm and match characteristics, K and $K^{u,d}$, and the number of quantiles p affect the number of moments. We use quantiles $(0.25, 0.5, 0.75)$ to construct the functions above for the biotech sector, where market sizes range between 2×2 to 9×11. For $K = 1$ and $K^{u,d} = 2$, the number of moments is 110. We use $(0.1, 0.3, 0.5, 0.7, 0.9)$ for quantiles for the medical sector, where market sizes are between 17×17 to 27×29. For $K = 1$ and $K^{u,d} = 2$, the number of moments is 166.

References

Ackerberg, Daniel A. and Maristella Botticini, "Endogenous Matching and the Empirical Determinants of Contract Form," *Journal of Political Economy*, 2002, *110* (2002), 564–591.

Agarwal, Nikhil, "An Empirical Model of the Medical Match," *American Economic Review*, 2015, *105* (7), 1939–1978.

_ **and William Diamond**, "Identification and Estimation in Two-Sided Matching Markets," 2013. working paper.

Akkus, Oktay, J. Anthony Cookson, and Ali Hortacsu, "The Determinants of Bank Mergers: A Revealed Preference Analysis," *Management Science*, forthcoming.

Baccara, Mariagiovanna, Ayse Imrohoroglu, Alistair Wilson, and Leeat Yariv, "A Field Study on Matching with Network Externalities," *American Economic Review*, 2012, *102* (5).

Bajari, Patrick, Han Hong, and Stephen Ryan, "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica*, September 2010, *78* (5), 1529–1568.

Becker, Gary S., "A Theory of Marriage: Part I," *Journal of Political Economy*, July-August 1973, *81* (4), 813–846.

Berry, S. and E. Tamer, "Identification in models of oligopoly entry," in Whitney Newey Richard K. Blundell and Torsten Persson, eds., *Advances in Economics and Econometrics: Theory and Applications: Ninth World Congress*, Cambridge University Press, 2006, chapter 2, p. 46.

Berry, Steven T., "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 1992, *60* (4), 889–917.

_ **and Philip A. Haile**, "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," 2010. Yale University working paper.

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

Bresnahan, Timothy F. and Peter C. Reiss, "Entry and Competition in Concentrated Markets," *The Journal of Political Economy*, October 1991, *99* (5), 977–1009.

Brock, W.A. and S.N. Durlauf, "Identification of binary choice models with social interactions," *Journal of Econometrics*, 2007, *140* (1), 52–75.

Brown, James N. and Harvey S. Rosen, "On the Estimation of Structural Hedonic Price Models," *Econometrica*, May 1982, *59* (3), 765–768.

Chen, Jiawei, "Two-Sided Matching in the Loan Market," *International Journal of Industrial Organization*, 2013, *31*, 145–152.

Chiappori, P.A., R.J. McCann, and L.P. Nesheim, "Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness," *Economic Theory*, 2010, *42* (2), 317–354.

Chiappori, Pierre-André, Bernard Salanié, and Yoram Weiss, "Partner Choice and the Marital College Premium: Analyzing Marital Patterns Over Several Decades," 2015. Columbia University working paper.

Choo, Eugene and Aloysius Siow, "Who Marries Whom and Why," *The Journal of Political Economy*, November 2006, *114* (1), 175–201.

Christakis, Nicholas, James Fowler, Guido W. Imbens, and Karthik Kalyanaraman, "An Empirical Model for Strategic Network Formation," 2010. working paper.

Ciliberto, Federico and Elie Tamer, "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, November 2009, *77* (6), 1791–1828.

Crawford, Gregory S. and Ali Yurokoglu, "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, 2012, *102* (2).

Crawford, Vincent P. and Elsie Marie Knoer, "Job Matching with Heterogeneous Firms and Workers," *Econometrica*, 1981, *49* (2), 437–450.

Dagsvik, John K., "Aggregation in Matching Markets," *International Economic Review*, February 2000, *41* (1), 27–57.

de Paula, Aureo and Xun Tang, "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," *Econometrica*, 2012, *80* (1), 143–172.

Dean, AM and JS Verducci, "Linear transformations that preserve majorization, Schur concavity, and exchangeability," *Linear Algebra and its Applications*, 1990, *127*, 121–138.

