Semiparametric Estimation of Two-Sided Matching Models with Endogenous Prices

Jeremy T. Fox*
University of Chicago

May 9, 2005

Abstract

I examine the identification and estimation of the payoffs of agents in two-sided matching models. Koopmans and Beckmann (1957), Shapley and Shubik (1972) and Becker (1973) introduce assignment models with endogenous prices that have the property that all stable matches are socially optimal. I show how to work with the social planning problem to ease estimation. I prove a new result that translates a local definition of social optimality into a statement about matching probabilities from the econometrician's viewpoint. Local social optimality then underlies the identification and consistency proof for a semiparametric maximum score estimator of total match payoffs for the assignment game. The estimator is for a market with prices, but does not require data on prices, so is useful for studying marriage and inter-firm contracts where the details of transfers and contracts are not public. Also, the estimator is computationally tractable because it avoids solving a matching mechanism. The estimator can be applied to markets where the number of potential matches is very large, including markets with one-to-many and many-to-many matching. Identification and estimation rely on an assumption of i.i.d. errors at a marketwide level, but the assumption can be significantly relaxed to allow for agent-specific fixed effects over matching partner nests specified by the econometrician. I also present Monte Carlo studies.

1 Introduction

Koopmans and Beckmann (1957), Gale and Shapley (1962), Shapley and Shubik (1972) and Becker (1973) introduced the study of two-sided matching of heterogeneous agents. Examples include the matching of workers to firms, men to women, families to houses, students to colleges, bidders to multiple objects for sale in an auction, and upstream to downstream firms. A desirable property of a matching assignment is that it be stable: no two parties should want to deviate from their assigned partners and match outside of the assignment. A typical data set for a matching market lists a series of observed matches and some characteristics about the parties in the match. Economists have interests in using the data to estimate the preferences of agents in the market over potential matches. Compared to a model of single-agent discrete choice, estimating a matching model presents

*Thanks to Victor Chernozukhov, Rosa Matzkin and Morten Sørensen for helpful comments. Email: fox@uchicago.edu.

additional complications because the actions of agents to match may preclude the possibility that other agents can match with the same parties. More simply, agents on the same side of the market are rivals, and the choice set of any agent is endogenously determined by those agents on the other side willing to match with it.

Applying single-agent methods, such as the well-known logit and probit discrete choice estimators, to matching markets may give misleading results. For example, it would be incorrect to infer from the fact that most college students do not attend elite universities that most students do not like such institutions. Another possibility is that most students do not have the credentials to attend such a university, so the elite university is not in their choice set. One approach to addressing the endogeneity and even the unobservability of choice sets is to make educated guesses about the equilibrium choice sets of agents. Even if these guesses are correct, the resulting single-agent estimator is inconsistent, because the choice sets are a function of unobserved factors affecting match payoffs. If the original matching model has i.i.d. errors over potential matches, the model's errors conditional on a certain equilibrium choice set are far from i.i.d.

In this paper, I show that the total payoffs created by a match in a two-sided matching model with endogenous prices and transferable utilities are semiparametrically identified. I study the assignment problem first introduced by Koopmans and Beckmann (1957) and generalized to game theoretic forms by Shapley and Shubik (1972) and Becker (1973) for one-to-one matching and Sotomayor (1992) for the case where agents on both sides of the market can make multiple matches. I also provide a tractable consistent estimator for the total surpluses of matches.

A pairwise match is an observed relationship from among a potentially large but finite set of possible matches. Therefore there are natural links between estimating two-sided matching models and estimating single-agent discrete choice models. In the context of multinomial discrete choice, Manski (1975) shows that, under an assumption of i.i.d. stochastic payoffs across discrete choices, choice probabilities are monotonic in deterministic payoffs. This monotonic relationship can be re-expressed locally: given two deterministic payoffs for discrete choices, the choice with the higher deterministic payoff will be made with higher unconditional probability. A literal extension of the single-agent monotonicity property to a two-sided matching game does not hold. Given two matches, the match with the higher deterministic payoff may actually be realized with lower probability if either of the two agents involved have attractive outside options. Matching a marginal student to an elite university may create more deterministic payoffs than matching the student to a community college, but the elite university may find it has better applicants for its available slots.

The theory of matching does, however, impose more subtle restrictions on matching probabilities. Koopmans and Beckmann (1957) and others prove a key property of an assignment game with endogenous prices, transferable utility and one-to-one matching: any stable match of physical pairings and transfers between agents is socially optimal. Social optimality means the economy-wide sum of match payoffs before transfers is the maximum achieved by any set of physical pairings. Social optimality can be seen as a version of the first welfare theorem for a matching market with transferable utility. The implication of social optimality is that the stable assignment in a decentralized market economy can be computed as a social planning problem, without the need to calculate the endogenous prices underlying the stable match in the decentralized market. Sotomayor (1992) extends the social optimality property of stable matches to the case where agents on both sides of the markets may make multiple matches, and payoffs for multiple matches are additive across matches.

This paper shows the notion that data on observed matches represent a social optimum places empirically relevant restrictions on the payoffs of agents. I base my identification and estimation strategies on a concept of local social optimality. Fix a quartet of four agents, two from each side of the market. For any two sets of matches, social optimality implies that the pair of observed matches give a weakly greater total payoff than the counterfactual situation where the four agents exchange partners. I prove that this statement about local social optimality survives translation into a statement about matching probabilities when an econometrician does not observe an i.i.d. error term at the marketwide outcomes level. Here, a probability is calculated unconditionally over all possible market outcomes. For example, a matching probability for two pairs of matches involving four agents in a market with 1000 agents on each of the two sides is a function of the $1000^2 = 1$ million observable covariates entering into match surpluses. The probability involves integrating out over error terms for each of the $1000! \approx 4.02 \times 10^{2567}$ unique collections of physical pairings possible in the market, if only one-to-one matches are allowed.

If the goal is to estimate the parameters in a deterministic index of match payoffs, I prove that the two matches that together give a higher deterministic payoff than the counterfactual exchange of partners must be made simultaneously with a greater probability than the probability of simultaneously observing the exchange of partners. Intuitively, an agent whose typical match partner is unavailable because of a marketwide stochastic payoff shock will more often than not pick the partner within the quartet that gives the agent a greater payoff. I can sign the difference between matching probabilities because the outside options available from partners other than the quartet under consideration contribute similarly to the probabilities of the two observed pairs matching and the counterfactual exchange of partners matching.

Signing the difference between the two probabilities requires that the stochastic components of payoffs are i.i.d, and enter at the marketwide outcome, rather than the individual match, level. The marketwide error structure places the marketwide matching problem into a single-agent, multinomial discrete choice framework I develop in Fox (2005), which itself is related to pioneering work on the semiparametric discrete choice models by Manski (1975).

My new mathematical result about the link between the local social optimality of deterministic payoffs and matching probabilities forms the basis for an identification argument. Again, fix a quartet of four agents, two on each side of the market. If the total payoffs of the matches have continuously varying observable covariates when asymptotically sampled across markets, then the total payoffs from a match are semiparametrically point-identified. If there are not continuously varying covariates, then the parameters in the deterministic payoff function are set-identified (bounded). Given my results about local social optimality and unconditional matching probabilities for a quartet of agents, the subsequent identification argument is similar to proofs in the single agent, semiparametric discrete choice literature (Manski, 1988).

A consistent estimator based on the identification argument involves nonparametrically estimating match probabilities that are functions of potentially millions or billions of observables. A major contribution of the paper is to introduce a tractable maximum score estimator for assignment games. The estimator is semiparametric because it does not require the assumption of parametric distributions for the stochastic portions of marketwide errors. Importantly, the matching estimator is for market with endogenous prices, like many of the markets studied by economists. However, prices do not enter into the definition of social optimality, so the estimator does not actually use data on the endogenous prices. Therefore, the estimator is an excellent tool to study contracting between firms, where typically the contracts involve prices that are not publicly disclosed, or marriage, where transfers between husbands and wives are often not observed.

Consistency of the maximum score estimator comes from the social optimality property of the stable match solution concept in assignment games. Basing estimation on the properties of the solution has several advantages over the nested solution methods used in two recent papers. Sørensen (2004) estimates the total payoffs of venture capitalists and entrepreneurial companies matching with each other, while Boyd, Lankford, Loeb and Wyckoff (2003) study the matching of teachers to public schools. Both of these papers employ Gale and Shapley (1962) matching models without endogenous prices, and make additional assumptions in order to guarantee uniqueness of the stable match for every parameter value and vector of error terms. Sørensen assumes that the preferences of venture capitalists and entrepreneurial companies are perfectly aligned, so that the unique stable match can be calculated by a top-down sorting algorithm. Boyd et al. study a model with non-aligned preferences, and impose a particular matching algorithm that picks one out of many stable matches. By contrast, assignment games have a unique stable allocation of matches without restrictions on preferences or the use of equilibrium selection algorithms, as the presence of endogenous prices assures that the socially optimal allocation is chosen.[1] Additionally, as the semiparametric matching estimator uses a local notion of social optimality, it does not require data on the quotas of individuals agents, and whether particular agents can choose to be unmatched.[2]

A major concern with the previous parametric nested solution approaches to two-sided matching is that they are computationally burdensome. A nested matching mechanism must be solved for all trial parameter values and, even more ambitiously, every combination of the set of all stochastic shocks needed for numerical integration in order to evaluate an objective function. In a market with 1000 agents on each side with match-specific shocks, there are 1 million such error terms. Further, the computational costs of solving the mechanism and evaluating the likelihood scale poorly in the number of agents in the market being studied. By assuming marketwide rather than match-specific errors, the assignment maximum score estimator introduced in this paper addresses both computational issues. First, the estimator uses only the revealed preference conditions inherent in social optimality, and avoids the need to nest a matching mechanism in the estimation procedure. Nested solution methods have not been used for models with endogenous prices as a linear programming problem would have to be solved for every trial parameter value and every vector of the stochastic errors. Sørensen (2004) computes that if he were to modify his nested solutions method to nest the linear programming problem inherent in assignment games, estimating a model on the same scale as his previous work (35 agents on each side of the market) would take 800 years, even at the upper bound of the performance of a recently introduced linear programming algorithm specialized for assignment games.

Second, the local notion of social optimality means that the maximum score estimator is consistent when only a subset of potential matches are entered into the objective function for computational reasons. Most nested solution estimators require covariate data on all possible matches, in order to solve the model and compute matching probabilities that enter the objective function. Matching probabilities computed without using the data on all available matches cannot be cleanly related to matching probabilities in the entire market. The maximum score estimator in this paper involves pairwise comparisons between two pairs of matching

[1]Only the set of stable matches (not necessarily the vector of corresponding prices) is unique. There is a unique social optimum with probability 1 because the stochastic portion of payoffs is assumed to have a distribution with no mass points.

[2]For example, in a labor market, in a nested solution approach, the matching algorithm requires the researcher to input the maximum number of employees that a firm can match with, even though this quota is usually not found in the data.

arrangements for a quartet of four agents. Local social optimality places sign restrictions on the difference between the unconditional probabilities of the two pairs of matching arrangements. Only the sign restrictions involving one quartet of agents are needed for identification and consistency. The assignment games maximum score estimator can be used for matching markets where the number of possible pairings exceeds the number of atoms in the universe, which is higher than the $35^2 = 1225$ matches per market that nested solutions methods have been used for.

A weakness of multinomial maximum score estimators is that consistency relies on the stochastic payoffs being i.i.d., an assumption originating in the single agent work of Manski (1975). In the context of the single-agent, binary-choice maximum score estimator, Abrevaya (2000) shows how to use panel data to consistently estimate parameters on time-varying covariates in the presence of fixed effects. This identification argument can be extended to multinomial cases using cross-sectional data if the researcher is willing to divide the potential match partners into nests. An agent has a common fixed effect for all the match partners in a given nest. Comparing match payoffs for partners within the same nest identifies the parameters on the observable covariates, as the fixed effect is held constant within the nest, and does not alter the ordering of unconditional matching probabilities for quartets that are the key implications of local social optimality.

