# Estimating Matching Games With Endogenous Prices

Jeremy T. Fox*
University of Chicago

December 9, 2005

**Abstract**

In matching games, agents must all agree for a match to be formed, and some agents can make only a finite number of matches. I examine the nonparametric identification and estimation of match production functions in matching games with endogenous prices and transferable utilities. Inequalities derived from single-agent best responses underly a nonparametric maximum score estimator of match production functions. The inequalities do not require data on prices, quotas, or production levels. The estimator does not suffer from a computational or data curse of dimensionality in the number of agents in a matching market as the estimator avoids solving for an equilibrium and estimating first-stage match probabilities. Further, using only a subset of the possible inequalities preserves consistency. The estimator allows markets with one-to-one, one-to-many, many-to-many and coalition formation matching, as well as externalities from agents outside a given match. For games with multiple equilibrium sets of matches, there is no need to estimate an equilibrium selection rule or computationally itemize the equilibria.

*Thanks to helpful comments from many colleagues as well as workshop participants at Chicago, Minnesota, the New York Fed, North Carolina, Northwestern, Stanford and Virginia. Email: fox@uchicago.edu.

# 1 Introduction

Becker (1973) introduces the use of two-sided matching theory to analyze empirical evidence on marriages between men and women. He models marriage as a competitive market with endogenous prices, or transfers between spouses. Other markets can be modeled as two-sided matching games with finite numbers of heterogeneous agents. Examples include the matching of workers to firms and upstream to downstream firms. Simpler matching games where one side of the market may care only about money include families to houses, members to exogenously specified clubs and bidders to multiple objects for sale in an auction. A key feature differentiating two-sided matching games from normal models of supply and demand is that, in matching, at least some agents on both sides of the market can make a limited number of trades. In marriage, each woman and have only one husband, so there is rivalry between men to marry the most attractive women.

Another matching framework is coalition formation, or one-sided matching. A group of agents divide themselves into mutually exclusive clubs, and the number and titles of clubs are not specified exogenously. Coalition formation can be used to study neighborhoods, political parties and industry alliances with horizontally differentiated firms. Theoretical work is also ongoing on models of many-sided matching, which can be used to study multi-tiered supply chains, for example. In the above games, this paper allows for externalities on payoffs arising from matches of agents outside of a given relationship.

Matching games are attractive frameworks for empirical work, as the models apply to a finite number of agents with flexible specifications for the productions functions generating match output. A typical data set for a matching market lists a series of observed matches and some characteristics about the parties in each match. Economists want to estimate the production function generating match output for observed and counterfactual matches. The research goal is primarily positive: to understand the relative importance of various observed agent characteristics in the equilibrium sorting of agents that we see in the data. Estimating match production functions can also produce an ordinal ranking of the efficiency of different match assignments for the same set of agents.

Matching games with endogenous prices have been used as a theoretical inspiration for univariate descriptive empirical work by Becker (1973) and others. However, despite their attractive theoretical properties, with one exception (which does not try to explain micro data on matches, see below) matching games with endogenous prices have not been used for formal structural estimation. A major impediment has been that standard maximum likelihood and method of moments estimators require a nested computation of an equilibrium for every realization of errors terms in order to evaluate the objective function for given parameter values. These complex equilibrium computations are nested within an integral over the unobserved error terms in the market, which should be of dimension equal to the number of potential matches in the market. For example, a simple marriage market with 100 men and 100 women nests the equilibrium computation inside a $100^2 = 10,000$ dimensional numerical integral, and this integral must be repeatedly evaluated at different trial parameter values in an outer optimization routine.

To address these practical concerns, this paper provides a consistent, nonparametric estimator for match production functions that is much easier to implement than the nested equilibria computation approach. The new estimator does not suffer from a curse of dimensionality in the number of agents in the market, and programming the objective function involves only evaluating the unknown production functions and checking inequalities. More abstractly, this paper presents conditions under which the match-specific production functions in a matching game with endogenous prices and transferable utilities are nonparametrically identified.

Some members of the class of matching games with transferable utilities are known as assignment games. The assignment problem was introduced by Koopmans and Beckmann (1957) and later examined by Shapley and Shubik (1972) and Becker (1973) for one-to-one matching, Kelso and Crawford (1982) for one-to-many matching, Leonard (1983) and Demange, Gale and Sotomayor (1986) for multiple-unit auctions, Sotomayor (1992) for many-to-many matching, Kovalenkov and Wooders (2003) for the case of arbitrary (no sides) coalition formation, and Ostrovsky (2004) for supply chain matching, among others. These models are applications of general equilibrium theory to games with finite numbers of agents. This paper allows for externalities, meaning the production form a match is a function both of the agents in the match and the matches formed by other agents. Overall, this paper uses the term "matching game" to encompass a broad class of models, including some games where the original theoretical analyses used different names.

One part of the outcome of a matching game is an assignment of matches for each agent, such as the identities of the suppliers for each retailer in a vertical market. This paper considers using data on only the assignment of matches, even though the matching game also has endogenous prices and production levels. A benefit to using assignments is that data on prices and production levels might not be available in many matching markets where economists believe prices are used. Becker (1973) studies marriage, a matching market where husbands and wives might exchange transfers, but those transfers are not recorded in typical data sets. Inter-firm contracting, such as vertical relationships between suppliers and retailers, also has the flavor that the contracting firms often transfer money, but the transfers are often confidential contractual details and not released to researchers.

An assignment is a qualitative outcome, so there are natural links between estimating matching games and single-agent discrete choice models, such as the well-known logit and probit discrete choice estimators. As in single-agent discrete choice, an agent picks the partner or partners that maximize the agent's payoffs from the partners that would agree to the match, all at the given prices. However, estimating a matching game presents additional complications because the actions of agents to match may preclude the possibility that other agents can match with the same parties. There are physical constraints (quotas) about how many matches each party can make. More simply, agents on the same side of the market are rivals, and the choice set of any agent is endogenously determined to be those agents on the other side willing to match with it.

Applying single-agent methods to matching games may give inconsistent parameter estimates. For example, a mine extracting a scarce mineral faces a capacity constraint and cannot sell to all customers. Consider a downstream manufacturer not being supplied with the mine's natural resource. It would be incorrect to apply a single-agent discrete choice model to the decision of a downstream manufacturer to source materials from the mine and infer that the manufacturer does not want the resource, as the true interpretation may be that other manufacturers have more valuable uses for the scarce resource. Even if the customer-specific price for every alternative to the mine is observed, in an econometric sense the prices are likely to be endogenous, as prices are correlated with unobserved components of the relationship specific costs and benefits from each customer-resource pair. The endogenous prices would make a single-agent discrete choice estimator with those prices included as a choice-specific regressor inconsistent.

The definition of transferable utility is that payoffs to an agent making a match are additively separable (quasi-linear) in the transfer paid to that agent by its partner. I also assume that all agents make a single-agent best response: each agent's profit from its current matches must exceed the profit if the agent were to match with a counterfactual partner, and pay that partner a transfer that makes the partner indifferent between the new match and the partner's old match. Transferable utility plus the assumption of single-agent best responses implies a condition I call local production maximization: two pairs of matches must together have a greater sum of production than if the four agents in question exchanged partners. This inequality forms the basis for identification and estimation.

The local production maximization inequality may be a necessary and not a sufficient condition for certain equilibrium concepts, so identification and estimation will not require computing equilibria. Further, I will show the useful result that estimation does not require the itemization of the inequalities for all pairs of observed matches, so estimation does not suffer from a computational curse of dimensionality in the number of agents in a matching market. Evaluating the inequality also does not require any first stage, nonparametric estimates of matching probabilities, so the estimator does not suffer from a data curse of dimensionality in the number of agents in the market.

The inequality involves only exchanging partners, so does not require the economist to consider alternative assignments with different numbers of matches for each agent. Therefore, the economist does not have to specify the maximum number of matches (the quota) each agent can make, which is an advantage as quotas are not found in many data sets.

Motivating the inequalities using single-agent best response actions, given price-taking agents, requires relatively weak assumptions. The inequalities can be written conditioning on the matches of agents not involved in a relationship, so the estimator can allow for arbitrary externalities. In a vertical market, the actions of two firms to match may create a cheaper product in the downstream market that competes with other firms and lowers the other firms' profits. Alternatively, two firms matching might represent a merger that lowers competition and raises profits for all firms. In a more traditional externality example, two firms matching may cause pollution that reduces the profits of other firms. As the externalities drop into the estimator's inequalities, adding externalities does not change the computational burden of the estimator.

For the dependent variable, the estimator in this paper uses assignment and not price data. There may be multiple assignments supporting equilibria, but as the estimator uses necessary conditions only, consistency is preserved under multiple equilibria. Unlike some recent approaches to estimating noncooperative, static Nash games, the estimator does not require estimation of an equilibrium selection rule and does not require computing all equilibrium assignments for a given parameter value and realization of error terms (Ciliberto and Tamer, 2003; Bajari, Hong and Ryan, 2004). No variables entering the equilibrium selection rule but excluded from the profits of agents in the market are required. Like the approach to estimating static Nash games of Pakes, Porter, Ho and Ishii (2005), the matching games estimator uses moment inequalities. However, this paper works with a maximum score estimator, which maintains consistency in the presence of unobservables that affect choices, an issue that troubles the estimator of Pakes et al. in many standard applications such as single-agent discrete choice.

A major contribution of the paper is to introduce a tractable maximum score estimator for matching games. Evaluating the objective function requires only computing match production levels and checking inequalities. The estimator is nonparametric because it does not require the assumption of parametric distributions for the stochastic portions of marketwide errors, and does not require a known parametric structure for the deterministic production functions. In practice, researchers will likely use a semiparametric version of the estimator: using economic theory to specify a parametric deterministic production function, but leaving the distribution of the error terms unspecified.