Echenique, Federico, SangMok Lee, Mattew Shum, and M. Bumin Yenmez, "The Revealed Preference Theory of Stable and Extremal Stable Matchings," *Econometrica*, 2013, pp. 153–171.

Eeckhout, Jan and Philipp Kircher, "Identifying Sorting – In Theory," *Review of Economic Studies*, 2011, *78*, 872–906. University Pompeu Fabra working paper.

Eisenhauer, Philipp, James J Heckman, and Stefano Mosso, "Estimation of dynamic discrete choice models by maximum likelihood and the simulated method of moments," *International Economic Review*, 2015, *56* (2), 331–357.

Ekeland, Ivar, James J. Heckman, and Lars Nesheim, "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 2004, *112* (S1), 60–109.

Fox, Jeremy T., "Estimating Matching Games with Transfers," 2010. University of Michigan working paper.

_ , "Identification in Matching Games," *Quantitative Economics*, 2010, *1* (2), 203–254.

_ **and Amit Gandhi**, "Nonparametric Identification and Estimation of Random Coefficients in Multinomial Choice Models," *RAND Journal of Economics*, forthcoming.

_ , **Kyoo il Kim, and Chenyu Yang**, "A Simple Nonparametric Approach to Estimating the Distribution of Random Coefficients in Structural Models," 2015. University of Michigan working paper.

Gale, David, *The Theory of Linear Economic Models*, McGraw-Hill, New York, 1960.

Galichon, Alfred and Bernard Salanie, "Cupid's Invisible Hand: Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," 2012. Ecole Polytechnique working paper.

Gautier, E. and Y. Kitamura, "Nonparametric Estimation in Random Coefficients Binary Choice Models," *Econometrica*, 2013, *81* (2), 581–607.

Gordon, Nora and Brian Knight, "A spatial merger estimator with an application to school district consolidation," *Journal of Public Economics*, 2009, *93*, 752–765.

Graham, Bryan S., "Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers," in A. Bisin J. Benhabib and M. Jackson, eds., *Handbook of Social Economics*, Elsevier, 2011, chapter 19.

Hajivassiliou, Vassilis A and Daniel L McFadden, "The method of simulated scores for the estimation of LDV models," *Econometrica*, 1998, pp. 863–896.

Hatfield, John William, Scott Duke Kominers, Alexandru Nichifor, Michael Ostrovsky, and Alexander Westkamp, "Stability and Competitive Equilibrium in Trading Networks," *Journal of Political Economy*, 2013, *121* (5), 966–1005.

_ , _ , _ , _ , and _ , "Full Substitutes," 2015. University of Texas working paper.

Heckman, James J., Rosa L. Matzkin, and Lars Nesheim, "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, September 2010, *78* (5), 1569–1591.

Hellmann, Thomas and Manju Puri, "The interaction between product market and financing strategy: The role of venture capital," *Review of Financial studies*, 2000, *13* (4), 959–984.

Hitsch, Günter, Ali Hortaçsu, and Dan Ariely, "Matching and Sorting in Online Dating," *American Economic Review*, 2010.

Ho, Katherine, "Insurer-provider networks in the medical care market," *American Economic Review*, 2009, *99* (1), 393–430.

Hsu, David H, "What do entrepreneurs pay for venture capital affiliation?," *The Journal of Finance*, 2004, *59* (4), 1805–1844.

_ **and Rosemarie H Ziedonis**, "Patents as Quality Signals for Entrepreneurial Ventures," in "Academy of Management Proceedings" number 1 Academy of Management 2008, pp. 1–6.

Ichimura, H. and TS Thompson, "Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution," *Journal of Econometrics*, 1998, *86* (2), 269–295.

Kelso, Alexander S. and Vincent P. Crawford, "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, November 1982, *50* (6), 1483–1504.

Koopmans, Tjalling C. and Martin Beckmann, "Assignment Problems and the Location of Economic Activities," *Econometrica*, January 1957, *25* (1), 53–76.

Lerner, Josh, "Venture capitalists and the oversight of private firms," *The Journal of Finance*, 1995, *50* (1), 301–318.