This paper concludes by offering practical implementation advice for statistical inference. Kim and Pollard (1990) show a class of models, that includes binary choice maximum score, converges at a $\sqrt[3]{n}$ rate, and the asymptotic distribution is not tractable for inference. Abrevaya and Huang (2005) further prove that the standard bootstrap is inconsistent in estimators in the Kim and Pollard class, which almost certainly includes the maximum score estimators in this paper. Therefore, I follow the advice of Delgado, Rodríguez-Poo and Wolf (2001) for the binary choice maximum score and suggest subsampling as a consistent estimator of test statistics. Finally, I present Monte Carlo evidence about the finite-sample performance of the assignment games maximum score estimator.

2 Identification in Assignment Games

This section discusses how the social optimality property of stable matches in assignment games can be used to identify the parameters in the total social surplus of a match.

2.1 Match Payoffs and Stable Matches

Consider a Koopmans and Beckmann (1957) assignment model. Let there be two sides to a matching market, upstream firms and downstream firms. More traditional examples of one-to-many and one-to-one matching, such as workers to employers and men to women, are special cases of the upstream-downstream model of many-to-many matching. The estimators of this paper apply to all of these markets.[3]

[3] Lucas (1995) discusses some known results about assignment games when there are more than two sides to a match. Unfortunately, the core of matching market can be nonempty in these cases. Abeledo and Isaak (1991) discuss how there may not exist a stable match in a market, like the roommates problem of Gale and Shapley (1962), where matches are between agents on the same side of a market. Ostrovsky (2004) shows how a special chain structure for n-way matching guarantees the existence of a stable chain. It is likely, although unproven, that the estimators in this paper could be modified to apply to the Ostrovsky (2004) supply chain case, when the payoffs satisfy the restrictions inherent in assignment games.

If upstream firm a matches with downstream firm i, it receives a structural payoff before transfers of $x_{ai}^{\text{up}\prime}\beta^{\text{up}}$, where x_{ai}^{up} represent the vector of the observable characteristics of the downstream firm i valued by upstream firms interacted with the characteristics of the upstream firm a, and β^{up} is a vector of parameters multiplying those characteristics. Similarly, downstream firm i receives a payoff before transfers of $x_{ai}^{\text{down}\prime}\beta^{\text{down}}$ if it matches with upstream firm a, where x_{ai}^{down} represents characteristics observable to the econometrician. When parties match with multiple agents, their payoff is the sum of the payoffs from each of the matches. In other words, all matching partners are perfect substitutes for each other, and externalities across matches are ruled out. Ruling out externalities is a serious restriction for applications to vertical relationships, as upstream and downstream firms cannot be worried about changes in post-match competition arising from matching. Extending the theory of two-sided matching to handle externalities is a holy grail of matching theory.

Firms matching together can exchange a transfer, p_{ai}. Transfers can be negative. Without loss of generality, say the downstream firm pays this transfer to the upstream firm.[4] Further, assume payoffs are additively separable between other characteristics and transfers, so that an upstream firm's final payoff is $x_{ai}^{\text{up}\prime}\beta^{\text{up}} + p_{ai}$, while its matched downstream firm receives $x_{ai}^{\text{down}\prime}\beta^{\text{down}} - p_{ai}$. Additive separability is often called transferable utility. A notion of equilibrium used in this class of models is a stable match, which is a collection of assigned matches and corresponding transfers where all matched parties receive greater benefit to being matched than remaining unmatched, and the total payouts to any pair of firms exceeds the payoffs they would receive if they deviated and matched outside of the mechanism.

Koopmans and Beckmann (1957) prove the amazing result that, under the assumption of transferable utility, any stable match is socially optimal, meaning that the total sum of non-transfer structural payoffs is the maximum that can be achieved by an allocation of upstream to downstream firms.[5] Social optimality can be seen as an implication of the first welfare theorem for markets with transferable utility. The proof of social optimality uses the fact, in linear programming, cost minimization is the dual to output maximization. The social efficiency result has been proved for the case where both sides of the market can make multiple matches (a number not greater than some fixed, agent-specific quota) and payoffs are additively separable across matches by Sotomayor (1992).[6]

The fact that in this class of models all stable matches are socially optimal leads me to focus on an identification method that relies on social optimality, rather than stable matches. Transfers do not affect a social planner's calculation of whether an assignment of physical pairings is socially optimal, so they play no more role in my identification and estimation strategy. Thus, my identification arguments are for markets with prices, but they do not use data on prices. The estimator is thus an excellent tool to study relationships between firms, where the financial terms of contracts are usually not disclosed publicly. The estimator can also apply to the marriage market case studied in Becker (1973), where economists believe there are procedures within marriages to transfer consumption between spouses, but data on such transfers are hard to collect.

Under this payoff structure, the total payoff from a match of upstream firm a and downstream firm i is

$$x_{ai}^{\text{up}\prime}\beta^{\text{up}} + x_{ai}^{\text{down}\prime}\beta^{\text{down}}.$$

[4]While Shapley and Shubik (1972) allow for prices to satisfy economy-wide feasibility constraints, individual rationality constraints inherent in the notion of a stable assignment imply that the total payments to agents in a match will equal the total surplus from that match, or that prices in a stable assignment can be thought of as transfers between two parties in a match.

[5]This result is Corollary 8.8 in the usual reference on two-sided matching, Roth and Sotomayor (1990).

[6]A reference that is more easily available online is Sotomayor (1999).

Only the total structural payoff of two agents in a match is relevant for computing whether a match could be part of a socially optimal allocation. It is not possible to use social optimality alone to distinguish whether large upstream and downstream firms are likely to match because large upstream firms value large downstream firms or because large downstream firms value large upstream firms.

Further, social optimality depends on only the portion of the payoff of a match that depends on the interaction of characteristics between the matching firms. Therefore, I introduce the new notation for the portion of match characteristics affecting social optimality

$$x_{ai}'\beta = x_{ai}^{\text{up}\prime}\beta^{\text{up}} + x_{ai}^{\text{down}\prime}\beta^{\text{down}}, \tag{1}$$

where it understood that the composite characteristic vector x_{ai} combines linearly dependent terms in x_{ai}^{up} and x_{ai}^{down} and β also combines the corresponding parameters in β^{up} and β^{down} for the linearly dependent terms of x_{ai}^{up} and x_{ai}^{down}.[7]

Note that I have not assumed that agents to a match have the same preferences over the match, or that agents split some surplus in fixed proportion, which together are assumptions Sørensen (2004) cleverly exploits to prove that his nested matching mechanism produces a unique stable match. A unique stable assignment is important when estimating a Gale and Shapley (1962) two-sided matching model without prices. By contrast, uniqueness arises naturally in assignment games with endogenous prices when covariates entering match payoffs have continuous distributions, so that feasible match assignments will sum to the same social optimal payoff value with probability 0.

2.2 Marketwide and Local Social Optimality

As discussed above, the presence of transferable utility (transfers are additively separable from other covariates) ensures that any stable assignment of upstream firms to downstream firms maximizes the sum of marketwide payoffs from all assigned matches. Matching theory takes as given a matrix of match payoffs, where an individual cell in the matrix is the payoff of a match between upstream firm a and downstream firm i, $x_{ai}'\beta$. The optimal assignment can then be computed as a social planning problem that maximizes total payoffs subject to feasibility constraints, namely the quota, or number of matches, that each agent can make.

Let U be the set of upstream firms in a matching market, and let D be the set of downstream firms in the same market. Define y_{ai} to be the indicator variable equal to 1 if firms a and i are assigned to match, and 0 otherwise. The social planning problem is easier to solve if y_{ai} can be any real number between 0 and 1. At the solution, y_{ai} will be either 0 or 1 for all matching possibilities. The social planner chooses the set of $y_{ai} \in [0,1]$ for all upstream and downstream firms to maximize

$$\sum_a \sum_i \left(x_{ai}'\beta\right) y_{ai}$$

[7] If the number of matches each firm will make at a stable allocation is preset, and no firm remains unmatched, then social optimality will not depend on the presence of unobserved intercept terms in match payoffs. The payoff of a downstream firm with intercepts can be written as $x_{ai}^{\text{up}\prime}\beta^{\text{up}} + \xi_a^{\text{up}} + \omega_i^{\text{up}}$, where ξ_a^{up} is a fixed payoff of firm a that it gets regardless of its match, and ω_i^{up} is a fixed payoff any upstream firm gets for matching with downstream firm i. Likewise, downstream firms can have payoffs $x_{ai}^{\text{down}\prime}\beta^{\text{down}} + \xi_a^{\text{down}} + \omega_i^{\text{down}}$. It is not possible to use local social optimality to identify these intercepts, but the parameter β in the portion of deterministic surplus that affects social optimality can still be identified despite the presence of the intercepts.

subject to the feasibility constraints that all agents are under their quotas, or

$$\begin{aligned}\sum_i y_{ai} &= q_a^{\text{up}}\ \forall a \in U\\ \sum_i y_{ai} &= q_i^{\text{down}}\ \forall i \in D.\end{aligned}$$

The quota of an upstream firm a, q_a^{up}, is the number of downstream firms that a can physically match with at one time. Because the coefficients and constraints are all linear in y_{ai}, this is a linear programming problem that is relatively easy to solve once.

Note that the above feasibility constraints allow agents to be unmatched. If an agent does not use up his quota, the agent is said to match with some finite number of dummy agents that represent that those quota spots are unfilled. Call these outcomes $y_{a0} = 1$, for some example upstream firm a and the dummy agent 0. The quota of the dummy agent 0 is unlimited. The payoff to being unmatched is normalized to 0 above, but this can be weakened in an empirical application, which can treat being unmatched as another matching partner with observable characteristics.

In this paper, I will solve an inverse problem. I am interested in taking data that I assume come from a (decentralized) solution to the linear programming problem, and produce estimates of the parameters β in $x'_{ai}\beta$. In markets with tens, hundreds, thousands, millions or billions of potential matches, repeatedly solving a linear programming problem is too computationally expensive to nest into an estimation algorithm.

The solution to the social planner's problem implies many restrictions for two upstream and two downstream firms at a time. If both matches of a and i and b and j are observed, then a (possible) local implication of social optimality is that the total surplus of two matches exceeds the total surplus from a matching with j and b matching with i.

Definition 1. *Matches between upstream firm a and downstream firm i, and upstream firm b and downstream firm j, are locally socially optimal when*

$$x'_{ai}\beta + x'_{bj}\beta \geq x'_{aj}\beta + x'_{bi}\beta. \tag{2}$$

Consider a hypothetical solution to the marketwide social planning problem where a matches with i and not j, and b matches with j and not i. If the local social optimality inequality condition is not satisfied, having a match with j and b match with i would improve total surplus from the quartet, without disturbing the matches of firms outside of the quartet. In a market where only one-to-one matching is allowed, itemizing over all possible quartets (a, b, i and j) will produce the definition of social optimality for the entire market, as long as remaining unmatched is considered a potential matching partner, where appropriate. In a market with many-to-one or many-to-many matching, local social optimality for all quartets is a weak implication of market-level social optimality, as local social optimality does not consider b deviating to match with i while a remains matched with i.

Definition 1 implies that components of payoffs that are not interactions between the characteristics of upstream and downstream firms do not contribute to social optimality. To see this, consider a payoff characteristic x_a, that only depends on a characteristic of the upstream firm a. One interpretation is that all downstream value the characteristic identically. In this case, all payoffs for matches with a increase by the same value, x_a times its parameter, and thus this term appears on both sides of the inequality in equation (2), and therefore cancels out from both sides. If there is a term such as x_a in the characteristic vector, then the parameter multiplying x_a cannot be identified, but as it cancels out from the local social optimality condition, other parameters can be identified. Thus payoffs are partially identified.[8]

For some policy questions, the cancellation of characteristics that are not interactions between the characteristics of upstream and downstream firms is an empirical advantage. Many datasets lack covariate data on all important characteristics of upstream and downstream firms. If some of these characteristics affect the level of match surplus of all matches equally, they difference out, and do not affect the optimal assignment of upstream to downstream firms. Therefore, if the policy questions of interest to the investigator are not functions of these unobserved characteristics, than differencing them out leads to a great deal of empirical robustness to missing data problems.