The matching maximum score estimator has already been used in an application. Bajari and Fox (2005) estimate the payoffs of bidders for items for sale in a multiple-unit auction. The application is to US auctions for distinct geographic markets for providing mobile phone service. A multiple-unit auction is a special case of a two-sided matching market where one side of the market, the items for sale, care only about the endogenous prices they sell for. A key issue in spectrum auctions is complementarities across items: is the valuation of a package of geographically near mobile phone markets greater than the sum of the valuations of the markets if won separately? Spectrum auctions highlight the maximum score estimator's ability to estimate nonlinearities in the payoffs across multiple geographic markets won by the same bidder.

Recently, several papers have proposed somewhat parametric approaches to estimating matching games, both with and without endogenous prices.[1] In contrast with the maximum score estimator in this paper, these methods all require parametric functional form assumptions on the distribution of the error terms, and most suffer from curses of dimensionality in the number of agents in the market. Many also make unattractive assumptions to resolve multiple equilibria issues.

Choo and Siow (2003) estimate a matching model with endogenous prices in a marriage application. In their model, the prices and error terms vary only over discrete classes of agents and are used to clear aggregate market clearing conditions, not to explain micro data on matches. For a related estimation approach in housing, see Bayer, McMillan and Reuben (2004). If it were to be applied outside of their application, Choo and Siow's estimator suffers from two curses of dimensionality: a data curse of dimensionality reflecting the need for first-stage nonparametric estimates of match probabilities, which are of high dimension if data across markets are used, and a second-stage need to solve a logit-derived system of nonlinear equations in the number of discrete types of agents in the market.

Whether a researcher should estimate a game with or without endogenous transfers depends on the researcher's understanding of the market in question. Boyd, Lankford, Loeb and Wyckoff (2003), Sørensen (2004) and Hitsch, Hortaçsu and Ariely (2005) estimate Gale and Shapley (1962) matching games where agents cannot transfer endogenous prices. All these approaches to estimating games without prices require additional assumptions to resolve a multiple equilibria problem in the set of matches. Boyd et al. and Sørensen, to differing degrees, require the nested solution of a matching mechanism; for every realization of the error terms for all matches in a market, the economist must compute an equilibrium. Hitsch et al. use data on both accepted and rejected matches from a website to estimate a dynamic programming search model, where some of the curse of dimensionality is alleviated because agents in the market share the econometrician's uncertainty about the unknown error terms. The current paper primarily consider perfect information matching games, although I discuss whether search is an candidate for the model's error terms in Section 5.7.

The rest of the text switches between less and more general models. Because the more general notation can obscure the basic ideas, Section 2 introduces the estimator for the simplest example of two-sided matching:

[1] In addition, estimating matching games is related to many other empirical literatures: auctions, supply and demand of homogeneous goods, non-cooperative static Nash games, models of sorting and public good provision, equilibrium models with differentiated products, hedonic models of equilibrium product characteristic choice, and models of price setting under search.

marriage. The main portions of the paper work with a more general class of matching games that incorporates important features of games, such as externalities, used in industrial organization, labor and public finance. Section 3 discusses identification, and Section 4 discusses the maximum score estimator. Work on the marriage, labor and financial investment markets has often focused on games without externalities, so Section 5 drops externalities to discuss additional results about games with a restrictive property: the decentralized equilibrium is in the core of the game.

## 2 The Simple Example of Marriage

Becker (1973) introduces a market model of marriage that suggests men with more education marry women with more education when the schooling of men and women are complements in production. However, having a spouse of the same religion may also be a factor in production, and religion may be correlated with schooling, so a researcher wants to estimate the complementarities between male and female schooling while controlling for the possibility that more surplus is generated in a marriage when spouses have the same religion. The presence of two inputs per agent requires a multivariate analysis, and hence formal estimation becomes attractive.

Consider a marriage market where a man $m$ marries a woman $w$ and they produce output according to the production function

$$f(m,w \mid \beta) = \beta_1 \times \text{school}_m \times \text{school}_w + \beta_2 \times 1\left[\text{religion}_m = \text{religion}_w\right],$$

where $\text{school}_m$ is the years of schooling of a man, $\text{religion}_m$ is the religion of a man, and the variables for women are similarly named. The indicator variable $1\left[\text{religion}_m = \text{religion}_w\right]$ is 1 when a hypothetical married couple share a religion, and 0 when they have different religions. A researcher wants to estimate the production function parameters such as $\beta_1$, which, if positive, means that male and female schooling are complements. Qualitative data on who matches with whom can only identify production functions up to scale normalizations, so $\beta_1 = \pm 1$. The parameter $\beta_2$ shows the benefit of having the same religion in schooling production units.[2]

Say marriages only happen within towns. Within each town, there will be a potential computational curse of dimensionality in the number of agents in a town. To understand the logic behind the combinatorics, let there be 3 men and 3 women in a town, none of whom can be single (for simplicity only). Let the notation 12 refer to a hypothetical marriage between man 1 and woman 2. It turns out that there are $3^2 = 9$ possible marriages that can happen, which are

$$11, 12, 13, 21, 22, 23, 31, 32, 33.$$

However, in an assignment of men to women for the entire market, each individual can join only one marriage. There are $3! = 6$ possible assignments for the entire market, which can be itemized as

$$\{11,22,33\}, \{11,23,32\}, \{12,21,33\}, \{12,23,31\}, \{13,21,32\}, \{13,22,31\}.$$

[2] The full production function might also have non-interacted schooling terms, such as $\beta_3 \times \text{school}_m$ and $\beta_4 \times \text{school}_w$. As I will explain later, these terms do not affect the matches that will form in equilibrium, and cannot be identified from data on realized matches. Their non-identification does not prevent the identification of the parameters multiplying the interaction of male and female characteristics.

To see how the combinatorics explode, now let there be 100 men and 100 women in a town. There are now $100^2 = 10{,}000$ matches and $100! = 9.33 \times 10^{157}$ marketwide assignments. By contrast, the number of atoms in the universe is much lower, at around $10^{79}$, than the number of assignments. Forming the probability that the observed assignment represents the market's equilibrium assignment will not be possible in a perfect information setup where a matching mechanism must be solved for every realization of the error terms.[3] Therefore, a standard likelihood estimator is not practical.

However, it is possible to derive inequalities that are necessary conditions for an equilibrium, and that are tractable to work with in estimation. Let the utility of man $a$ from matching with his observed wife $w_a$ be be $v^m(a, w_a) - t_{aw_a}$, where $v^a(a, w_a)$ is man $a$'s experience utility, and $t_{aw_a}$ is a (possibly negative) transfer paid by his wife to him. Let the payoffs of man $b$ from his wife $w_b$ have a similar functional form, $v^m(b, w_b) - t_{bw_b}$. The payoff from woman $w_a$ of matching with her husband $a$ is $v^w(a, w_a) + t_{aw_a}$, where $v^w(a, w_a)$ is woman $w_a$'s experience utility and $t_{aw_a}$ is the transfer she pays her husband.

Transfers are not observed in the data, so the goal will be to derive an inequality restriction involving total match production functions of the form $f(a, w_a) \equiv v^m(a, w_a) + v^w(a, w_a)$. It will be a result that assignment data alone can only identify match production functions, not the utilities of men and women separately.

Single-agent best response indicates that the total utility of man $a$ from marrying his observed wife $w_a$ exceeds his utility from instead marrying woman $w_b$ at a transfer level equal to the level that would make $w_b$ switch from her observed husband $b$,

$$v^m(a, w_a) - t_{aw_a} \geq v^m(a, w_b) - \tilde{t}_{aw_b}, \tag{1}$$

where $\tilde{t}_{aw_b}$ is the price that makes $w_b$ indifferent between $a$ and $b$:

$$v^w(a, w_b) + \tilde{t}_{aw_b} = v^w(b, w_b) + t_{bw_b}.$$

Substituting in the definition of $\tilde{t}_{aw_b}$ into (1) gives

$$v^m(a, w_a) - t_{aw_a} \geq v^m(a, w_b) - \left(v^w(b, w_b) + t_{bw_b} - v^w(a, w_b)\right). \tag{2}$$

Repeating the above algebra for the decision of $b$ to marry $w_b$ instead of $w_a$ gives

$$v^m(b, w_b) - t_{bw_b} \geq v^m(b, w_a) - \left(v^w(a, w_a) + t_{aw_a} - v^w(b, w_a)\right). \tag{3}$$

Adding the inequalities in (2) and (3) leaves, as the transfers $t_{aw_a}$ and $t_{bw_b}$ cancel,

$$v^m(a, w_a) + v^m(b, w_b) \geq v^m(a, w_b) - \left(v^w(b, w_b) - v^w(a, w_b)\right) + v^m(b, w_a) - \left(v^w(a, w_a) - v^w(b, w_a)\right).$$

Rearranging two of the $v$'s and substituting in the definition of a production function, $f(a, w_a) \equiv v^m(a, w_a) + v^w(a, w_a)$, leaves

$$f(a, w_a) + f(b, w_b) \geq f(a, w_b) + f(b, w_a).$$

This condition says that if the marriages $aw_a$ and $bw_b$ are observed, then single-agent best responses under price

[3] In an assignment game where an equilibrium assignment is in the core of the game, a matching mechanism can be implemented as a linear programming problem.

taking behavior imply that the sum of the match production levels from $aw_a$ and $bw_b$ must exceed the production levels from the exchange of spouses: $aw_b$ and $bw_a$. I call this inequality *local production maximization*, as the price-taking best responses of $a$, $b$, $w_a$ and $w_b$ ensure that production is maximized within the two marriages.