_ , **Hilary Shane, and Alexander Tsai**, "Do equity financing cycles matter? Evidence from biotechnology alliances," *Journal of Financial Economics*, 2003, *67* (3), 411–446.

Levin, Richard C, Alvin K Klevorick, Richard R Nelson, Sidney G Winter, Richard Gilbert, and Zvi Griliches, "Appropriating the returns from industrial research and development," *Brookings papers on economic activity*, 1987, pp. 783–831.

Levine, Anna A., "Licensing and Scale Economies in the Biotechnology Pharmaceutical Industry," April 2009. Washington University in St Louis working paper.

Lewbel, Arthur, "Semiparametric latent variable model estimation with endogenous or mismeasured regressors," *Econometrica*, 1998, *66* (1), 105–121.

_ , "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 2000, *97* (1), 145–177.

Logan, J.A., P.D. Hoff, and M.A. Newton, "Two-sided estimation of mate preferences for similarities in age, education, and religion," *Journal of the American Statistical Association*, 2008, *103* (482), 559–569.

Manski, Charles F., "Identification of Binary Response Models," *Journal of the American Statistical Association*, September 1988, *83* (403), 729–738.

Masten, Matthew, "Random coefficients on endogenous variables in simultaneous equations models," 2015. Working paper.

Matzkin, Rosa L., "Heterogeneous Choice," in W. Newey and T. Persson, eds., *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of the Econometric Society*, Cambridge University Press., 2007.

Mazzeo, Michael J., "Product Choice and Oligopoly Market Structure," *The RAND Journal of Economics*, 2002, *33* (2), 221–242.

McFadden, Daniel L., "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometricac*, September 1989, *57* (5), 995–1026.

Menzel, Konrad, "Large matching markets as two-sided demand systems," *Econometrica*, 2015, *83* (3), 987–941.

Pakes, A. and D. Pollard, "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 1989, *57* (5), 1027–1057.

Park, Minjung, "Understanding Merger Incentives and Outcomes in the Mutual Fund Industry," *Journal of Banking and Finance*, 2013, *37*, 4368–4380.

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

Seim, Katja, "An empirical model of firm entry with endogenous product-type choices," *The RAND Journal of Economics*, 2006, *37* (3), 619–640.

Shapley, Lloyd S. and Martin Shubik, "The assignment game I: the core," *International Journal of Game Theory*, 1972, *1*, 111–130.

Sheng, Shuyang, "A Structural Econometric Analysis of Network Formation Games," 2014. UCLA working paper.

Siow, Aloysius, "Testing Becker's Theory of Positive Assortative Matching," 2009. University of Toronto working paper.

Sørensen, Morten, "How Smart is Smart Money? A Two-Sided Matching Model of Venture Capital," *Journal of Finance*, December 2007, *LXII* (6), 2725–2762.

Sorenson, Olav and Toby E. Stuart, "Syndication Networks and the Spatial Distribution of Venture Capital Investments," *American Journal of Sociology*, May 2001, *106* (6), 1546–1588.

Sotomayor, Marilda, "The multiple partners game," in Mukul Majumdar, ed., *Equilibrium and dynamics: essays in honour of David Gale*, Macmillan, 1992, chapter 17, pp. 322–336.

_ , "The lattice structure of the set of the stable outcomes of the multiple partners assignment game," *International Journal of Game Theory*, November 1999, *28* (4), 567–583.

Tamer, Elie, "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 2003, *70* (1), 147–165.

Thompson, T. Scott, "Identification of Semiparametric Discrete Choice Models," 1989. University of Minnesota Discussion Paper 249.

Uetake, Kosuke and Yasutora Watanabe, "Entry by Merger: Estimates from a Two-Sided Matching Model with Externalities," 2012. Northwestern University working paper.

Weese, Eric, "Political Mergers as Coalition Formation: An Analysis of the Heisei Municipal Amalgamations," *Quantitative Economics*, forthcoming.

Yang, Yupin, Mengze Shi, and Avi Goldfarb, "Estimating the Value of Brand Alliances in Professional Team Sports," *Marketing Science*, November-December 2009, *28* (6), 1095–1111.