2.3 Marketwide Assignment Error Terms

This paper solves an inverse problem to identify the parameters β in the deterministic payoffs $x_{ai}'\beta$. In any empirical investigation, there is not data on all characteristics relevant to an observed outcome. Therefore, the econometrician needs to introduce errors so that the model can fit the data.

There are several approaches one could take to incorporating error terms into a two-sided matching model. The nested matching mechanism estimators used by Boyd et al. (2003) and Sørensen (2004) make the match payoffs equal to $x_{ai}'\beta + \varepsilon_{ai}$, where ε_{ai} is an i.i.d. across matches error term. Unfortunately, repeatedly solving the entire optimal assignment linear programming problem for all hypothetical combinations of values of the ε_{ai} for upstream firms and downstream firms is not possible. If there are 1000 upstream firms and 1000 downstream firms, there are 1 million unobserved random variables ε_{ai}. Even if there are only around 35 upstream and downstream firms, Sørensen suggests nesting a linear programming problem will require 800 years for estimation, as evaluating the linear programming problem even once is somewhat expensive (on the order of 0.7 seconds for a specialized assignment games routine), and the linear programming problem must be evaluated thousands or millions of times in numerical integration over the $35^2 = 1225$ errors in the 35 firm case.

I address this scalability problem by placing the error terms at the marketwide allocation level. Each set of assignments of upstream to downstream firms, including the probability of being unmatched, results in a payoff, for marketwide assignment E, of $\sum_{cg\in E} x_{cg}'\beta$, where the index $cg \in E$ represents an assigned match in E. If there are 1000 upstream firms and 1000 downstream firms in a one-to-one matching market without the option of remaining unmatched, the total number of marketwide assignments of physical pairings E is 1000!, or 4.02×10^{2567}, which is a very large number. I introduce an error term for each marketwide allocation, so the final perceived marketwide payoff of assignment E is

$$\sum_{cg\in E} x_{cg}'\beta + \varepsilon_E.$$

[8]One way of identifying (or at least bounding, if covariate assumptions below are not met) the coefficient on x_a in payoffs is if unmatched upstream firms are observed, and unmatched firms do not value their own characteristic x_a. Then x_a is implicitly multiplied by an indicator variable equal to 1 if a match partner is not the null set, and 0 if the match partner is remaining unmatched. In this case, x_a is truly not a characteristic valued equally by all downstream firms, as being unmatched is treated as a type of firm.

This transforms a complex two-sided matching market estimation problem into a single-agent discrete choice problem. A social planner considers the sum of deterministic payoffs generated by any marketwide matching assignment, $\sum_{cg \in E} x'_{cg}\beta$, and adds a random error term to the final payoff.

For a minute, consider the case of single-agent discrete choice, and say agent a decides between discrete choices such as i and j. The multinomial maximum score estimators of Manski (1975), Matzkin (1993) and Fox (2005) allow the estimation of single-agent discrete choice models without imposing a particular parametric functional form for the disturbance term. An important assumption is, however, that the error terms ε_{ai} are i.i.d. across choices for a given agent. The functional form for the disturbances can be completely different across observationally distinguishable agents, so that agents from Texas might have Laplace errors, and agents from Illinois might have multimodal, mixed normal errors with much smaller variances. In notation, the error terms have a common distribution function $F(\varepsilon_{ai} \mid x_a)$, where x_a is all of the covariates for agent a.

Returning to the two-sided matching case. Adding, say, an outrageous1000! error terms at the marketwide allocation E level instead of "only" 1 million error terms at the individual match ai level is somewhat arbitrary from the viewpoint of matching theory, as the random errors ε_E are not assigned to the payoffs of any individual agents. However, tractable estimation prevents me from itemizing all 1000! outcomes, or even all components entering a single outcome in markets with many upstream firms and downstream firms. This paper will show estimation can proceed with data on only two upstream firms and two downstream firms per market, even if the true size of the markets are much larger. This requires that the errors be i.i.d. across marketwide assignments.

If instead I had included match specific error terms, the payoffs to overall marketwide outcomes would be statistically correlated. The same match-specific error term ε_{ai} would appear in multiple outcomes, as a can match with i in many situations involving changes in the matching arrangements of the other firms in the market. Eliminating correlation across market outcomes, and therefore the assumption of marketwide assignment errors ε_E is critical to the semiparametric identification and estimation strategy taken by this paper, as the result that the probability of observing a given marketwide outcome is monotone in $\sum_{cg \in E} x'_{cg}\beta$ requires i.i.d. errors across outcomes.

Assumption 1. *For all marketwide assignments E of upstream to downstream firms, and including the option of remaining unmatched where appropriate, let the random variable ε_E be i.i.d. across choices for a given market. Let the random variable ε_E have a continuous distribution $F(\varepsilon_E \mid X_h)$ with full support and no mass points. Let ε_E have a corresponding density $f(\varepsilon_E \mid X_h)$.*

The distribution function of the error terms can vary across markets with potential market observables X_h. I define X_h in more detail in the next section.

While i.i.d. stochastic error terms is a restrictive assumption if the observable covariates have low explanatory power for predicting matches, Section 3.5 discusses how to relax Assumption 1 by allowing for firm-specific fixed effects over pre-specified nests of match partners.

2.4 The Definition of a Market

The use of asymptotic theory to prove identification of the assignment maximum score estimator requires me to choose whether the limiting population is observing a matching market with an infinite number of agents, or observing an infinite number of matching markets, each with a finite number of firms. A market with an infinite number of firms changes the character of the matches that will be observed; it is much simpler to consider a population with an infinite number of markets.

Each market h is distinguished by its characteristics, X_h, its observed set of matches, and the unobserved stochastic error terms generating the observed matches. The vector X_h is a particularly important construct in understanding the theoretical properties of the estimator I will introduce below. X_h contains most of the exogenous characteristics of a matching market.

Definition 2. *The vector of most of the exogenous characteristics of matching market h is X_h.*

- *X_h contains the number of upstream, U_h, and the number of downstream firms, D_h, in market h.*
- *For each pair of an upstream firm a and downstream firm i, X_h contains the vector of (potentially) observable characteristics x_{aih} entering into the total match surplus.*
- *Characteristics entering the value of remaining unmatched also enter into X_h.*
- *X_h also contains the quota, q_{ih}^{down} or q_{ah}^{up}, the number of matches each firm can make.*

If there are 1000 upstream firms and 1000 downstream firms, X_h contains 1 million vectors of covariates x_{ai} as well as other data. The stochastic payoff terms ε_E are exogenous from a matching theory standpoint, but are specifically excluded from X_h.

In order to solve the social planner's linear programming problem for a given realization of all the error terms, every component of X_h must be observable. Consider a (computationally intractable) parametric or semi-parametric pseudo-maximum likelihood procedure that involves a nested solution to the linear programming problem for all combinations of error term values and unknown payoff parameters β. Every component of X_h would be needed to solve the model and therefore for parametric identification and consistency. Identification in this paper relies on the local social optimality property of the observed stable match, and I will show in Section 3 that not every component of X_h must be observable for identification and estimation. One important example is the quota of each agent is not needed, as I will discuss in Section 3.4.

In the theory of two-sided matching, a market is the collection of agents who may physically match with each other. In many applications, the definition of a market may be unclear to the econometrician. In fact, the definition of the relevant market is an important issue in most anti-trust litigation. However, the economic theory of two-sided matching is only developed for the case where a market is well-defined, and that is how this paper will proceed.

2.5 Unconditional Quartet Match Probabilities in Assignment Games

The local social optimality condition in Definition 1 needs to be transformed into a statement about matching probabilities in order to be useful for empirical work, as the stochastic components of marketwide payoffs, the ε's, are not observed in data.[9]

[9] As the econometrician does not observe ε_E, the matching probabilities are from the econometrician's point of view, not the agents or the social planner in the model, who can implicitly calculate the gains to any potential match. This distinguishes assignment games from search models, where an agent must spend time to sample more match partners.

If there are 1000 upstream firms and 1000 downstream firms in a one-to-one matching market without the option of remaining unmatched, the total number of marketwide allocations of physical pairings is 1000!, or 4.02×10^{2567}. Let the massive set of all feasible marketwide allocations be $\mathcal{Z}_h$. The unique socially optimal allocation E_h satisfies,

$$\sum_{ckh \in E_h} x'_{ckh}\beta + \varepsilon_{E_h} \geq \sum_{ckh \in Z_h} x'_{ckh}\beta + \varepsilon_{Z_h} \ \forall \text{ allocations } Z_h \neq E_h, Z_h \in \mathcal{Z}_h, \tag{3}$$

which states that the sum of the payoffs of a socially optimal allocation is greater than other feasible allocations, where feasible allocations enforce the quotas of agents. The inequality in equation (3) transforms the computation of the social optimum into a single-agent discrete choice problem with extra i.i.d. additive errors.

The errors need to be integrated out to calculate the probability, from the point of view of an econometrician, of an assignment maximizing the deterministic payoffs of agents and the marketwide error term. The definition of local social optimality involves a quartet of agents. Therefore, I introduce the concept of a quartet matching probability.

Definition 3. *In a matching market h with characteristics X_h, consider the quartet of upstream firms a and b and downstream firms i and j. Let $\mathcal{E}_h(ai, bj) \subseteq \mathcal{Z}_h$ be the set of all allocations in h where a matches with i and b matches with j. Then the unconditional quartet match probability is*

$$P(ai, bj \mid X_h) = \sum_{E \in \mathcal{E}_h(ai,bj)} \text{Prob}_{\varepsilon_h}(E_h \textit{ solves social planner's discrete choice problem}), \tag{4}$$

where the social planner's discrete choice problem is defined in equation (3).

$P(ai, bj \mid X_h)$ is the unconditional probability, from the econometrician's point of view, of a matching with i and b matching with j. Calculating $P(ai, bj \mid X_h)$ involves integrating out the vector of all marketwide assignment specific error terms, ε_E^h, or, alternatively, all allocations in $\mathcal{Z}_h$ where a does not match with i and b does not match with j.

The quartet matching probabilities are computed holding fixed the large vector X_h of all observable market characteristics. As X_h can have millions or billions of elements, an estimator that involves computing quartet match probabilities will not be tractable. Quartet match probabilities are well defined if there is a unique socially optimal match with probability 1.

The reason for specifying the joint probability of two matches is because the estimator focuses on the local social optimum property of exchanging two matches. More formally, proving that an extremum estimator is consistent requires showing that the probability of the objective function has a unique relevant extremum at the true parameter value. The probability limit of the maximum score objective function will involve terms such as $P(ai, bj \mid X_h)$.

2.6 Monotonicity of Matching Probabilities under Local Social Optimality

A key insight of Manski (1975) for single-agent discrete choice models is that the i.i.d. property implies that choice probabilities $P(i \mid X)$ are monotone in the deterministic part of utility $x'_{ai}\beta$, so observed choices should, more often than not, have greater deterministic linear indices than unobserved choices. Consider a agent making a standard, single-agent, multinomial discrete choice from a set J of choices. Fix two choices, i and j, from the set J of all choices. Under a single-agent version of Assumption 1, $x_i'\beta > x_j'\beta$ if and only $P(i \mid X) > P(j \mid X)$.[10]

A literal extension of the monotonicity of outcome probabilities does not hold in matching models, as a match ai that gives a higher deterministic payoff $x_{ai}'\beta$ than another match bj may not be observed with higher frequency if either a or i has good outside options. However, for semiparametric identification, I need to find a similar monotonicity property for an assignment game. It turns out that social optimality of an assignment game implies that there will be local socially optimal matching in a probabilistic sense. Given two upstream firms and two downstream firms, it is more likely that the combination of two matches with the higher deterministic payoff, say $x_{ai}'\beta + x_{bj}'\beta$, will be observed than the alternative combination, which in this case has a total payoff of $x_{aj}'\beta + x_{bi}'\beta$. The following theorem is to my knowledge new and is the key mathematical property of assignment models that allows their semiparametric identification and estimation. The theorem holds for one-to-one and many-to-one matching.