Some numbers may help crystallize why local production maximization can identify match production functions. Say there are two men and two women, and that $\beta_2 = 0$, so that religion does not enter production. Let one man and one woman each have a schooling level of 10, and let the other man and other woman have a schooling level of 1. The data are that agents assortatively match: the two high schooling people marry each other, and similarly the two low schooling agents marry. Then the local production maximization inequality is

$$\beta_1 \times 10 \times 10 + \beta_1 \times 1 \times 1 \geq \beta_1 \times 10 \times 1 + \beta_1 \times 1 \times 10,$$

or, simplifying, $\beta_1 101 > \beta_1 20$, which implies that $\beta_1 > 0$, or $\beta_1 = 1$ given the scale normalization that $\beta_1 = \pm 1$. Note that while the inequalities in estimation involve only observable characteristics, I will discuss how the matching maximum score estimator is consistent when matches are probabilistic to the econometrician because of unobservables.

Readers familiar with Becker's original model will recognize that local production maximization is consistent with Becker's theorem that assortative matching happens when production functions are complementary (supermodular) in a matched pair of inputs, one per agent. A supermodular, or complementary, production function has a positive cross-derivative in the inputs of men and women, or $\frac{\partial f(x_m, x_w)}{\partial x_m \partial x_w} > 0$, so that incremental marriage production is especially high when a man with high education marries a woman with high education. However, the local production maximization inequality underlying my formal estimation procedure generalizes to more than Becker's case of one continuously varying input per agent in a one-to-one two-sided matching game. I will show that local production maximization allows many inputs per agent (in this section, schooling and religion), unordered inputs (religion), production functions that are not globally super or submodular in pairs of inputs, interactions between different characteristics of the same agent, many-to-many two-sided matching with nonlinearities across payoffs of agents on the same side of the market, one-sided matching, multiple equilibrium assignments, and games with externalities. Local production maximization will apply to many games where the decentralized equilibrium does not maximize total marketwide (global) production, which is the property Becker uses to prove his theorem.

A researcher has data from 10 towns, indexed by $h$. The observed wife of man $a$ in town $h$ is $w_a^h$. The researcher knows the years of schooling and religion of all adults. For this data, the maximum score objective function introduced in this paper is

$$Q(\beta) = \sum_{h=1}^{10} \sum_{a=1}^{100} \sum_{b=a+1}^{100} 1 \left[ f\left(a, w_b^h \mid \beta\right) + f\left(b, w_b^h \mid \beta\right) > f\left(a, w_b^h \mid \beta\right) + f\left(b, w_a^h \mid \beta\right) \right].$$

The maximum score estimator imposes the best responses assumption and finds the production function parameters most consistent with local production maximization. For a market, the estimator itemizes over all pairs of men $a$ and $b$ and their observed wives $w_a^h$ and $w_b^h$. For a given vector of production function parameters, the estimator asks whether the sum of the productions of two marriages exceeds the sum of the productions when the two couples exchange partners. If the deterministic production from the observed marriages is larger, the score of correct predictions according to local production maximization, and hence the maximum score objective function, increases by 1.

A researcher uses a global optimization routine to numerically maximize $Q(\beta)$ to find the vector of production function parameters that make the observed marriages have the greatest score of correct predictions according to Becker's model. Evaluating the objective function requires only computing production for given marriages and checking inequalities. No complex matching algorithms must be solved and the estimator avoids any first stage estimation of matching probabilities. The estimator is semiparametric because it does not require the researcher to assume a particular functional form for the distribution of the error terms. The maximum score estimator in this paper is nonparametric in its most general form, as the estimator will also not require a functional form assumption for the production function.

# 3 Identification in Matching Games

This section proves that production functions in matching games can be nonparametrically identified. Identification results, in contrast to estimation, assume that a researcher has an infinite amount of data, and computational concerns are not relevant. Estimation will be discussed in Section 4.

This paper considers matching games with transferable utility and endogenous prices. Matching games have been fairly well studied in the theoretical literature. For expositional purposes, I divide matching games into two subcategories: multi-sided matching and coalition formation games. In a coalition formation game (or one-sided matching), any arbitrary subset of agents in the economy may form a match. Examples include a group of people choosing roommates, local residents forming clubs, and so on. In this framework, the number of clubs arises endogenously.

By contrast, in multi-sided matching agents in exclusive "sides" of the market match at least with some agents from other other sides. The most common example of multi-sided matching is two-sided matching, which I use here for simplicity. As its name implies, two-sided matching relies on dividing agents into two exclusive groups, such as upstream and downstream firms. Agents on one side of the market, such as upstream firms, can only match with the agents on the other side, in this case downstream firms.

Two-sided matching is itself divided into one-to-one, many-to-one and many-to-many matching. A one-to-one matching market is like the marriage market in Western society: each man can marry only one woman. A many-to-one matching market is like a stylized version of a multiple-unit auction or labor market: each bidder can win multiple items but each item can be one only once, and each worker can have only one job, while each employer can hire multiple workers, perhaps up to some employer-specific limit, or quota. Finally, in many-to-many matching, both sides of the market can make multiple matches up to some agent-specific number of matches, or quota. An example is the matching of upstream firms (suppliers) and downstream firms (retailers). Each retailer may stock items from multiple suppliers, and each supplier is likely to sell to multiple retail outlets.

Throughout the discussion of two-sided matching, I use notation based upon the example of upstream and downstream firms, as many-to-many matching is more general than one-to-one and many-to-one matching, and important for applications to supply chains.

## 3.1 Agents and Match Production

Let there be two sides to a market, upstream firms and downstream firms. There are $U$ upstream firms, indexed by $a = 1,\ldots,U$. The other side of the market is the $D$ downstream firms, indexed by $i = 1,\ldots,D$. $U$ and $D$ refer to both the sets and numbers of firms.

Each firm has a quota, the number of matches that it can physically make. The quota of upstream firm $a$ is $q_a^u$, and likewise the quota of downstream firm $i$ is $q_i^d$. An agent can often make fewer matches than its quota, although that feature of the model is not essential. A quota can be set to $+\infty$, so that a firm may be unconstrained in the number of matches it can make. $M_a$ is a set of downstream firms that may hypothetically match with upstream firm $a$. If $q_a^u = q_i^d = 1 \ \forall a \in U, d \in D$, the game is the familiar one-to-one or marriage model.

Each agent has a vector of characteristics in the data. For upstream firm $a$, $x_a^u$ is a vector of $r^u$ different characteristics. For example, we could observe characteristics such as $a$'s location and product quality. Likewise for downstream firm $i$, $x_i^d$ is a vector of $r^d$ observable characteristics.

The production function from a match is the key structural primitive that drives the pattern of matching and is the goal of estimation. A production function takes the observed characteristics of the parties in a match and creates some level of output. In one-to-one matching, if man $a$ marries woman $i$, total match production is $f\left(x_a^u, x_i^d\right)$. In many-to-one matching, if downstream firms $i$ and $j$ match with upstream firm $a$, total production is $f\left(x_a^u, x_i^d, x_j^d\right)$. If upstream firm $a$ has a quota of $q_a = 3$, then the firm can supply up to three downstream firms, so a more general notation for output when firm $a$ matches $i$ and $j$ is $f\left(x_a^u, x_i^d, x_j^d, \emptyset\right)$, where the empty set $\emptyset$ stands in for the idea that slot 3 is not filled. As above, $M_a = \{i, j\}$ is the set of downstream firms supplied by firm $a$. Another way of writing production functions is then $f\left(x_a^u, \{x_k^d\}_{k \in M_a}\right)$, where the empty set for the vacant slot is suppressed for convenience, and the set notation for the covariates for the downstream firms is meant to be expanded to be equal to $f\left(x_a^u, x_i^d, x_j^d\right)$.

This notation quickly becomes cumbersome. In many instances, I use the shorthand notation $f(a,i,j)$ or $f(a, M_a)$ to stand in for $f\left(x_a^u, x_i^d, x_j^d\right)$. Remember that production functions are always functions of the observable characteristics of agents, even if the characteristics themselves are suppressed in the notation. Sometimes, $a$ matching with $i$ is written as $ai$.

As with the previous marriage example, we can microfound match production functions with the non-transfer revenue functions over various matching agents, as in

$$f\left(x_a^u, x_i^d, x_j^d\right) \equiv v^u\left(x_a^u, x_i^d, x_j^d\right) + v^d\left(x_a^u, x_i^d, x_j^d\right) + v^d\left(x_a^u, x_j^d, x_i^d\right),$$

where $v^u$ is the pre-transfer revenue function for an upstream firm, and $v^d$ is the revenue function for a downstream firm.

It is usual in matching games to normalize the production from remaining single to be zero, or $f\left(x_a^u\right) = f\left(x_i^d\right) = 0$ $\forall i \in D, a \in U$. This normalization is not necessary, but it does make theoretical results in matching easier to derive. In some cases, a researcher might want to include match-specific covariates $\left(x_{ai}^{ud}\right)$, such as the distance between two firms. Estimation can proceed as long as the data contain match-specific covariates for both observed and counterfactual matches. Match-specific covariates make the identification problem easier by having the production of individual matches shift around in an flexible way. I will not consider match-specific covariates further in this section to prove that they are not required for identification.

Writing production functions in many-to-many matches is slightly more complicated. Consider a market where upstream firm $a$ sells to downstream firms $i$ and $j$ and $i$ also receives product from upstream firm $b$. If we allow for arbitrary nonlinearities across these relationships, the production function should be written as $f(a,b,i,j)$. We need a mechanism to distinguish the fact that $a$ does not supply $j$. Typically in an application, a researcher will choose a parametric functional form for $f$ so that nonlinearities in $i$'s profits across its supplier's characteristics are distinguished from $a$'s nonlinearities across the retailers it sells to. However, when being completely nonparametric, some additional assumption needs to be placed on $f$. For simplicity of exposition, in many-to-many two-sided matching I assume that the revenue function for the downstream firm is additively separable across suppliers, or

$$v^d\left(\{x_k^u\}_{k\in\{a,b\}}, x_i^d\right) = \sum_{k\in\{a,b\}} v^d\left(x_k^u, x_i^d\right) = \sum_{k\in\{a,b\}} v^d(k,i),$$

where in this case downstream firm $i$'s payoffs are evaluated at its matches with suppliers $a$ and $b$.[4]

I should emphasize that this assumption is made for simplicity and is not related to any deep limitation from matching theory. Again, in practice a researcher will put a parametric structure on $f$ that handles the nonlinearities across multiple partners separately for each agent.