Theorem 1. *Consider two upstream firms, a and b, and two downstream firms, i and j, all in a one-to-one or many-to-one matching market h with endogenous prices, transferable utility, and utility for an agent who makes multiple matches that is additive across matches. Under Assumption 1,*

$$x_{aih}'\beta + x_{bjh}'\beta > x_{ajh}'\beta + x_{bih}'\beta,$$

if and only if

$$P(ai, bj \mid X_h) > P(aj, bi \mid X_h).$$

The proofs of theorems are collected in Appendix A. The intuition for the proof is understandable. Under the conditions in the theorem, there are the same number of marketwide assignments E_1 where a matches with i and b matches with j as assignments E_2 when a matches with j and when b matches with i. These assignments have payoffs of the form

$$x_{aih}'\beta + x_{bjh}'\beta + \sum_{cg \in E_1 \setminus \{ai, bj\}} x_{cgh}'\beta + \varepsilon_{hE_1}$$

and

$$x_{ajh}'\beta + x_{bih}'\beta + \sum_{cg \in E_1 \setminus \{aj, bi\}} x_{cgh}'\beta + \varepsilon_{hE_2}.$$

As the maximum number of possible matches, the quota, of firms i and j is not violated by switching the matching partners of a and b, for every set of other matches $E_1 \setminus \{ai, bj\}$, there is an identical set $E_1 \setminus \{aj, bi\}$. It can be shown by simple integration that the marketwide assignments E_1 and E_2 happen with the same probability.

For many-to-many matching, Theorem 1 does not hold, because the theorem does not rank assignments of the form

$$x_{aih}'\beta + x_{bjh}'\beta + x_{ajh}'\beta + \sum_{cg \in E_1 \setminus \{ai, bj\}} x_{cgh}'\beta + \varepsilon_{hE_1},$$

[10] The proof is Case (b) of Step 2 on pages 212-213 of the consistency theorem in Manski (1975), and relies on writing the functional form for choice probabilities in terms of an integral over the error terms in the model.

which is an event that contributes to the calculation of $P(ai, bj \mid X_h)$ but not $P(aj, bi \mid X_h)$, vs. assignments of the form

$$x'_{ajh}\beta + x'_{bih}\beta + x'_{bjh}\beta + \sum_{cg \in E_1 \setminus \{aj, bi\}} x'_{cgh}\beta + \varepsilon_{hE_2},$$

which contributes to the calculation of $P(aj, bi \mid X_h)$ but not $P(ai, bj \mid X_h)$.

The arguments from the cases of one-to-one and many-to-one matching about assignments where a matches with i and not j and b matches with j and not i still hold. Therefore, let

$$\bar{P}(aj, bi \mid X_h) = \sum_{E \in \mathcal{E}_h(ai,bj) \cap \mathcal{E}_h(ai,bj,aj) \cap \mathcal{E}_h(ai,bj,bi)} \text{Prob}_{\varepsilon_h}\left(E_h \text{ solves social planner's discrete choice problem}\right)$$

be the probability of a matching with i, b matching with j, but a not matching with j and b not matching with i. Here, $\mathcal{E}_h(ai, bj, aj)$ is the set of marketwide assignments where a matches with i, b matches with j, and a matches with j. Then the following corollary to Theorem 1 holds.

Corollary 1. *Consider two upstream firms, a and b, and two downstream firms, i and j, all in a many-to-many matching market h with endogenous prices, transferable utility, and utility for an agent who makes multiple matches that is additive across matches. Under Assumption 1,*

$$x'_{aih}\beta + x'_{bjh}\beta > x'_{ajh}\beta + x'_{bih}\beta,$$

if and only if

$$\bar{P}(ai, bj \mid X_h) > \bar{P}(aj, bi \mid X_h).$$

The proof is also in the appendix.

2.7 Identification Using Covariates for Only One Quartet

The following identification arguments are written for one-to-one and many-to-one matching. However, all the following arguments hold for many-to-many matching if $\bar{P}(aj, bi \mid X_h)$ replaces $P(aj, bi \mid X_h)$.

Point identification is showing that an there is only one parameter β^0 that could generate the data for an infinite number of observed markets. If there are an infinite number of markets, there are also an infinite number of identical markets, and the matching probabilities $P(ai, bj \mid X_h)$ are observable. Given the matching probabilities, Theorem 1 places restrictions on the set of deterministic payoffs $x'_{aih}\beta$ that are consistent with the data. Without additional assumptions than made in Theorem 1, the identified set of parameter vectors is

$$\mathcal{B} = \left\{\beta \in \Theta \mid x'_{aih}\beta + x'_{bjh}\beta > x'_{ajh}\beta + x'_{bih}\beta \text{ when } P(ai, bj \mid X_h) > P(aj, bi \mid X_h)\ \forall X_h, a, b, i, j\right\}. \tag{5}$$

Without any restrictions on X_h, I can only prove that this set $\mathcal{B}$ exists, and that it is not the entire space Θ of theoretically possible parameters. In other words, β is set-identified, and we can use Theorem 1 to identify bounds on β, or the boundaries of $\mathcal{B}$. Note that the outcomes of discrete choice models (in this case, matches) are qualitative and have no natural cardinalization. Therefore, Θ will impose location and scale normalizations, as only the parameter vector $\beta / \|\beta\|$ is identifiable from discrete matches.

Most applied economists prefer to report point estimates rather than estimates of sets. Manski (1975), Manski (1988) and other authors have discussed the semiparametric point-identification of discrete choice models, where semiparametric means that the distribution of the stochastic error terms ε_{hE} are not specified. This section follows that tradition by showing sufficient conditions on the variation in the data that allow point identification of the parameters. Point identification means that the identified set $\mathcal{B}$ is a single vector, and that vector is β^0, the parameter that generates the data.

The mathematical argument for point identification focuses on the detailed covariates of only one quartet of firms in each market. To first-time readers of this paper, this may seem shocking, but the intuition follows naturally from the intuition for the identification of single-agent discrete choice models, as I will discuss below. To this end, make the following assumption about the identities of the relevant quartet and the corresponding variation in the observable data. I first need to split the vector of characteristics x_{ai} entering into the total match payoff of upstream firm a and downstream firm i into $x_{ai} = (x_{1,ai}, x_{-1,ai})$, where $x_{1,ai}$ is the first component of the vector, and $x_{-1,ai}$ is all other covariates.

Assumption 2. *For every matching market, the econometrician sees the market characteristics X_h, and in particular observes two particular upstream firms a and b, and two particular downstream firms i and j. The joint distribution of the vectors of the characteristics entering into match surpluses is $g\left(x_{ai}, x_{aj}, x_{bi}, x_{bj}\right)$. Furthermore, define the random variable $y \equiv x_{1,ai} + x_{1,bj} - x_{1,aj} - x_{1,bi}$, where $x_{1,ai}$ is the first covariate of the vector of characteristics.*

- *The random variable y is assumed to have a continuous conditional density $g\left(y \mid x_{-1.ai}, x_{-1.aj}, x_{-1.bi}, x_{-1.bj}\right)$ with positive support on the entire real line and no mass points.*
- *The parameter β_1 on characteristic 1 is nonzero.*
- *The support of the distribution of the entire set of characteristics $\left\{x_{ai}, x_{aj}, x_{bi}, x_{bj}\right\}$ does not lie in a proper linear subspace of $\mathbb{R}^d$, where $d = \dim\left\{x_{ai}, x_{aj}, x_{bi}, x_{bj}\right\}$.*

The sampling rule for the data, g, should be seen as an implication of the sampling rule for the characteristics of all matches in the entire market X_h. This includes whatever rule is being used to assign firms in different markets to the slots of the abstract firms a, b, i and j. The special random variable y is assumed to be freely varying conditional on the other characteristics of the matches. The existence of such a freely varying covariate is required for point identification of semiparametric discrete choice models (Manski, 1988; Horowitz, 1998).

Intuitively, the support condition for y means there exist a continuum of moment restrictions (one for each value of the characteristics), and moment restrictions that are relevant for every potential value of the unknown parameters β. In the case of two-sided matching, the number of possible matches in an entire matching market is large, but still finite. Itemizing over the entire set of possible match quartets only provides the finite number of inequality moments from Theorem 1. On the other hand, adding additional observations with new continuous characteristics y from an infinite number of new markets (the exercise in identification) creates a continuum of restrictions from Theorem 1. Thus, semiparametric point identification takes advantage of continuously varying covariates such as y and does not require examination of the entire set of possible matches.

Assumption 2 states that the vector of the characteristics of all firms and matches X_h is observable. This assumption is made to conceptualize the set of observationally equivalent markets in order to observe quartet match probabilities $P(ai, bj \mid X_h)$. Section 3.4 argues that, in the case of the quota of matches each firm can make, the observability of all elements of X_h is a convenience, and not a necessity.

Identification is stated in the following theorem, and the theorem is proved in Appendix A.4.

Theorem 2. *Under the assumptions of Theorems 1 and and Assumption 2, the true parameter β^0 from the data generating process is uniquely identified up to the choice of location and scale normalizations in the set Θ.*

3 Semiparametric Estimation of Total Match Payoffs

The previous section shows that the parameters multiplying the observable characteristics in total match payoffs are semiparametrically identified under the assumptions of i.i.d. stochastic payoffs across marketwide assignments and the availability of a particular freely varying covariate. By following the identification argument, one can construct a potentially consistent estimator of the parameter in the data generating process, β^0. The researcher nonparametrically estimates quartet match probabilities $P(ai, bj \mid X_h)$ by using data across markets. Then the researcher uses the estimates of quartet match probabilities to estimate $\mathcal{B}$, the identified set using the conditions of Theorem 1. As data on more markets appear, the estimate of $\mathcal{B}$ converges to the true parameter β^0, if the conditions for identification are met, and under possible additional regularity conditions.

Given that X_h may have millions of elements, the dimensionality of $P(ai, bj \mid X_h)$ means that nonparametric estimation is not a tractable strategy for typical datasets. This section provides a more practical maximum score estimator. The maximum score estimator works directly with the parameter vector β, and does not involve auxiliary nonparametric estimates or estimates of sets. A small downside is that consistency of the maximum score estimator requires somewhat stronger properties for the covariates. In particular, most of Assumption 2 must hold for all quartets of firms that enter the objective function, rather than just the one quartet needed for identification in Theorem 2.

3.1 The Assignment Maximum Score Estimator for One-to-One and One-to-Many Matching

First consider the case of one-to-one and one-to-many matching. Define the assignment maximum score estimator to be any parameter vector $\beta \in \Theta$ that maximizes the objective function

$$Q_H(\beta) = \frac{1}{H} \sum_{h=1}^{H} \sum_{a \in U_h} \sum_{b=a+1}^{U_h} \sum_{i \in D_h} \sum_{j \in D_h, j \neq i} 1[ai, bj \text{ match in } h]\, 1\left[x'_{aih}\beta + x'_{bjh}\beta > x'_{ajh}\beta + x'_{bih}\beta\right], \tag{6}$$

where H is the number of such markets observed by the econometrician, U_h is the set of upstream firms in market h, and likewise D_h is the set of downstream firms in market h. The term U_h is also the number of upstream firms in market h, and is included to prevent duplicate quartets of firms from appearing in the objective function. The terms $1[\cdot]$ are indicator functions equal to 1 when the condition in brackets is true, and 0 otherwise. The dependent variable from the assignment game is $1[ai, bj \text{ match in } h]$, which is equal to 1 when firms a and i and also firms b and j match simultaneously in market h.

Consider the case of one-to-one matching. A non-stochastic notion of social optimality implies that if upstream firm a matches with downstream firm i, and upstream firm b matches with downstream firm j, then the sum the payoffs for the observed matches must be greater than the payoffs from the quartet exchanging partners. Theorem 1 extends local social optimality to the stochastic case where there are i.i.d. error terms across marketwide assignments. If the local social optimality condition is met for an observed pair of matches at some trial vector of parameters β, the score of correct predictions within the quartet increases by 1. The assignment maximum score estimator is any vector of payoff parameters that receives the highest score for not violating predictions of Theorem 1's version of local social optimality for observed match quartets. As the objective function is a step function, there will always be more than one global maximum; finding one is sufficient for estimation.

As proved below, maximizing $Q_H(\beta)$ produces a consistent estimator of the true population parameter vector β^0. In practice, one uses a numerical optimization package to compute a maximum of the objective function. In a finite sample, only the observed match quartets make positive contributions to the objective function. However, in many applications the number of observed match quartets will be so large that the evaluation of the function $Q_H(\beta)$ will be too computationally expensive for this estimator to be practical. For example, if there are 1000 upstream firms and 1000 downstream firms in a one-to-one matching market without the option of remaining unmatched, there are 1000 observed matches and the number of distinct quartets making positive contributions to the objective function is $999+998+\ldots+1$, or 499,500.

Fox (2005) proves that a single-agent multinomial maximum score estimator is consistent when using only a subset of all of the theoretical model's choices in estimation. Here, I extend the single-agent results to show that the assignment games maximum score estimator is consistent when only a subset of potential matches are considered in estimation. Given that the identification argument in Section 2.7 requires data on only one quartet per market, it is not too surprising that subset estimation is possible with an appropriately constructed estimator. Note that even for a small number of choices, evaluation of the matching maximum score objective function will be much quicker computationally than nested solution estimators, which require the nested solution to a matching algorithm for every parameter vector and vector of stochastic payoff terms, and to date do not allow for endogenous prices. The maximum score objective function involves only multiplication, addition and pairwise comparisons.

The assignment games maximum score estimator using only a subset of potential matches in estimation is defined to be any parameter vector $\beta \in \Theta$ that maximizes

$$Q_H^{\text{sub}}(\beta) = \frac{1}{H}\sum_{h\in H}\sum_{a\in U_h^{\text{sub}}}\sum_{b=a+1,b\in U_h^{\text{sub}}}^{U_h}\sum_{i\in D_h^{\text{sub}}}\sum_{j\in D_h^{\text{sub}},j\neq i} 1\left[ai,bj \text{ match in } h\right] 1\left[x'_{aih}\beta + x'_{bjh}\beta > x'_{ajh}\beta + x'_{bih}\beta\right], \quad (7)$$

where again U_h and D_h are the sets of upstream and downstream firms in a matching market h, and U_h^{sub} and D_h^{sub} are subsets of those firms that have been arbitrarily selected by the econometrician to enter the objective function in order to reduce the computational cost of evaluating the objective function.[11] If $U_h^{\text{sub}} = U_h$ and $D_h^{\text{sub}} = D_h$, the subset estimator simplifies to the non-subset matching maximum score estimator introduced above. It is easiest to think of the subset maximum score objective function as simply the full maximum score

[11] The summation $\sum_{b=a+1,b\in U_h^{\text{sub}}}^{U_h}$ means that terms not in U_h^{sub} are skipped in computation of the subset objective function.

objective function with some match quartets dropped. If the subsets U_h^{sub} and D_h^{sub} are fixed (as a function of market characteristics X_h), I prove below that the matching maximum score estimator using only a subset of potential match quartets in estimation consistently estimates β^0.

To clarify, the subset matching maximum score estimator does not rely on randomly sampling an estimation choice set for each agent, nor computing choice probabilities conditioning on a subset of choices, as the single-agent logit sampling estimator of McFadden (1978) does. Instead, for all firms (in a market captured by X_h) there is one common set of quartets that enters the objective function. Some pairwise comparisons between the deterministic payoffs of quartet matches are excluded for computational or data reasons, and the identification argument in Section 2.7 shows that identification does not require covariates data on all quartets.

3.2 The Assignment Maximum Score Estimator for Many-to-Many Matching

The estimator for many-to-many matching is somewhat more complex, as the estimator must rule out assignments where there are three or four matches for quartet members. The assignment games estimator for many-to-many matching is any parameter vector $\beta \in \Theta$ that maximizes

$$\begin{aligned} Q_H^{\text{sub}}(\beta) \quad = \quad & \frac{1}{H} \sum_{h \in H} \sum_{a \in U_h^{\text{sub}}} \sum_{b=a+1, b \in U_h^{\text{sub}}}^{U_h} \sum_{i \in D_h^{\text{sub}}} \sum_{j \in D_h^{\text{sub}}, j \neq i} 1\left[\text{in}\, h;\ ai, bj \text{ match},\ aj\, \text{do not match},\ bj\, \text{do not match}\right] \\ \cdot \quad & 1\left[x'_{aih}\beta + x'_{bjh}\beta > x'_{ajh}\beta + x'_{bih}\beta\right]. \end{aligned} \tag{8}$$

Corollary 1 does not consider the situations where more than two matches happen within the quartet, so the assignment games maximum score estimator must not either.

Note that the many-to-many objective collapses to equation (7) in the case of one-to-one or one-to-many matching. Therefore, the many-to-many objective function can be considered to be more general.

3.3 Consistency

To apply a lemma from asymptotic theory,[12] we need the following assumptions about the data generating process.

Assumption 3. *The number of possible matching quartets does not have an infinite mean across markets. Further the distribution $G(X_h)$ of observable covariates, including the numbers of upstream and downstream firms, is identical and independent across markets.*

It is a standard technical requirement for extremum estimators that the parameter space must be compact.

Assumption 4. *The parameter vector β is known to lie in a compact space Θ, which also imposes location and scale normalizations.*

[12] Lemma 2.4 from Newey and McFadden (1994), which appears in the proof of Theorem 3.

Following Assumption 2, one suggestion for the scale normalization is to normalize the parameter β_1, which multiplies the components of continuously varying characteristic for a specified quartet y and is already assumed to be nonzero, to values of ± 1. If the sign of β_1 is not known from economic theory, it can be superconsistently estimated by estimating the model twice, once where β_1 is fixed at -1, and once where β_1 is fixed at 1. The final estimates for all parameters correspond to the sign of β_1 with the highest objective function value.

Consistency of the maximum score estimator requires stronger assumptions than identification. Identification requires Assumption 2, which requires there to be a freely varying function of covariates y for at least one quartet of firms. Consistency requires that such a y exist for all quartets that have been entered into the objective function. For any quartet, there are likely to be markets such that $x'_{ai}\beta^0 + x'_{bj}\beta^0 = x'_{aj}\beta^0 + x'_{bi}\beta^0$. Without the existence of such a y for one quartet, there may be a positive probability for the set of markets where, at the true parameter vector, the total deterministic payoffs from one set of match equals the payoffs from the exchange of partners. Markets where $x'_{ai}\beta^0 + x'_{bj}\beta^0 = x'_{aj}\beta^0 + x'_{bi}\beta^0$ do not make positive contributions to the objective function evaluated at the true parameter value. Choosing some alternative parameter value, $\tilde{\beta} \in \Theta$, may make these markets contribute positively to the objective function.[13] If so, it is possible that $\tilde{\beta}$ maximizes the probability limit of the objective function, which violates a condition for proving consistency of the maximum score estimator.

With this argument in mind, I make the following covariate assumption.

Assumption 5. *The conditions of Assumption 2 hold for every quartet of two upstream and two downstream firms that enters the objective function. One exception is that data on whether a firm may remain unmatched, or on the total quota of matches that a firm may make, is no longer required to be observable for any quartet. Another exception is that covariate data not entering the objective function for computational or pure unavailability reasons is also not required to be observable.*

If the researcher knows that, ex ante, some quartets have covariates with only discrete distributions, for example, then the researcher can ensure consistency by excluding those problematic quartets. Adding more quartets to the model may make the estimator inconsistent if the quartets themselves are not sufficient for identification. This lack of consistency is a property of the maximum score objective function, and does not affect the identification arguments in Section 2.7. Section 3.4 discusses quotas in more detail.

The following theorem states that semiparametric assignment maximum score estimator is consistent, including when a subset of possible match quartets are used in estimation.

Theorem 3. *Under the assumptions of Theorem 1 and Corollary 1, as well as Assumptions 3, 4 and 5, any argument $\beta_H \in \Theta$ that maximizes the subset assignment games maximum score objective function, equation (8), is a consistent estimator for β^0, the true parameters in the total deterministic payoff of a match.*

As discussed above, the many-to-many objective function, equation (8), is a generalization of equation (7).

[13] The specific argument relies on the fact that strict inequalities, $>$, enter the maximum score objective function. If weak inequalities, $\geq$, entered the maximum score objective function instead, markets where $x'_{ai}\beta^0 + x'_{bj}\beta^0 = x'_{aj}\beta^0 + x'_{bi}\beta^0$ would make contributions of $P(ai, bj \mid X_h) + P(aj, bi \mid X_h)$ to the probability limit of the objective function. An alternative parameter vector $\tilde{\beta} \in \Theta$ that maximized the instances of such equalities might raise the total objective function value, and be a global maximum of the probability limit of the objective function. See the proof of Theorem 3 for more details about proving the probability limit of the objective function has a unique maximum at the true parameter value, β^0.

The proof is in Appendix A.4. The most economically interesting part of the proof proves the true parameter value β^0 maximizes the probability limit of the objective function. The probability limit has many terms that look like $P(ai,bj \mid X_h) \cdot 1\left[x'_{ai}\beta + x'_{bj}\beta > x'_{aj}\beta + x'_{bi}\beta\right]$. Each quartet in the estimation subset appears twice: once as just listed, and once with the matches within the quartet exchanged, leaving $P(aj,bi \mid X_h) \cdot 1\left[x'_{aj}\beta + x'_{bi}\beta > x'_{ai}\beta + x'_{bj}\beta\right]$. With the support condition, Assumption 5, ties in the indicator functions happen with probability 0. So for any parameter vector, one of the indicator variables must be 1, and the other then become 0. The highest objective function value can be achieved by having the larger of the two probabilities multiply the 1, and the smaller of the two probabilities multiply the 0. Therefore, a global maximum of the function is found by using Theorem 1 to show that the true parameter vector β_0, which generates $P(ai,bj \mid X_h)$, implements this assignment. Note that nowhere in the argument does the itemization of all matches play a role, so the estimator is consistent when only a subset of matches are included for computational or data unavailability reasons.

3.4 Quotas

A quota is a fixed number of other agents that an agent may match with. As a requirement of nesting a matching mechanism into a parametric estimator, a researcher must make often unverifiable assumptions about the size of the quota of each agent in their estimation sample. Sørensen (2004) assumes that all agents (venture capitalists, in his example) use all of their quota, so the quota is equal to the number of observed matches for each venture capitalist. Boyd et al. (2003) study the hiring of public school teachers, and argue that state laws mandate that a fixed number of teachers must be hired based upon an exogenously specified number of students attending a school.

By contrast, the method of estimating based upon properties of the solution considered in this paper does not require that the econometrician specify the quota of each agent. The estimator only compares exchanging two matches at a time. Given that the existing number of matches of any firm is under its quota, switching its matching partner does not increase the number of partners it is matching with, and therefore does not violate any quota.