For coalition formation, agents are not exogenously divided into sides of the market. Therefore, there is no need to distinguish between downstream and upstream firms. The notation is the same as many-to-one (if each agent can join only one coalition) or many-to-many matching (if agents can join multiple coalitions), without $u$ and $d$ superscripts.[5] The production if agents $a$, $b$ and $c$ all match together is $f(a,b,c)$.

The discussion of the payoff framework of a matching game can be formalized into an assumption.

**Assumption 1.**

1. *Agents care only about payments / profits, or alternatively transfers enter profits quasi-linearly (transferable utility).*

2. *If $f$ is nonparametric, for many-to-many matching, one side of the market has revenues that are additively separable across multiple matches. Without loss of generality, label the side of the market with revenues that are additive across multiple matches the "downstream firms".*

An outcome of a matching game is a set of physical matches for all firms and a set of monetary transfers between matched firms. The main dependent variable in my analysis is the set of physical matches for an entire matching market, which I label an assignment.

**Definition 1.** *Let an* ***assignment*** $\{M_a\}_{a\in U}$ *be a physically possible set of physical matches, where $M_a$ is the set of downstream firms matching with upstream firm $a$. Physically possible means quotas are satisfied: $|M_a| \le q_a^u \, \forall a \in U$ and $\left|M_i^d\right| \le q_i^d \, \forall i \in D$ and $M_i^d$ is the implied (by $\{M_a\}_{a\in U}$) set of upstream firms matching with downstream firm $i$.*

[4] Sotomayor (1992) and Sotomayor (1999) study many-to-many matching under the stronger assumption that payoffs for both upstream and downstream firms are additively separable across multiple matches.

[5] I restrict attention to markets where agents can join only a single coalition at a time, although the the equivalence between a social planning problem and the decentralized equilibrium should extend to the more general case, given appropriate notation.

In many games, it will be important to allow for externalities. For example, if firms are competing in a product market, then the decision of some firms to match (cooperate, merge) may raise (if the merger reduces competition) or lower (if cooperation produces a lower cost competitor) the profits of other firms. To allow for externalities, I write that a match's production is a function of its own characteristics, and the characteristics and matches of all the firms in an assignment. Let $E$ be an assignment. If $a$ matches with the downstream firms and externalities are important, $a$'s match production can be written as $f(a, M_a \mid E)$, where the conditioning notation implies that the matches of other firms may exert an externality on the firms matching with $a$.

## 3.2 Single-Agent Best Responses and Local Production Maximization

This paper identifies the production function in a matching game using a system of inequalities derived from revealed preference arguments. The condition I use is called local production maximization. It is a consequence of single-agent best responses under price taking behavior, as I have already shown for the marriage example. If both matches of upstream firm $a$ with downstream firm $i$ and upstream firm $b$ with downstream firm $j$ are observed, then a local implication of production maximization is that the total production of the two matches exceeds the total production from the exchange of partners, $aj$ and $bi$. Otherwise, matches $aj$ and $bi$ could form without disturbing any other matches and without changing the total number of matches of any agent. The formal definition of local production maximization is as follows.

**Definition 2.** 

- *In a two-sided matching game with assignment $E$, consider two upstream firms, $a$ and $b$, two groups of downstream firms $M_a$ and $M_b$, and two downstream firms, $i \in M_a$ and $j \in M_b$, all in a matching market $h$. Further let $i \notin M_b$ and $j \notin M_a$, and let $\tilde{E}$ be the assignment $E$ except that $aj$ and $bi$ match and $ai$ and $bj$ do not match. The matches $ai$ and $bj$ satisfy* ***local production maximization*** *when*

$$f(a, M_a \mid E) + f(b, M_b \mid E) \geq f\left(a, (M_a \setminus \{i\}) \cup \{j\} \mid \tilde{E}\right) + f\left(b, (M_b \setminus \{j\}) \cup \{i\} \mid \tilde{E}\right). \tag{4}$$

- *In a coalition formation game with assignment $E$, consider two coalitions $M_a$ and $M_b$, and one agent from each coalition: $M_a \ni i \notin M_b$ and $M_b \ni j \notin M_a$. Let $\tilde{E}$ be the overall assignment when $M_b \ni i \notin M_a$ and $M_a \ni j \notin M_b$. The observed matches $ai$ and $bj$ satisfy* ***local production maximization*** *when*

$$f(M_a \mid E) + f(M_b \mid E) \geq f\left((M_a \setminus \{i\}) \cup \{j\} \mid \tilde{E}\right) + f\left((M_b \setminus \{j\}) \cup \{i\} \mid \tilde{E}\right).$$

In the rest of the paper, I will focus on two-sided many-to-many matching for conciseness. Coalition formation just drops the distinction between upstream and downstream firms, and Definition 2 sufficiently documents how the inequality in the estimator changes for the case of coalition formation. Once one understands the derivation of the inequalities, it is easy to extend the analysis to other matching games with endogenous transfers, such as the many-sided chain matching studied by Ostrovsky (2004).

I will derive the local production maximization inequalities from an assumption of single-agent best responses for the case of upstream and downstream firms. Let $\bar{M}_a = M_a \setminus \{i\}$ be the downstream firms other than $i$ matching with $a$, and similarly let $\bar{M}_b = M_b \setminus \{j\}$. The profit of upstream firm $a$ from matching with the downstream firms $M_a$ is $v^u(a, i, \bar{M}_a \mid E) - t_{ai} - \sum_{k \in \bar{M}_a} t_{ak}$, where $v^u(a, i, \bar{M}_a \mid E)$ are the revenues from the market assignment, and

$t_{ai}$ is a payment from upstream firm $a$ to downstream firm $i$. The profits and revenues of $b$ are similar. Let the profit of downstream firm $i$ from matching with upstream firm $a$ and the other downstream firms in $M_a$ be $v^d(a,i,\bar{M}_a \mid E)+t_{ai}$, where $v^d(a,i,\bar{M}_a \mid E)$ is $i$'s revenues from its business at the market assignment, and $t_{ai}$ is the transfer payment from $a$. The goal will be to derive a local production maximization condition involving production functions of the form $f(a,i,\bar{M}_a \mid E) \equiv v^u(a,i,\bar{M}_a \mid E)+v^d(a,i,\bar{M}_a \mid E)$.

Single-agent best response indicates that the profit of $a$ from matching with $i$ exceeds the profit from matching with $j$ in $i$'s place at the transfer $j$ would require to switch from $b$

$$v^u(a,i,\bar{M}_a \mid E)-t_{ai}-\sum_{k\in\bar{M}_a} t_{ak} \geq v^u\left(a,j,\bar{M}_a \mid \tilde{E}\right)-\tilde{t}_{aj}-\sum_{k\in\bar{M}_a} t_{ak}, \tag{5}$$

where $\tilde{t}_{aj}$ is the price that makes $j$ indifferent between $a$ and $j$'s observed partner $b$:

$$v^d\left(a,j,\bar{M}_a \mid \tilde{E}\right)+\tilde{t}_{aj}=v^d(b,j,\bar{M}_b \mid E)+t_{bj}.$$

Note that I do not consider $a$ adding $j$ while $a$ remains matched to $i$. Such a move is possible if $a$ has unused quota, but I do not consider that deviation in the derivation of the local production maximization condition because data on quotas are often not available to a researcher. Substituting in the definition of $\tilde{t}_{aj}$ into (5) and cancelling the duplicate transfers $\sum_{k\in\bar{M}_a} t_{ak}$ gives

$$v^u(a,i,\bar{M}_a \mid E)-t_{ai} \geq v^u\left(a,j,\bar{M}_a \mid \tilde{E}\right)-\left(v^d(b,j,\bar{M}_b \mid E)+t_{bj}-v^d\left(a,j,\bar{M}_a \mid \tilde{E}\right)\right). \tag{6}$$

Repeating the above algebra for decision of $b$ to match with $j$ instead of $i$ gives

$$v^u(b,j,\bar{M}_b \mid E)-t_{bj} \geq v^u\left(b,i,\bar{M}_b \mid \tilde{E}\right)-\left(v^d(a,i,\bar{M}_a \mid E)+t_{ai}-v^d\left(b,i,\bar{M}_b \mid \tilde{E}\right)\right). \tag{7}$$

Adding the inequalities in (6) and (7) leaves, as the transfers $t_{bj}$ and $t_{ai}$ cancel out,

$$\begin{aligned} v^u(a,i,\bar{M}_a \mid E)+v^u(b,j,\bar{M}_b \mid E) \geq\ & v^u\left(a,j,\bar{M}_a \mid \tilde{E}\right)-\left(v^d(b,j,\bar{M}_b \mid E)-v^d\left(a,j,\bar{M}_a \mid \tilde{E}\right)\right)+ \\ & v^u\left(b,i,\bar{M}_b \mid \tilde{E}\right)-\left(v^d(a,i,\bar{M}_a \mid E)+t_{ai}-v^d\left(b,i,\bar{M}_b \mid \tilde{E}\right)\right). \end{aligned}$$

Rearranging two of the $v$'s and substituting in the definition of a production function, $f(a,i,\bar{M}_a \mid E) \equiv v^u(a,i,\bar{M}_a \mid E)+v^d(a,i,\bar{M}_a \mid E)$, leaves

$$f(a,i,\bar{M}_a \mid E)+f(b,j,\bar{M}_b \mid E) \geq f\left(a,j,\bar{M}_a \mid \tilde{E}\right)+f\left(b,i,\bar{M}_b \mid \tilde{E}\right),$$

which is just Definition 2, local production maximization.