A subtlety is that the identification argument in Section 2.7 requires knowledge of quotas. The reason is that the quotas of firms affect quartet match probabilities, and such probabilities appear in the identification argument based upon Theorem 1. As part of the proof of consistency, I show that the probability limit of the maximum score objective function has a unique global maximum at the true parameter value, β^0. The consistency proof can be seen as an alternative, constructive proof of semiparametric identification. As consistency does not require knowledge of the exact values for quotas, the assumptions of Theorem 2 are stronger than required. A weaker, constructive version of Theorem 2 can be proved by using Theorem 1 and Assumption 2 to prove that $\beta^0 \in \Theta$ is the unique global maximum of

$$E_{X_h}\left\{P(ai,bj \mid X_h,\beta_0)\,1\left[x'_{aih}\beta + x'_{bjh}\beta > x'_{ajh}\beta + x'_{bih}\beta\right] + P(aj,bi \mid X_h,\beta_0)\,1\left[x'_{ajh}\beta + x'_{bih}\beta > x'_{aih}\beta + x'_{bjh}\beta\right]\right\},$$

which is just the probability limit of the maximum score objective function when using data on only one quartet of firms. This objective function can be estimated by a maximum score objective function, where data on X_h, such as quotas, do not enter the objective function directly. The constructive identification argument is a special case of the proof of Theorem 3.

3.5 Relaxing i.i.d. Errors Through Nest-Specific Fixed Effects

The original paper by Manski (1975) on multinomial maximum score estimation relies on the assumption of i.i.d. error across choices for a given agent, but not across agents. The two-sided matching maximum score estimators in this paper rely on an assumption of i.i.d. errors as well. The assumption of i.i.d. errors is particularly unappealing in matching contexts where marketwide assignments can be identical except for the exchange of partners within a quartet. It might be reasonable to suspect that the random error terms are correlated across similar marketwide assignments. Also, it is reasonable to suspect that individual agents might have unobserved payoff terms that vary across match partners. The assumption of marketwide errors, needed for the i.i.d. across marketwide outcomes property, prevents match-specific errors. Further, match-specific errors might be correlated with partner characteristics, so that higher-quality firms also have higher levels of unmeasured quality. This endogeneity problem goes further than just allowing for correlation of errors across partners.

Luckily, a researcher can consistently estimate the parameters in the linear index while allowing for agent-specific fixed effects that are constant across nests specified by the researcher. The fixed effects can be correlated with included covariates. Identification and estimation then proceeds by comparing alternative match partners within the same nest, where the fixed effect is held constant and does not affect the relative ranking of alternative match partners. This within estimator was proposed for the case of single-agent ordered choice models using panel data by Abrevaya (2000), but the argument extends to the case of cross-sectional nests in single-agent multinomial choice maximum score and the two-sided matching models considered in this paper.

For a assignment market with prices h, let there be a set of nests for upstream firms $\mathcal{N}_h^{\text{up}}$, and let the corresponding set of nests for downstream firms be $\mathcal{N}_h^{\text{down}}$. Let n_h^{up} be an individual nest for upstream firms, and likewise let n_h^{down} be a nest of downstream firms. In an abuse of notation, let $n_h^{\text{down}}(i)$ be a function that gives the nest of downstream firm i. An upstream firm a receives a payoff of

$$x_{aih}^{\text{up}\prime}\beta^{\text{up}} + \xi^{\text{up}}_{an_h^{\text{down}}(i)} + p_{aih}$$

for matching with downstream firm i, where $\xi^{\text{up}}_{an_h^{\text{down}}(i)}$ is a's fixed effect for matching with downstream firms in the $n_h^{\text{down}}(i)$ nest. There is a symmetric payoff function for downstream firms.

Consider the quartet of upstream firms a and b and downstream firms i and j. Assume that a and b are in the same nest, and i and j are in the same nest. As an example, consider the payoffs inequality Theorem 1, a version of local social optimality. After allowing for fixed effects, local social optimality becomes

$$x_{aih}'\beta + \xi^{\text{up}}_{an_h^{\text{down}}(i)} + \xi^{\text{down}}_{in_h^{\text{up}}(a)} + x_{bjh}'\beta + \xi^{\text{up}}_{bn_h^{\text{down}}(j)} + \xi^{\text{down}}_{jn_h^{\text{up}}(b)} > x_{ajh}'\beta + \xi^{\text{up}}_{an_h^{\text{down}}(j)} + \xi^{\text{down}}_{jn_h^{\text{up}}(a)} + x_{bih}'\beta + \xi^{\text{up}}_{bn_h^{\text{down}}(i)} + \xi^{\text{down}}_{in_h^{\text{up}}(b)}$$

Under the assumption that a and b as well as i and j are in the same nests, the inequalities on either side of the above inequality are the same and therefore cancel, leaving the original deterministic notion of local social optimality in Theorem 1. Thus, by looking within nests, a researcher can identify the unknown total surplus parameters β using within-nest variation in covariates, while allowing the unobserved payoffs of firms to be correlated with covariates, and to be correlated across similar match partners.[14]

The fixed effects approach is very powerful, but there are two major downsides. First, the method is only consistent if the researcher correctly specifies the nests. Second, the inclusion of fixed effects means that the researcher cannot identify the parameters on covariates that do not vary within nests. It should be noted that these two drawbacks also apply to the use of fixed effects in linear regression models, and are not unique to two-sided matching models.

4 Subsampling for Asymptotic Inference

Aside from the original work of Manski (1975) and a few others such as Matzkin (1993) and Fox (2005), the single-agent maximum score literature has focused on the binary-choice estimator.

A typical approach in a nonlinear estimation problem is to derive the variance-covariance matrix of the limiting normal distribution of the estimator, and to use a consistent estimator of the matrix for inference. In cases where either the derivation or the estimation of the limiting distribution is difficult, researchers often employ a resampling procedure known as the bootstrap to construct asymptotically valid inference.

Neither estimating the asymptotic distribution or using the bootstrap are currently useful methods for inference for maximum score estimators. Kim and Pollard (1990) show that the binary-choice maximum score estimator converges at the rate of $\sqrt[3]{n}$ (instead of the more typical $\sqrt{n}$) and that its limiting distribution is too complex for use in inference. Abrevaya and Huang (2005) show that the typical bootstrap procedure is not consistent in the case of the class of $\sqrt[3]{n}$-consistent estimators studied by Kim and Pollard.

Delgado, Rodríguez-Poo and Wolf (2001) show that an alternative resampling procedure, subsampling, consistently estimates the asymptotic distribution of test statistics for the class of $\sqrt[3]{n}$-consistent estimators studied by Kim and Pollard. Subsampling was developed by Politis and Romano (1994), and is a procedure that, in contrast to the bootstrap, does not rely on the smoothness of an objective function. A key step in subsampling is that artificial datasets with fewer observations than the original data are sampled from the original data. Delgado, Rodríguez-Poo and Wolf provide a Monte Carlo study of the relation between the finite-sample and asymptotic coverage properties of subsampling for the binary-choice maximum score case, which appears to be good, and provide an algorithm for selecting an appropriate number of observations in each of the subsamples.

While the literature has not provided a proper modification of the proofs of Kim and Pollard (1990) and Abrevaya and Huang (2005) for the multinomial maximum score, it is almost surely the case that the assignment game maximum score estimator, under suitable regularity conditions, is a member of the class of $\sqrt[3]{H}$-consistent estimators studied by Kim and Pollard.

An alternative procedure to subsampling is to estimate a smoothed version of the maximum score estimator. For the single-agent binary-choice maximum score estimator, Horowitz (1992) proves that, under additional

[14]Theorem 1 still applies to the full quartet social surplus $x'_{aih}\beta + \xi^{\text{up}}_{an_h^{\text{down}}(i)} + \xi^{\text{down}}_{in_h^{\text{up}}(a)} + x'_{bjh}\beta + \xi^{\text{up}}_{bn_h^{\text{down}}(j)} + \xi^{\text{down}}_{jn_h^{\text{up}}(b)}$, where the fixed effects are treated as nuisance parameters.

smoothness assumptions about the underlying model, a smoothed version converges at a rate close to $\sqrt{n}$ (the exact rate depends on the smoothing parameter) and, more importantly, is asymptotically normal with a variance-covariance matrix than can be estimated and used for inference. Unfortunately, Monte Carlo studies show the finite-sample performance of the asymptotic distribution is poor, and Horowitz (2002) proves the applicability of the bootstrap to refine the estimates of individual components of the variance formula. Horowitz (2002) presents Monte Carlo evidence that the coverage properties of the bootstrap-refined asymptotic distribution approximates the finite-sample distribution well. I conjecture the Horowitz results could be extended to the current two-sided matching estimators.[15]

5 Conclusions

This paper's main purpose is to prove the identification of and introduce a new semiparametric maximum score estimator for Koopmans and Beckmann (1957), Shapley and Shubik (1972), Becker (1973) and Sotomayor (1992) assignment games. The main assumptions for this class of models are the presence of endogenous prices, additive payoffs across multiple matches, and additive separability between transfers and other parts of payoffs. Under these assumptions, linear programming arguments show that all stable assignments of prices and matches are socially optimal. I translate a notion of local social optimality into a statement about matching probabilities under the assumption that there are marketwide i.i.d. errors. This assumption translates the matching market into a single-agent discrete choice problem by a social planner. The theorem states that a pair of matches that together give a higher deterministic sum of payoffs than the payoffs for the pair of matches where the agents exchange partners will be observed more frequently than the exchange of partners. The probabilistic local social optimality condition forms the basis for identification and the consistency of a semiparametric estimator.

The assignment maximum score estimator has many practical advantages over currently used nested solutions methods. The maximum score estimator is semiparametric, meaning that a parametric distribution for the errors does not need to specified. The assignment estimator is for a market with prices, like many matching markets are in practice, but does not use data on prices. The assignment estimator does not require additional assumptions about matching mechanisms and strong assumptions about quotas for individual agents. Computationally, evaluating the maximum score objective is much simpler than evaluating nested solution objective functions, as no stable match needs to be computed. Further, the maximum score estimator is consistent when the number of matches exceeds the number of atoms in the universe, as the maximum score estimator is consistent when a subset of matching quartets are entered into the objective function. Finally, a researcher can weaken the i.i.d. errors assumption by identifying and estimating unknown model parameters in the presence of agent-specific fixed effects for specified nests of matching partners.

[15] Smoothing the maximum score step function does not solve the main issue in the computational cost of numerically maximizing the objective function: the presence of local hills providing tempting regions for a greedy optimization routine to converge to.

A Proofs

A.1 Theorem 1 (Quartet Match Probabilities)

Consider the state of being unmatched as a partner (who specially can make multiple matches) if it is included in the underlying model. Also, I drop the market index h on payoff terms in what follows.

A.1.1 One-to-One Matching

Consider the case of one-to-one matching. I first derive an explicit formulation for choice probabilities in terms of the density and distribution function for the i.i.d. errors. The condition for an assignment E_h to be optimal is seen in equation (3). Writing this out in more detail gives

$$
\begin{aligned}
P(E \mid X_h) &= \int_{-\infty}^{\infty} \left\{ \prod_{Z \neq E} \int_{-\infty}^{\sum_{ck\in E} x'_{ck}\beta - \sum_{ck\in Z} x'_{ck}\beta + \varepsilon_E} f(\varepsilon_Z \mid X)\, d\varepsilon_Z \right\} f(\varepsilon_E \mid X)\, d\varepsilon_E \\
&= \int_{-\infty}^{\infty} \left\{ \prod_{Z \neq E} F\left(\sum_{ck\in E} x'_{ck}\beta - \sum_{ck\in Z} x'_{ck}\beta + \varepsilon_E \mid X \right) \right\} f(\varepsilon_E \mid X)\, d\varepsilon_E.,
\end{aligned}
$$

where the first equality integrates out over all marketwide error terms other than ε_E, and the second equality uses the fact that the integral of a density function with a non-infinite upper limit is the distribution function evaluated at the upper limit.