Notice how local production maximization depends on only the production functions $f(a,i,\bar{M}_a \mid E)$. Using data on assignments alone, I cannot hope to identify the revenue functions for upstream and downstream firms separately. This is not necessarily a weakness, as $f(a,i,\bar{M}_a \mid E)$ is more general than $v^u(a,i,\bar{M}_a \mid E)$ and $v^d(a,i,\bar{M}_a \mid E)$ as separate functions, as $f(a,i,\bar{M}_a \mid E)$ does not require that the production from a match be additively separable across agents. Also note that transfers do not enter the local production maximization inequality. This too may be an advantage if the researcher does not have data on transfers between agents. In many inter-firm contracting applications, firms exchange transfers but do not release the details of the transfers to researchers.

Local production maximization was not derived from a complete equilibrium concept, but from only the definitions of single-agent best response under price taking behavior and transferable utility. As $E$ includes the actions of other agents, the framework can handle quite general forms of externalities. For example, two firms matching may create a competitive pressure for the other firms in the downstream market if the matching firms can offer a cheaper product.[6]

An equilibrium that satisfies local production maximization (or single-agent best responses) for all pairs may not satisfy global production maximization: the sum of marketwide match production may not be maximized. For example, $ai$ and $bj$ may find it in their private interests to match, but they may impose an externality on other firms, such as the match $ck$. If the externality on $ck$ is large enough, it could be that this equilibrium produces a lower sum of marketwide production than if all six firms had formed a grand coalition and chosen an alternative matching arrangement. Local production maximization is not derived from a restriction involving group decision making.

### 3.3 Firm Beliefs About Counterfactual Externalities

Definition 2, local production maximization, allows a researcher to include externalities in the payoff of agents. The inequalities as written assume that upstream firm $a$ believes that when $a$ replaces its partner $i$ with a new downstream firm $j$, $i$ will choose to match with $j$'s old partner $b$. The counterfactual assignment with matches $aj$ and $bi$ is labeled $\tilde{E}$. The inclusion of $\tilde{E}$ in Definition 2 can be motivated by any of three assumptions: 1) Upstream $a$ is naive and believes $i$ will match with $b$ when $a$ drops $i$; 2) There is not time for $i$ to find any other partner than the now available $b$ after $a$ drops $i$, so $i$ must match with $b$; and 3) All firms are small in the calculation of externalities, and the assignment $\tilde{E}$ is $a$'s approximation as to what will happen. Any of the three arguments lead to Definition 2 as written.

In some games it may be possible that $i$ will choose to match with some third firm $c$ after being dropped by $a$. To facilitate the match with $i$, $c$ might drop its partner $k$, producing a chain of disruptions. As this new chain was precipitated by an out-of-equilibrium deviation by $a$, the chain is unlikely to lead to a new equilibrium and hence end. It does not seem logical that $a$ could work out a reasonable counterfactual implication of the chain of disruptions it causes, and incorporate how those disruptions will affect the externalities imposed on $a$ when it drops $i$ for $j$.

Chains of disruptions do not arise when testing whether an set of strategies is a pure strategy Nash equilibrium to a static noncooperative game. In a normal form Nash game, each agent's strategy space is unrivaled: any strategy is physically possible. If the game is an entry game, each agent can choose to enter the market regardless of what other agents do. To check if a deviation is profitable, a researcher fixes the actions of all other players.[7] On the other hand, in a matching game, the actions of firms are constrained to by physically

[6] An externality caused by the internal operations of an agent on other aspects of the same agent's operations will not be observable, if the internal operations of an agent are not modeled in the matching game.

[7] The closest approximation to holding the actions of other parties fixed in a matching game is to consider counterfactuals evaluated at externalities $E$ instead of $\tilde{E}$. However, it is physically impossible for matches $aj$ and $bi$ to form as part of the original assignment $E$, because in $E$ (the data) matches $ai$ and $bj$ but not $aj$ and $bi$ form. The identification strategy in this paper works by comparing the probabilities of different assignments. A physically impossible assignment has a probability of zero.

possible: a firm can only make more matches if it has unused quota. Upstream firm $a$ must, if its quota is constrained, drop $i$ to match with $j$. Having an agent make a deviation while holding the actions of the other involved agents fixed does not make sense in a matching game with externalities.

The focus on the beliefs of firms making a deviation about counterfactual externalities only arise in matching games where production is a function of the matches of agents outside of the current match. Without externalities, Definition 2 is derivable from single-agent best responses without arguments about a firm's beliefs about counterfactual externalities. With externalities, the equilibrium concept for the game in question will determine whether Definition 2 is applicable. Again, any of the three explanations listed above provides a motivation for Definition 2.

## 3.4 Equilibrium Existence and Uniqueness

I have not introduced a definition of an equilibrium, and I certainly have not imposed sufficient conditions to ensure the existence of an equilibrium. The concept that I informally label "single-agent best responses" is closely related to the notion of a "stable allocation" in Hatfield and Milgrom (2005). In many-to-one two-sided matching with complementarities across matches on the same side of the market, Hatfield and Milgrom (2005) present a constructive theorem that demonstrates that preference profiles can be found for where there is no stable allocation. The result of Hatfield and Milgrom does not mean that an equilibrium does not exist for any empirical application, just that very general existence theorems cannot be proved.

Another solution concept is known as the core. Lucas (1995) discusses some core non-emptiness results for assignment games when there are more than two sides to a match. Ostrovsky (2004) shows how a special chain structure for $n$-way matching guarantees the existence of a stable chain under endogenous prices and restrictions on payoffs. 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. Kovalenkov and Wooders (2003) use a weaker notion of the core, the $\varepsilon$-core, and derive conditions on parameterized families of games for the non-emptiness of the $\varepsilon$-core for coalition formation games.

Many interesting empirical applications require investigating possibilities outside of the scope of current existence theorems. I make the assumption that the data on the assignment are generated by a matching game with endogenous prices, and represent an equilibrium for the game. At this equilibrium, the local production maximization condition holds for all subsets of firms satisfying the conditions of Definition 2. Mathematically, this existence assumption is nested into the forthcoming Assumption 3.

Many games with where an equilibrium is defined to be robust to pairwise deviations will have multiple equilibria, including multiple equilibrium assignments, in addition to multiple vectors of transfers between matched partners in the assignment. As estimation will rely on the local production maximization inequalities, rather than computing an equilibrium, the multiple equilibrium assignments property of many matching games will not pose a problem for estimation.[8]

[8] A common solution concept in match games without externalities is the core. A core outcome is robust to deviations by coalitions of agents. As the core solution concept is used in important classes of matching games, I discuss the core more formally in Section 5.

### 3.5 A Class of Production Functions for Identification

In a single-agent discrete choice model, the discrete choice observed in data is a qualitative outcome. The preference ordering generating the discrete choice can be represented on the real line using a utility function, with preferred choices offering higher utility values. It is a commonly known result in choice theory that any positive monotonic transformation of an agent's utility function will produce the same preference ordering, and thus discrete choice data can only identify utility functions up to positive monotonic transformations. In single-agent, semiparametric discrete choice estimation, where deterministic payoffs are assumed to be of the form $x'\beta$, location and scale normalizations must be imposed on $\beta$. Common scale normalizations impose that $\beta'\beta = 1$ and, alternatively, normalize the value of one element of the vector $\beta$ to be $\pm 1$. In a generalization, Matzkin (1993) considers nonparametric identification of the payoff function, and must restrict attention to a class of utility functions where no function is a positive monotonic transformation of another function in the class.

In the matching estimators introduced in this paper, the only outcome data being used are the assignments. An assignment is a qualitative outcome, and so production functions are only identifiable in classes that give unique predictions about local production maximization.

**Assumption 2.** *Let $f \in \Theta$, where $\Theta$ is a set of match production functions satisfying the following properties.*

1. *For each $f \in \Theta$, there is no $\tilde{f} \in \Theta$ such that for all two vectors of characteristics for upstream firms $x_a^u$ and $x_b^u$,*

$$f\left(x_a^u, \vec{x}_1 \mid E\right) + f\left(x_b^u, \vec{x}_2 \mid E\right) \geq f\left(x_a^u, \vec{x}_3 \mid E\right) + f\left(x_b^u, \vec{x}_4 \mid E\right) \Longleftrightarrow \tilde{f}\left(x_a^u, \vec{x}_1 \mid \tilde{E}\right) + \tilde{f}\left(x_b^u, \vec{x}_2 \mid \tilde{E}\right) \geq \tilde{f}\left(x_a^u, \vec{x}_3 \mid \tilde{E}\right) + \tilde{f}\left(x_b^u, \vec{x}_4 \mid \tilde{E}\right),$$

   *where for feasible groups of downstream firms $\vec{x}_1$, $\vec{x}_2$, $\vec{x}_3$, and $\vec{x}_4$, $\vec{x}_3$ is formed from $\vec{x}_1$ by exchanging one partner from $\vec{x}_2$, and $\vec{x}_4$ is formed from $\vec{x}_2$ by exchanging one partner with $\vec{x}_1$. For games with externalities, $E$ is an assignment where matches $ai$ and $bj$ form, and $\tilde{E}$ is the same assignment except that matches $aj$ and $bi$ form instead of $ai$ and $bj$.*

2. *For each $f \in \Theta$, $f$ is continuous in all of its arguments.*

3. *$\Theta$ is compact.*

Data on assignments identify tradeoffs of inputs in determining the rank ordering of production from rearrangements of matches in a matching game, not the cardinality of production.