Writing out the definition of a quartet matching probability, equation (3), gives

$$
\begin{aligned}
P(ai, bj \mid X_h) &= \sum_{E \in \mathcal{E}(ai,bj)} \mathrm{Prob}_{\varepsilon_h}\left(E_h \text{ solves social planner's discrete choice problem}\right) \\
&= \sum_{E \in \mathcal{E}(ai,bj)} \int_{-\infty}^{\infty} \left\{ \prod_{Z \neq E} F\left(\sum_{ck\in E} x'_{ck}\beta - \sum_{ck\in Z} x'_{ck}\beta + \varepsilon_E \mid X \right) \right\} f(\varepsilon_E \mid X)\, d\varepsilon_E \\
&= \int_{-\infty}^{\infty} \left\{ \sum_{E \in \mathcal{E}(ai,bj)} \prod_{Z \neq E} F\left(\sum_{ck\in E} x'_{ck}\beta - \sum_{ck\in Z} x'_{ck}\beta + \varepsilon \mid X \right) \right\} f(\varepsilon \mid X)\, d\varepsilon \\
&= \int_{-\infty}^{\infty} \left\{ \sum_{E \in \mathcal{E}(ai,bj)} \prod_{Z \neq E} F\left(x'_{ai}\beta + x'_{bj}\beta + \sum_{ck\in E\setminus\{ai,bj\}} x'_{ck}\beta - \sum_{ck\in Z} x'_{ck}\beta + \varepsilon \mid X \right) \right\} f(\varepsilon \mid X)\, d\varepsilon
\end{aligned}
$$

where the third equality uses the fact that ε is i.i.d. across marketwide assignments E, and the fourth equality uses the fact that all assignments in $P(ai, bj \mid X_h)$ have a matching with i and b matching with j. By a symmetric argument,

$$
P(aj, bi \mid X_h) = \int_{-\infty}^{\infty} \left\{ \sum_{E \in \mathcal{E}(aj,bi)} \prod_{Z \neq E} F\left(x'_{aj}\beta + x'_{bi}\beta + \sum_{ck\in E\setminus\{aj,bi\}} x'_{ck}\beta - \sum_{ck\in Z} x'_{ck}\beta + \varepsilon \mid X \right) \right\} f(\varepsilon \mid X)\, d\varepsilon.
$$

In one-to-one matching, for every marketwide assignment where a matches with j and b matches with i, there is another assignment where a matches with i and b matches with j, and all other matches not involving the quartet mentioned in the statement of the theorem are the same. Therefore, for each assignment in $\mathcal{E}(ai,bj)$, there is an equivalent assignment in $\mathcal{E}(aj,bi)$ with the same terms $\sum_{ck\in E\setminus\{aj,bi\}} x'_{ck}\beta$.

For $P(ai,bj \mid X_h)$, some of the alternatives Z are in $\mathcal{E}(aj,bi)$. For those alternatives, the sum $x'_{aj}\beta + x'_{bi}\beta$ enters strictly negatively. There are usually many assignments not in either $\mathcal{E}(ai,bj)$ or $\mathcal{E}(aj,bi)$. These assignments appear as alternatives in both $P(ai,bj \mid X_h)$ and $P(aj,bi \mid X_h)$.

The above arguments show that the functional forms of $P(ai,bj \mid X_h)$ and $P(aj,bi \mid X_h)$ are the same, except for where the sums $x'_{ai}\beta + x'_{bj}\beta$ and $x'_{aj}\beta + x'_{bi}\beta$ enter. For $P(ai,bj \mid X_h)$, the distribution functions F are monotonically increasing in the sum $x'_{ai}\beta + x'_{bj}\beta$, and monotonically decreasing in the sum $x'_{aj}\beta + x'_{bi}\beta$. Because of the continuous support with no mass points part of Assumption 1, $P(ai,bj \mid X_h)$ is strictly increasing in $x'_{ai}\beta + x'_{bj}\beta$ and strictly decreasing in $x'_{aj}\beta + x'_{bi}\beta$. Likewise, $P(aj,bi \mid X_h)$ is strictly increasing in $x'_{aj}\beta + x'_{bi}\beta$ and strictly decreasing in $x'_{ai}\beta + x'_{bj}\beta$.

If two functions are the same, except for components in the first function resulting in a larger value, the first function will have a larger value. Therefore, if $x'_{ai}\beta + x'_{bj}\beta > x'_{aj}\beta + x'_{bi}\beta$, then $P(ai,bj \mid X_h) > P(aj,bi \mid X_h)$. Likewise, the "only if" portion of the theorem is proved because the only way an otherwise identical function can have a larger value is if it has a larger argument, as it does when $x'_{ai}\beta + x'_{bj}\beta > x'_{aj}\beta + x'_{bi}\beta$.

A.1.2 Many-to-One Matching

For many-to-one matching, without loss of generality assume that downstream firms, such as i and j from the theorem, can make only one physical match at a time. Each upstream firm can make one or more matches. If a can make at least two matches, there are now assignments where both i and j match with a, and therefore neither i nor j matches with b. These assignments are not in either $\mathcal{E}(ai,bj)$ or $\mathcal{E}(aj,bi)$, and so appear as alternatives in both $P(ai,bj \mid X_h)$ and $P(aj,bi \mid X_h)$. Therefore, the functional forms of $P(ai,bj \mid X_h)$ and $P(aj,bi \mid X_h)$ are still the same except where the sums $x'_{ai}\beta + x'_{bj}\beta$ and $x'_{aj}\beta + x'_{bi}\beta$ from the statement of the theorem enter, and the earlier arguments still apply.

A.2 Proof of Corollary 1 (Many-to-Many Matching)

Now both upstream firms and downstream firms can make multiple matches. Compared to one-to-one and many-to-many matching, there are additional marketwide assignments where a matches with i and b matches with j at the same time as a matches with j as well, for example. This allocation is in $\mathcal{E}(ai,bj)$ but not $\mathcal{E}(aj,bi)$. However, the definition of $\bar{P}(ai,bj \mid X_h)$ implies that these assignments are excluded from the relevant set

$$\mathcal{E}_h(ai,bj) \cap \mathcal{E}_h(ai,bj,aj) \cap \mathcal{E}_h(ai,bj,bi).$$

Likewise, these same assignments where more than two matches are made in the quartet are excluded from $\bar{P}(aj,bi \mid X_h)$. Therefore, by the arguments in the proof of Theorem 1, the functional forms of $\bar{P}(ai,bj \mid X_h)$ and $\bar{P}(aj,bi \mid X_h)$ are the same except where the sums $x'_{ai}\beta + x'_{bj}\beta$ and $x'_{aj}\beta + x'_{bi}\beta$ appear in the total payoffs of assignments where where only two matches within the quartet are observed and all four members of the quartet match. Therefore, by earlier arguments, the corollary is true.

A.3 Theorem 2 (Identification)

I will write the proof for the case of one-to-one and one-to-many matching. For many-to-many matching, the proof is the same, except that I replace references to Theorem 1 with references to Corollary 1.

I will choose convenient location and scale normalizations for the proof, which I can do without further loss of generality. For the location normalization, I assume I will not try to identify a constant term in payoffs. Following Assumption 2, for the scale normalization I will assume that the parameter β_1, which multiplies the components of y and is already assumed to be nonzero, has a value of ± 1. If the sign of β_1 is not known from economic theory, it can be trivially identified, by seeing whether the match quartet probabilities are consistent with $\beta_1 = +1$ or $\beta_1 = -1$.

We want to show that the identified set $\mathcal{B} \subseteq \Theta$ is a singleton vector, using the data on the observed quartet match probabilities $P(ai, bj \mid X_h)$. Assume to the contrary. Then there is a $\tilde{\beta} \in \Theta$ such that $\tilde{\beta} \neq \beta^0$. I will state the proof for $\beta_1 = 1$. The argument for $\beta_1 = -1$ is symmetric.

Theorem 1 shows that, within the quartet of a, b, i and j, the pair of matches with the higher deterministic social surplus are more likely to be observed than the exchange of partners. As β^0 generates $P(ai, bj \mid X_h)$, it is in $\mathcal{B}$ and is consistent with the implications of Theorem 1 for the observable quartet match probabilities. I introduce the set of markets where the two parameter vectors, $\tilde{\beta}$ and β^0 give different implications about the rank ordering of the match probabilities for the quartet of a, b, i and j, using the rank orderings generated by Theorem 1,

$$\begin{aligned} S_{abij}\left(\beta^0\tilde{\beta}\right) &= \left\{X_h \mid x'_{ai}\beta^0 + x'_{bj}\beta^0 > x'_{aj}\beta^0 + x'_{bi}\beta^0 \cup x'_{aj}\tilde{\beta} + x'_{bi}\tilde{\beta} > x'_{ai}\tilde{\beta} + x'_{bj}\tilde{\beta}\right\} \\ &\cup \left\{X_h \mid x'_{aj}\beta^0 + x'_{bi}\beta^0 > x'_{ai}\beta^0 + x'_{bj}\beta^0 \cup x'_{ai}\tilde{\beta} + x'_{bj}\tilde{\beta} > x'_{aj}\tilde{\beta} + x'_{bi}\tilde{\beta}\right\}, \end{aligned}$$

where the union sign, $\cup$, means that both inequalities are true. Using the definition of y in Assumption 2, one can treat separately the first element of the vector of covariates, and some algebra shows the set can be rewritten as

$$\begin{aligned} S_{abij}\left(\beta^0,\tilde{\beta}\right) &= \left\{X_h \mid x'_{-1,aj}\tilde{\beta}_{-1} + x'_{-1,bi}\tilde{\beta}_{-1} - x'_{-1,ai}\tilde{\beta}_{-1} - x'_{-1,bj}\tilde{\beta}_{-1} > y > x'_{-1,aj}\beta^0_{-1} + x'_{-1,bi}\beta^0_{-1} - x'_{-1,ai}\beta^0_{-1} - x'_{-1,bj}\beta^0_{-1}\right\} \\ &\cup \left\{X_h \mid x'_{-1,aj}\tilde{\beta}_{-1} + x'_{-1,bi}\tilde{\beta}_{-1} - x'_{-1,ai}\tilde{\beta}_{-1} - x'_{-1,bj}\tilde{\beta}_{-1} < y < x'_{-1,aj}\beta^0_{-1} + x'_{-1,bi}\beta^0_{-1} - x'_{-1,ai}\beta^0_{-1} - x'_{-1,bj}\beta^0_{-1}\right\}. \end{aligned}$$

Here is the subvector comprising the parameters that do not multiply the first element of the vector of covariates. For any set of $x_{-1,aj}$, $x_{-1,bi}$, $x_{-1,ai}$ and $x_{-1,bj}$ such that $x'_{-1,aj}\tilde{\beta}_{-1} + x'_{-1,bi}\tilde{\beta}_{-1} - x'_{-1,ai}\tilde{\beta}_{-1} - x'_{-1,bj}\tilde{\beta}_{-1} \neq x'_{-1,aj}\beta^0_{-1} + x'_{-1,bi}\beta^0_{-1} - x'_{-1,ai}\beta^0_{-1} - x'_{-1,bj}\beta^0_{-1}$, we can find a set of combinations of $x_{1,aj}$, $x_{1,bi}$, $x_{1,ai}$ and $x_{1,bj}$ such that the resulting points X_h are in $S_{abij}\left(\beta_0,\tilde{\beta}\right)$. This is due to Assumption 2, which states that $y \equiv x_{1,ai} + x_{1,bj} - x_{1,aj} - x_{1,bi}$ is freely varying conditional on the other covariates, and has full-support on the real line.