The interaction of characteristics of upstream and downstream firms drives the patterns of matches that are the endogenous variables used in identification. Definition 2 makes this clearer. As the same set of firms appear on either side of the local production maximization inequality, terms that do not involve interactions between the characteristics of firms difference out. For example, in a one-to-one marriage market without externalities, if $f\left(x_a^u, x_i^d\right) = \beta_u x_a^u + \beta_d x_i^d$, then the local production maximization inequality in Definition 2 reduces to

$$\beta_u x_a^u + \beta_d x_i^d + \beta_u x_b^u + \beta_d x_j^d \geq \beta_u x_b^u + \beta_d x_i^d + \beta_u x_a^u + \beta_d x_j^d,$$

or $0 \geq 0$, so the definition has no empirical content. Theoretically, the uninteracted characteristics are valued equally by all potential partner firms, and are priced out in equilibrium. Only the differential production levels of various matches affect assignments.

If there is a observed characteristic such as $x_a^u$ that is not interacted with the characteristics of downstream firms, then the parameter $\beta_u$ multiplying $x_a^u$ cannot be identified from the qualitative assignment data. However, as $\beta_u x_a^u$ cancels out from the local production maximization condition, other parameters can be identified. Thus production functions are identified only within a class $\Theta$ satisfying Assumption 2.[9]

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 match production of all matches equally, the characteristics difference out, and do not affect the assignment of upstream to downstream firms. Therefore, if the policy questions of interest to the investigator are not functions of these unobserved characteristics, then differencing them out leads to empirical robustness to missing data problems.[10]

For researchers interested in nonparametric identification, further examples can clarify what properties of production can be identified in a class $\Theta$. Consider the marriage model in Becker (1973) with one characteristic for each spouse. Focus on the two seemingly very different production functions $f(a,i) = 2x_a^u x_i^d$ and $\tilde{f}(a,i) = -\left(x_a^u - x_i^d\right)^2$. These production functions have the same cross-derivative, $\frac{\partial f(a,i)}{\partial x_a^u \partial x_i^d} = 2$, so the production functions are both globally supermodular. Both production functions imply that, when agent characteristics are schooling levels, highly educated men should marry highly educated women. The reasoning is different, as in $f$ matching two agents with high schooling levels creates a large amount of production, and in $\tilde{f}$ not matching two agents with the same schooling creates a loss. It is not possible to use data on assignments to distinguish between $f$ and $\tilde{f}$, which are both supermodular over their entire support and have the same predictions for observed assignments.

However, even for the case of only one characteristic per agent, formal estimation can identify more than Becker (1973). Again, consider the case of one-to-one two-sided matching (marriage). Then, in Part 1 of Assumption 2, $\vec{x}_1 = \vec{x}_4$ and $\vec{x}_2 = \vec{x}_3$. The condition in Part 1 can be rewritten as

$$f(x_a^u, \vec{x}_1) - f(x_a^u, \vec{x}_2) \geq f(x_b^u, \vec{x}_1) - f(x_b^u, \vec{x}_2) \Longleftrightarrow \tilde{f}(x_a^u, \vec{x}_1) - \tilde{f}(x_b^u, \vec{x}_2) \geq \tilde{f}(x_a^u, \vec{x}_1) - \tilde{f}(x_b^u, \vec{x}_2),$$

which, if $x_a^u$ is a scalar, is a non-derivative based definition of supermodularity (increasing differences). The condition states that there must be at least one point with different implications for supermodularity. Assumption 2 requires only one point of disagreement between $f$ and $\tilde{f}$, not the supermodularity of $f$ over its entire support, as in Becker (1973). So, nonparametrically, it is possible to identify the sign of the cross-derivative

[9] One way of identifying the coefficient $\beta_u$ on $x_a^u$ in payoffs is if unmatched upstream firms are observed, and unmatched firms do not value their own characteristic $x_a^u$. Then $x_a^u$ 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, $x_a^u$ is truly not a characteristic valued equally by all downstream firms, as being unmatched is treated as a type of firm.

[10] In demand estimation methods such as Berry, Levinsohn and Pakes (1995), investigators are often concerned that the endogenous prices are correlated with unobserved product characteristics. As instruments are hard to find, typically researchers assume that the observed product characteristics are independent of the unobserved characteristics. By contrast, prices do not appear in the definition of local production maximization, so the advantage of differencing out unobserved product characteristics involves a concern that the unobserved characteristics are correlated with the observed characteristics.

of $f$ (for continuous characteristics) over its entire support, rather than assuming $f$ is either globally sub or supermodular, as in production functions such as $f(a,i) = 2x_a^u x_i^d$.

This paper considers many other classes of games than just marriage. For these games, generalizations of concepts such as local supermodularity still are some of the properties of production functions that arise from single-agent best responses and therefore drive equilibrium assignments. Part 1 of Assumption 2 uses the local production maximization condition to precisely state what can be identified using data on assignments alone.

One limitation of Assumption 2 is that two class members $f$ and $\tilde{f}$ must be distinguishable by exchanges of only downstream firm per upstream firm. Single-agent best responses imply the local production maximization inequality, which involves exchanges of one downstream firm per upstream firm. It is hypothetically possible that $f$ and $\tilde{f}$ could produce the same ranking for the relative sums of production of all exchanges of one downstream firm, but not agree on the ranking of the sums of production from exchanges of two downstream firms per upstream firm. For example, $f$ and $\tilde{f}$ might give different implications for the inequality (written for $f$)

$$f(a,i,j) + f(b,k,l) > f(a,k,l) + f(b,i,j),$$

as one cannot typically derive whether the inequality is $>$ or $\leq$ using a series of exchanges of one downstream firm per upstream firm $a$ or $b$. Under Assumption 2, the class $\Theta$ must rule out the possibility that $f$ and $\tilde{f}$ only disagree over exchanges of two downstream firms per upstream firm.

Most parametric forms chosen by researchers in applications will not involve such complex nonlinearities in production functions, and the local identification considered in Assumption 2 will not pose an empirical obstacle. For example, Bajari and Fox (2005) estimate complementarities across multiple geographic markets for sale in a government auction. Complementarities are proxied by a measure of the geographic closeness of a collection of geographic markets. Exchanging one geographic market per competing bidder provides local variation in the closeness measure for each bidder, and this local variation is enough for identification using the local production maximization inequalities in the parametric class of production functions considered.

Note that the limitation to local identification in Assumption 2 is a consequence of not making strong assumptions about equilibria. Section 5 shows that a game where the equilibrium is in the core allows global identification, as Assumption 2 can be extended to allow exchanges of more than one downstream firm per upstream firm.[11]

Assumption 2 also states that the class of production functions for identification is compact. Compactness is required for the consistency of many extremum estimators, but otherwise is not deeply related to identification.

## 3.6 The Definition of a Market

The use of asymptotic theory to prove identification of match production functions 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

[11] Alternatively, one could specify a larger class $\Theta^+$ and then argue the local production maximization inequalities identify a subset of $\Theta^0 \subset \Theta^+$ with similar implications for local production maximization as the otherwise point identified $f^0 \in \Theta$, where $\Theta$ satisfies Assumption 2. This paper does not pursue set identification because this type of non-identification seems far removed from any empirical example, and in any case working with games in the core allows for global identification.

of firms changes the character of the matches that will be observed; it is much simpler to consider a limiting population with an infinite number of markets.

Note that while asymptotics in the number of agents in a market has a lot of practical appeal to researchers with data on only one market, it does not have a tight link to the concept of a law of large numbers in statistics. Laws of large numbers are used to prove consistency. Usually laws of large numbers consider adding more data, holding the previous data constant. However, in a matching market, adding another agent to the market alters the matches of the existing agents. So not only is more data added, but the data the researcher already has is altered. Nevertheless, Section 5.6 presents a short discussion relating increasing the number of agents in a market to the empirical industrial organization demand estimation literature on increasing the number of products in a differentiated products market. Also, I later present a Monte Carlo study that includes an examination of the properties of the estimator when using data from only one market.

So this paper primarily considers asymptotics in the number of markets. Each market $h$ is distinguished by its observed characteristics, its observed set of matches and the potentially unobserved endogenous prices, and the yet-to-be-introduced unobserved stochastic error terms generating the observed matches. The collection $X_h$ is an 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 3.** *The collection 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 $h$, or the total number of agents in a coalition formation game.*
- *For each upstream firm $a$ , $X_h$ contains the (potentially) observable vector of $r_u$ characteristics $x_a^u$ entering match production functions. $X_h$ also contains the quota $q_a^u$, the number of physical matches $a$ can make.*
- *For each downstream firm $i$, $X_h$ contains the (potentially) observable vector of $r_d$ characteristics $x_i^d$ entering match production functions. $X_h$ also contains the quota $q_i^d$, the number of physical matches $i$ can make.*
- *Any characteristics entering the value of remaining unmatched also enter into $X_h$.*
- *In models with multiple physical assignments than can support equilibria, the equilibrium selection rule enters $X_h$.*
- *The agent-specific nest fixed effects introduced in the next section are in $X_h$.*

If there are 1000 upstream firms and 1000 downstream firms, $X_h$ contains 2000 vectors of covariates as well as other data. The stochastic payoff terms are exogenous from a matching theory standpoint, but are specifically excluded from $X_h$.

In order to compute an equilibrium assignment for a given realization of all the error terms, every component of $X_h$ must be observable. Consider a marriage market, where Koopmans and Beckmann (1957) show that a linear program can be used to compute an equilibrium assignment. Consider a (computationally intractable) parametric maximum likelihood procedure that involves a nested solution to the linear programming problem for random combinations of error term values and trial guesses for the unknown production function $f$. Every component of $X_h$ would be needed to compute the equilibrium assignment and therefore for feasible estimation.

In many matching games, there can be multiple physical assignments that support equilibria. In this case, $X_h$ contains the exogenous process that selects an equilibrium physical assignment. Such a rule computes a solution to the market for a given realization of the error terms.

Identification in this paper relies on the necessary but not always sufficient local production maximization property of the observed equilibrium assignment. Section 4 shows that not every component of $X_h$ must be observable for identification and estimation. An important example is the quota of each agent is not needed for estimation. Another example is that the equilibrium selection is rule is not required.

In matching theory, 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. The definition of the relevant market is an important issue in most anti-trust litigation. The economic theory of matching is only developed for the case where a market is well defined. However, as Section 4 discusses, consistency will often be maintained if a researcher defines a market conservatively, and uses only a subset of the restrictions imposed by the theory.

### 3.7 Agent-Specific Nest Fixed Effects

The previous discussion of matching games has focused on deterministic models without error terms. However, such purely deterministic models will often not be flexible enough to perfectly fit the assignments from realistic data sets. Properly specified econometric models make the model consistent with arbitrary outcome data by adding error terms to the model. There are two basic approaches to adding error terms to discrete choice models. The parametric approach assumes a known functional form for the error terms. By contrast, the semiparametric and nonparametric approaches derive identification and consistent estimators that are valid for any error distribution satisfying broad properties. This paper uses the nonparametric approach, as functional forms for the error terms and the production functions are not assumed.

Typically, a researcher will want to allow agent and match-specific unobservable components of production to be correlated across similar match partners. A researcher can consistently estimate production functions while allowing for agent-specific fixed effects that are constant across nests specified by the researcher. The fixed effects represent unobservables in agent revenues. 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. Fixed effects for different nests can be correlated, and fixed effects for a given agent can be correlated with that agent's characteristics, as well as the characteristics and fixed effects of other agents.

For a matching market $h$, let there be a set of nests for upstream firms $\mathcal{N}_h^u$, and let the corresponding set of nests for downstream firms be $\mathcal{N}_h^d$. Let $n_h^u$ be an individual nest for upstream firms, and likewise let $n_h^d$ be a nest for downstream firms. In an extension of notation, let $n_h^d(i)$ be a function that gives the nest of downstream firm $i$. The production function for the match of upstream firm $a$ with set of downstream firms in $M_a$ at the market assignment $E$ is

$$f(a, M_a \mid E) + \sum_{k \in M_a} \left( \xi^u_{a n_h^d(k)} + \xi^d_{k n_h^u} \right)$$

where $a \in n_h^u$, $\xi^u_{an_h^d}$ is upstream firm $a$'s unobserved fixed effect for a downstream firm $k \in n_h^d(k)$, $\xi^d_{kn_h^u}$ is downstream firm $k$'s fixed effect for upstream firms such as $a$ in the nest $n_h^u$. The fixed effects enter additively separably into match production.[12]

Consider an inequality focusing on the upstream firms $a$ and $b$, the set of downstream firms matched with $a$ and $b$, $M_a$ and $M_b$, and downstream firms $M_a \ni i \notin M_b$ and $M_b \ni j \notin M_a$. Also assume that $a$ and $b$ are in the same nest $n_h^u$, and $i$ and $j$ are in the same nest $n_h^d$. The sum of the payoffs of the observed matches are

$$f(a, M_a \mid E) + \sum_{k \in M_a} \left( \xi^u_{an_h^d(k)} + \xi^d_{kn_h^u} \right) + f(b, M_b \mid E) + \sum_{k \in M_b} \left( \xi^u_{bn_h^d(k)} + \xi^d_{kn_h^u} \right),$$

while the sum of the payoffs when instead $a$ matches with $j$ and $b$ matches with $i$ is

$$f\left(a, \left(M_a^h \backslash \{i\}\right) \cup \{j\} \mid \tilde{E}\right) + \sum_{k \in \left(M_a^h \backslash \{i\}\right) \cup \{j\}} \left( \xi^u_{an_h^d(k)} + \xi^d_{kn_h^u} \right) + \\ f\left(b, \left(M_b^h \backslash \{j\}\right) \cup \{i\} \mid \tilde{E}\right) + \sum_{k \in \left(M_b^h \backslash \{j\}\right) \cup \{i\}} \left( \xi^u_{bn_h^d(k)} + \xi^d_{kn_h^u} \right).$$

By the assumption that $a$ and $b$ and also $i$ and $j$ are in the same nests, the sums of fixed effects of the form

$$\sum_{k \in M_a} \left( \xi^u_{an_h^d(k)} + \xi^d_{kn_h^u} \right) + \sum_{k \in M_b} \left( \xi^u_{bn_h^d(k)} + \xi^d_{kn_h^u} \right)$$

are identical under the observed matches and the exchange of upstream firms for $i$ and $j$. Thus, the fixed effects cancel out from Definition 2, local production maximization. The necessary condition for local production maximization depends only on $f$ when the two upstream firms are in the same nest the and two downstream firms are part of the same nest.

By looking within nests, a researcher can identify the unknown production function $f$ using within-nest variation in characteristics, while allowing the unobserved payoffs of firms to be correlated with covariates, and to be correlated across similar match partners. The fixed effects approach is powerful, but there are two downsides. First, the method is only consistent if the researcher does not define the nest too broadly. As with the definitions of markets that I discussed above, using too narrow nests preserves consistency. 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 variants on these two drawbacks also apply to the use of fixed effects in linear regression models, and are not unique to matching games.

### 3.8 Within Nests: The Rank Order Property for Local Production Maximization

The previous section introduces agent-specific fixed effects over nests of choices. Such fixed effects add error terms to the model and will often explain a good deal of residual variation in the data. Indeed, i.i.d. logit errors over broad types/nests of potential spouses are the only error terms in Choo and Siow (2003), meaning that the nest fixed effects already add a richer stochastic structure than some earlier work. If a researcher is happy with

[12]For single-agent discrete choice versions of the maximum score estimator, the payoffs only need to enter weakly separably (Fox, 2005). The additive separability here comes from the need to add the production from the sets of matches of two upstream firms.

agent-specific nest fixed effects, than nothing more needs to be done to ensure consistency. However, in any data set it is likely that suboptimal (in observable characteristics) matches will be seen within a nest, making the introduction of additional errors necessary to make the model able to explain the within-nest data.

An insight of Manski (1975) for single-agent discrete choice models is that under some conditions choice probabilities $P(i \mid X)$ are rank ordered by the deterministic part of utility $x_i'\beta$, so observed choices should, more often than not, have greater deterministic linear indices than unobserved choices. Consider an agent making a standard, single-agent, multinomial discrete choice from a set $J$ of choices. A choice $i$ gives payoff $x_i'\beta + \varepsilon_i$, where $\varepsilon_i$ is a choice-specific error term. Fix two choices, $i$ and $j$, from the set $J$ of all choices. If the joint density of the error terms for all choices is exchangeable, $x_i'\beta > x_j'\beta$ if and only $P(i \mid X) > P(j \mid X)$.[13]

A literal extension of the rank ordering of outcome probabilities does not hold in matching games, as a match $ai$ that gives a higher deterministic payoff $f(a,i)$ than another match $aj$ may not be observed with higher frequency if $i$ has attractive outside options. For nonparametric identification, I need to find a similar rank ordering property for matching games.

I extend Definition 2, local production maximization, to the case where the econometrician does not observe the stochastic error terms for partners with the same nests. The matching game will have local socially optimal matching in a probabilistic sense. Given two upstream firms within a nest and two downstream firms within a nest, it is more likely that the combination of two matches with the higher deterministic payoff will be observed than the alternative combination.

In many matching games, there can be multiple equilibrium assignments. Recall that the equilibrium assignment selection rule is in $X_h$. For any realization of the error terms, the equilibrium selection rule finds a physical set of matches. When the econometrician integrates out over the error terms for a given selection rule, the following property is assumed to hold. As externalities can enter payoffs, in the most general form the property refers to the probability of an entire assignment arising.

**Assumption 3.** *In a two-sided matching game with assignment $E$, consider two upstream firms, $a$ and $b$, two groups of downstream firms $M_a$ and $M_b$, and two downstream firms, $i \in M_a$ and $j \in M_b$, all in a matching market $h$. Further let $i \notin M_b$ and $j \notin M_a$, and let $\tilde{E}$ be the assignment $E$ except that $aj$ and $bi$ match and $ai$ and $bj$ do not match. Finally, $a$ and $b$ are in the same nest of upstream firms, and $i$ and $j$ are in the same nest of downstream firms. Assume*

$$f(a, M_a \mid E) + f(b, M_b \mid E) \geq f\left(a, (M_a \setminus \{i\}) \cup \{j\} \mid \tilde{E}\right) + f\left(b, (M_b \setminus \{j\}) \cup \{i\} \mid \tilde{E}\right)$$

*if and only if*

$$P(E \mid X_h, f) > P_{ab}\left(\tilde{E} \mid X_h, f\right),$$

*where $P(E \mid X_h, f)$ is the probability to the econometrician that the assignment $E$ happens, conditional on the potentially observable exogenous market characteristics in $X_h$ and the match production function $f$.*

The rank order assumption is key to identification. The assumption uses the notion of an assignment matching probability, or the probability that a set of matches involving all agents in a market happen at the same time.

[13] If the errors are independent and identically distributed across choices, 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. Fox (2005) discusses the extension to exchangeability.

$P(E \mid X_h, f)$ is the probability, from the econometrician's point of view, of the assignment $E$ arising. Calculating $P(E \mid X_h, f)$ involves integrating out the vector of all error terms over the region where the assignment $E$ is optimal. Proving that an extremum estimator is consistent requires showing that the probability limit 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 probabilities of the form $P(E \mid X_h, f)$.

The assignment matching probabilities are computed holding fixed the collection $X_h$ of all potentially observable exogenous market characteristics. As $X_h$ can have thousands of elements, an estimator that involves repeatedly solving for an equilibrium for realizations of the error terms to compute assignment match probabilities will not be tractable.