Further, this argument shows $S_{abij}\left(\beta_0,\tilde{\beta}\right)$ has positive probability if there is at least one collection of $x_{-1,aj}$, $x_{-1,bi}$, $x_{-1,ai}$ and $x_{-1,bj}$ such that $x'_{-1,aj}\tilde{\beta}_{-1} + x'_{-1,bi}\tilde{\beta}_{-1} - x'_{-1,ai}\tilde{\beta}_{-1} - x'_{-1,bj}\tilde{\beta}_{-1} \neq x'_{-1,aj}\beta^0_{-1} + x'_{-1,bi}\beta^0_{-1} - x'_{-1,ai}\beta^0_{-1} - x'_{-1,bj}\beta^0_{-1}$. While not needed for identification, the positive probability result is necessary for consistency of the maximum score estimator. If there is no such collection, then the covariates belong in a proper linear subspace, which violates another condition in Assumption 2.

A.4 Theorem 3 (Consistency)

The proof of the theorem is based upon the standard consistency theorem in the econometrics literature, Theorem 2.1 in Newey and McFadden (1994). The theorem has four conditions:

1. The probability limit of the subset maximum score objective function, $Q_{\infty}^{\text{sub}}(\beta)$, has a unique global maximum at the true parameter vector, β_0 (constructive identification).
2. The parameter space B is compact.
3. The probability limit of the objective function, $Q_{\infty}^{\text{sub}}(\beta)$, is continuous in β.
4. The objective function converges uniformly in probability to its limit.

Condition 2 is satisfied by Assumption 4.

I will write the proof for the case of one-to-one and one-to-many matching. For many-to-many matching, the proof is the same, except that I replace references to Theorem 1 with references to Corollary 1.

A.4.1 Constructive Identification

The economically interesting condition to verify is Condition 1, which is a constructive identification condition. As the number of markets, H, goes to infinity, we observed infinitely many markets with the same number of firms and identical characteristics, all captured by X_h. By a law of large numbers and the law of iterated expectations,

$$\text{plim}_{H\to\infty}\left(\frac{1}{H}\sum_{h=1}^{H} 1\left[ai, bj \text{ match}\right]\right) = E_{X_h,\varepsilon_h}\left\{1\left[ai, bj \text{ match}\right]\right\} = E_{X_h}E_{\varepsilon_h}\left\{1\left[ai, bj \text{ match}\right] \mid X_h\right\} = E_{X_h}\left\{P\left(ai, bj \mid X_h, \beta_0\right)\right\},$$

where ε_h is the vector of all stochastic terms in the market, and the true parameter vector β_0 has been added to the notation for matching probabilities in order to emphasize that the probability limit is calculated using the sampling rule of the true data generating process. Therefore, the limit of $Q_H^{\text{sub}}(\beta)$ as the number of markets, H, goes to infinity is

$$\begin{aligned} Q_{\infty}^{\text{sub}}(\beta) &= E_{X_h}\left\{\sum_{a\in U_h^{\text{sub}}}\sum_{b=a+1, b\in U_k^{\text{sub}}}^{|U_h|}\sum_{i\in D_h^{\text{sub}}}\sum_{j\in D_h^{\text{sub}}, j\neq i} E_{\varepsilon_h}\left\{1\left[ai, bj \text{ match}\right]\cdot 1\left[x'_{aih}\beta + x'_{bjh}\beta > x'_{ajh}\beta + x'_{bih}\beta\right] \mid X_h\right\}\right\} \\ &= E_{X_h}\left\{\sum_{a\in U_h^{\text{sub}}}\sum_{b=a+1, b\in U_h^{\text{sub}}}^{|U_h|}\sum_{i\in D_h^{\text{sub}}}\sum_{j\in D_h^{\text{sub}}, j\neq i} P\left(ai, bj \mid X_h, \beta_0\right) 1\left[x'_{aih}\beta + x'_{bjh}\beta > x'_{ajh}\beta + x'_{bih}\beta\right]\right\}, \quad (9) \end{aligned}$$

where the first equality uses the a law of large numbers and the law of iterated expectations, and the second equality uses the fact that $1\left[x'_{aih}\beta + x'_{bjh}\beta > x'_{ajh}\beta + x'_{bih}\beta\right]$ does not depend on ε_h and can be factored out of the expectation with respect to ε_h, conditional on X_h.

I prove that $Q_{\infty}^{\text{sub}}(\beta)$ has a global maximum at the true parameter vector β_0 by first proving the integrand evaluated at particular set of the characteristics of all agents in a market, X_h, is globally maximized at β_0. If the integrand is indeed maximized for all X_h, except for a set with probability 0, then when $Q_{\infty}^{\text{sub}}(\beta)$ is computed by integrating out X_h, it will be maximized at β_0.

Therefore, fix X_h. Each matched quartet a, b, i and j formed from elements in U_h^{sub} and D_h^{sub} appears twice in $Q_{\infty}^{\text{sub}}(\beta)$: once as $P(ai, bj \mid X_h, \beta_0)$ times $1\left[x'_{aih}\beta + x'_{bjh}\beta > x'_{ajh}\beta + x'_{bih}\beta\right]$ and once as $P(aj, bi \mid X_h, \beta_0)$ times $1\left[x'_{ai}\beta + x'_{bj}\beta > x'_{aj}\beta + x'_{bi}\beta\right]$. First, under the covariate Assumption 2, $x'_{ai}\beta_0 + x'_{bj}\beta_0 = x'_{aj}\beta_0 + x'_{bi}\beta_0$ with probability 0, as each match has a freely varying characteristic conditional on the other matches. As the inequalities in $Q_{\infty}^{\text{sub}}(\beta)$ are strict, such points do not contribute to the objective function, but as they occur with probability 0, choosing an alternative parameter vector $\tilde{\beta}$ to make one or the other match have a greater surplus will not increase the value of $Q_{\infty}^{\text{sub}}(\beta)$.

I can restrict attention to the cases where one of the sums of match surpluses is strictly greater than the match surplus with the exchange of partners. Notice that the inequalities $x'_{ai}\beta + x'_{bj}\beta > x'_{aj}\beta + x'_{bi}\beta$ and $x'_{aj}\beta + x'_{bi}\beta > x'_{ai}\beta + x'_{bj}\beta$ are mutually exclusive, so one of the two indicator functions has value 1 and the other has value 0. An assignment where the value of 1 multiplies the higher of the two probabilities for all quartets and is a global maximum of the integrand evaluated at X_h. By Theorem 1, the parameter vector β_0 implements this assignment. As X_h arbitrary, the integrand for a given quartet a, b, i and j is globally maximized at all points, other than a set of measure 0, by β_0. As the quartet a, b, i and j is arbitrary, $Q_{\infty}^{\text{sub}}(\beta)$ is globally maximized at β_0.

Note that there is an strong inequality in the indicator function in the objective function, so that $\beta = 0$ is a global minimum and not a global maximum.

The next step of the proof is to show that the global maximum of $Q_{\infty}^{\text{sub}}(\beta)$, β_0, is unique. This argument is the same as the proof of Theorem 2, identification. For some possible other global maximum, $\tilde{\beta} \in \Theta$, the proof of Theorem 2 derives the set $S_{abij}\left(\beta_0, \tilde{\beta}\right)$ where $\tilde{\beta}$ gives implications about the choice probabilities that are inconsistent with Theorem 1. By Assumption 2, this set has positive measure for at least one set of covariates for each market. So $\tilde{\beta}$ implement a sub-optimal series of quartet match probabilities to enter $Q_{\infty}^{\text{sub}}(\beta)$, and thus cannot be a global maximum of $Q_{\infty}^{\text{sub}}(\beta)$.

A.4.2 Continuity of the Limiting Objective Function and Uniform Convergence

The assignment games maximum score objective function is not continuous in β. Condition 3 is that the probability limit of the objective function, $Q_{\infty}^{\text{sub}}(\beta)$, is continuous in β. Lemma 2.4 from Newey and McFadden (1994) can be used to prove continuity of $Q_{\infty}^{\text{sub}}(\beta)$ as well as uniform in probability convergence of $Q_H^{\text{sub}}(\beta)$ to $Q_{\infty}^{\text{sub}}(\beta)$, which is Condition 4. Remember that the asymptotics are in the number of markets. The conditions of Lemma 2.4 are that the data (across markets) are i.i.d., which can hold if we view the number of upstream and downstream firms as random; that the parameter space B is compact (Assumption 4), that the terms for each market are continuous with probability 1 in β; and that the terms for each market are bounded by a function whose mean is not infinite. While the terms for each market,

$$\sum_{a\in U_h^{\text{sub}}}\sum_{b=a+1,b\in U_h^{\text{sub}}}^{|U_h|}\sum_{i\in D_h^{\text{sub}}}\sum_{j\in D_h^{\text{sub}},j\neq i} 1\left[ai,bj \text{ match in } h\right] 1\left[x'_{aih}\beta + x'_{bjh}\beta > x'_{ajh}\beta + x'_{bih}\beta\right], \tag{10}$$

are not continuous in β because of the indicator functions, they are continuous with probability 1 by the support condition on the covariates, Assumption 5. As the continuous covariate $x_{1,aih}$ is freely varying conditional on the other covariates, $x'_{aih}\beta + x'_{bjh}\beta = x'_{ajh}\beta + x'_{bih}\beta$ with probability 0. The other condition we need to verify to apply Lemma 2.4 is that the market-specific terms in equation (10) are bounded by a function with a non-infinite mean. Equation (10) can be at most the number of observed quartets. Assumption 3 states that the mean number of such quartets is not infinite.

References

Abeledo, Hernán and Garth Isaak, "A characterization of graphs that ensure the existence of stable matches," *Mathematical Social Sciences*, August 1991, *22* (1), 93–96.

Abrevaya, Jason, "Rank estimation of a generalized fixed-effects model," *Journal of Econometrics*, 2000, *95*, 1–23.

_ **and Jian Huang**, "On the bootstrap of the maximum score estimator," *Econometrica*, 2005.

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

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff, "Analyzing Determinants of the Matching of Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets," 2003.

Delgado, Miguel A., Juan M. Rodríguez-Poo, and Michael Wolf, "Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator," *Economics Letters*, 2001, *73*, 241–250.

Fox, Jeremy T., "Semiparametric Estimation of Multinomial Discrete Choice Models Using a Subset of Choices," 2005.

Gale, David and Lloyd Shapley, "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 1962, *69.*

Horowitz, Joel L., "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, May 1992, *60* (3), 505–551.

_ , *Semiparametric Methods in Econometrics*, Vol. 131 of *Lecture Notes in Statistics*, Springer, 1998.

_ , "Bootstrap critical values for tests based on the smoothed maximum score estimator," *Journal of Econometrics*, 2002, *111*, 141–167.

Kim, Jeankyung and David Pollard, "Cube Root Asymptotics," *The Annals of Statistics*, 1990, *18* (1), 191–219.

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

Lucas, William F., "Core Theory for Multiple-Sided Assignment Games," *Duke Mathematical Journal*, 1995, *81* (18), 55–65.

Manski, Charles F., "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 1975, *3*, 205–228.

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

Matzkin, Rosa L., "Nonparametric identification and estimation of polychotomous choice models," *Journal of Econometrics*, 1993, *58*, 137–168.

McFadden, Daniel L., "Modelling the Choice of Residential Location," in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, Vol. I, North Holland, 1978, pp. 75–96.

Newey, Whitney K. and Daniel McFadden, "Large Sample Estimation and Hypothesis Testing," in "Handbook of Econometrics," Vol. IV, Elsevier, 1994, chapter 36, pp. 2111–2245.

Ostrovsky, Michael, "Stability in Supply Chain Networks," December 2004.

Politis, Dimitris N. and Joseph P. Romano, "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," *The Annals of Statistics*, December 1994, *22* (4), 2031–2050.

Roth, Alvin E. and Marilda A. Oliveira Sotomayor, *Two-Sided Matching* number 18. In 'Econometric Society Monographs.', Econometric Society, 1990.

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

Sørensen, Morten, "How Smart is Smart Money? An Empirical Two-Sided Matching Model of Venture Capital," December 2004. working paper.

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.