Again, the property in Assumption 3 holds $X_h$ fixed. $X_h$ includes the equilibrium selection rule as a function of the unobservables. Therefore, the equilibrium selection rule can be different across markets, as well as correlated with the other exogenous characteristics of the market in $X_h$, as well as the unknown and market-specific distribution function of the error terms.[14]

The possibility of multiple equilibrium assignments prevents any formal analysis of sufficient conditions for Assumption 3. Assumption 3 is more likely to hold if the equilibrium assignment selection rule selects "nearby" assignments when the realizations of the error terms are close. However, the concept of "nearby" equilibrium assignments is not formal, and so no formal analysis can be undertaken under multiple equilibria. Assumption 3 should be seen as a primitive assumption on both the equilibrium selection rule in $X_h$ and the distribution of unobservable error terms.

When the core is the solution concept, in many games there will be a unique equilibrium assignment with probability 1, and it is possible to discuss sufficient conditions for Assumption 3 to hold. Therefore, Section 5 discusses a set of sufficient conditions that generate Assumption 3.

## 3.9 Identification Through Covariate Variation

Point identification proves that there is only one production function $f^0$ in the class $\Theta$ 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 $P(E \mid X_h, f)$ are observable. Given the matching probabilities, Assumption 3 places restrictions on the set of production functions $f$ that are consistent with the data. Without additional assumptions than Assumption 3, the identified set of production functions comprises the production functions consistent with rank ordering.

**Definition 4.** *The identified set $\mathcal{F}$ of production functions comprises functions $f$ such that Assumption 3 holds for all possible markets $X$, pairs of upstream firms $a, b \in U_h$ that are in the same nest, feasible downstream matches for $a$ and $b$ $M_a \subseteq D_h$ and $M_b \subseteq D_h$, pairs of downstream firms $M_a \ni i \notin M_b$ and $M_b \ni j \notin M_a$ that are in the same nest, and assignments $E$.*

Without any restrictions on $X$, I can only prove that this set $\mathcal{F}$ exists, and that it is not the entire space $\Theta$ of theoretically possible parameters. In other words, $f$ is set-identified, and we can use the rank order property to identify bounds on $f^0$: the boundaries of $\mathcal{F}$.

[14]Notationally, the distribution of the error terms should be in $X_h$, but I want to emphasize that identification does not require specifying the distribution of the error terms.

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 is not specified. Matzkin (1993) extends the identification and estimation results to the nonparametric identification of the deterministic portion of utility as well. This section follows Matzkin by showing sufficient conditions on the variation in the data that allow point identification of the production functions.

**Definition 5.** *The match production function is point identified if there exists a set of markets* $h$ *with positive measure such that for any* $\tilde{f} \in \Theta$, $\tilde{f} \neq f^0$, *there exists assignments* $E$ *such that* $P\left(E \mid X_h, f^0\right) \neq P\left(E \mid X_h, \tilde{f}\right)$.

The mathematical argument for point identification focuses on varying the characteristics of two upstream firms in each market, as seen in the local production maximization inequality. To this end, make the following assumption about the identities of the relevant two upstream and two downstream firms and the corresponding variation in the observable data. I first need to split the vector of characteristics $x_a^u$ for upstream firms entering into production into $x_a^u = \left(x_{1,a}^u, x_{-1,a}^u\right)$, where $x_{1,a}^u$ is the first, scalar component of the vector and $x_{-1,a}^u$ is the vector of all other covariates.

**Assumption 4.** *For every market* $h$, *there are two particular upstream firms* $a$ *and* $b$, *which are always in the same nest. The joint distribution of the vectors of the characteristics of* $a$ *and* $b$ *entering into match production, conditional on other market characteristics* $X$, *is* $g\left(x_a^u, x_b^u \mid X \setminus \{x_a^u, x_b^u\}\right)$.

- *The joint density of the first elements of* $a$ *and* $b$*'s characteristics conditional on the other characteristics for* $a$ *and* $b$ *and all other market characteristics,* $g\left(x_{1,a}^u, x_{1,b}^u \mid X \setminus \left\{x_{1,a}^u, x_{1,b}^u\right\}\right)$, *has an everywhere positive density in* $\mathbb{R}^2$.
- *The data across markets are sampled statistically independently.*

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 abstract firm indices such as $a$, $b$, $i$ and $j$. The special random variable $x_{1,a}^u$ is assumed to be freely varying conditional on the other characteristics of the upstream and downstream firms. The existence of such a freely varying covariate is required for point identification of discrete choice models (Manski, 1988; Horowitz, 1998).

Intuitively, the support condition for $x_{1,a}^u$ and $x_{1,b}^u$ 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 production function $f$. In the case of 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 Assumption 3. On the other hand, adding additional observations with new continuous characteristics $x_{1,a}^u$ and $x_{1,b}^u$ from an infinite number of new markets (the exercise in identification) creates a continuum of restrictions from Assumption 3. Thus semiparametric point identification takes advantage of continuously varying covariates such as $x_{1,a}^u$ and $x_{1,b}^u$, and identification does not require examination of the entire set of possible matches.[15]

[15]The assumption that the support of $x_{1,a}^u$ and $x_{1,b}^u$ is $\mathbb{R}^2$, rather than some compact subset of $\mathbb{R}^2$, is made for convenience. Manski

The following assumption 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 matching probabilities. The consistency proof of the estimator shows that, in the case of the quota of matches each firm can make and the equilibrium selection rule, the observability of all elements of $X_h$ is a convenience, and not a necessity for identification.

**Assumption 5.** *For every matching market* $h$*, the econometrician observes the matrix of market characteristics* $X_h$.

Identification is stated in the following theorem, and the theorem is proved in Appendix A.1. The proof works by finding a point where the rank order property does not hold, and exploiting the continuous characteristic to show that a set around that point also does not satisfy the rank order property, so that the points not satisfying the rank order property have positive measure.[16]

**Theorem 1.** *Under Assumptions 1, 2, 3, 4 and 5, the true production function* $f^0$ *from the data generating process is point identified in the set* $\Theta$.

# 4 Estimation of Match Production Functions

The previous section shows that match production functions are nonparametrically identified within a class $\Theta$ under the assumption that match probabilities are ranked by their deterministic contributions to output maximization. Ignoring issues with multiple equilibria, by following the identification argument, one can construct a potentially consistent estimator of the production function, $f^0$. The researcher nonparametrically estimates assignment probabilities $P\left(E \mid X_h, f^0\right)$ by using data across markets. Then the researcher uses the estimates of simultaneous match probabilities to estimate $\mathcal{F}$, the identified set using the conditions of Assumption 3. As data on more markets appear, the estimate of $\mathcal{F}$ converges to the true production function $f^0$, if the conditions for identification are met, and under possible additional regularity conditions.

Given that $X_h$ may have thousands or even millions of elements, and the number of markets in the data may be low, the dimensionality of match probabilities implies that first-stage nonparametric estimation of match probabilities 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 production function $f$, and does not involve auxiliary nonparametric estimates of matching probabilities.

## 4.1 Choosing Inequalities

Before estimating the production function, the researcher must choose a set of inequalities to form the objective function. Let $B^{\text{sub}}(X_h)$ be a set-of-sets-valued function returning the matches to use in estimation for a market

(1988) and Horowitz (1998) show how to relax the full support assumption for the identification of single agent binary choice models. The identification arguments in this paper are not related to the identification at infinity arguments made in the literature on selection and the related work on the special regressor estimator of Lewbel (2000).

[16] Identification can proceed using other assumptions than the rank-order property, Assumption 3. For example, in the unlikely even that there are match-specific regressors ($x_{ai}$) with full support that are independent of the error terms, identification based on the "special regressor" arguments of Lewbel (2000) might be possible. The Lewbel single-agent, multinomial choice estimator requires multidimensional density estimation and therefore suffers from a curse of dimensionality in the number of choices. Likewise, any matching estimator based on the "special regressor" identification argument will not be tractable in markets of reasonable size.

with characteristics $X_h$. An element of $B^{\text{sub}}(X_h)$ comprises a quartet of two upstream firms from the same nest and two downstream firms from the same nest to focus on, and a set of other downstream firms potentially matched to the two upstream firms at the same time and an assignment. In notation, an element $z$ of the set $B^{\text{sub}}(X_h)$ is a set

$$z = \{a, b, i, j, M_a, M_b, E\} \text{ where } a \neq b, a, b \in U_h, a, b \in n_h^u, i, j \in n_h^d, M_a \subseteq D_h, M_b \subseteq D_h,$$
$$M_a \ni i \notin M_b, M_b \ni j \notin M_a, |M_a| \leq q_a^u, |M_b| \leq q_b^u, q_k^d \geq 2 \, \forall k \in M_b \cap M_a,$$

where $E$ is an assignment, $U_h$ is the set of upstream firms in market $h$, and likewise $D_h$ is the set of downstream firms in market $h$. Here $M_a$ is a hypothetical set of matches for upstream firm $a$, and $M_b$ is a set of downstream firms for upstream firm $b$. Only cases where $M_a$ has weakly fewer downstream firms than the quota of $a$ and $M_b$ has fewer downstream firms than the quota of $b$ are considered. Further, only sets $M_a$ and $M_b$ where all downstream firms in both sets have quotas of at least two are considered at the same time.

There are several things to keep in mind when choosing $B^{\text{sub}}(X_h)$. First, the researcher should include some inequalities involving exchanging agents with very different characteristics, in order to rule out production functions $f$ that are far from the truth. Second, the researcher should attempt to use a deterministic rule to select inequalities, so as to aid replication by other researchers. One typical deterministic rule includes all inequalities formed by exchanges of one downstream firm each between two upstream firms. Third, if the researcher is unsure whether an exchange of partners is physically possible, the researcher should consider not including the corresponding inequality. For example, if the researcher is unsure of the exact definition of the matching market, using a conservative definition of the market to form pairwise comparisons will preserve consistency if the formal conditions on $B^{\text{sub}}(X_h)$ listed below are satisfied. Including inequalities involving exchanges that are not physically possible will break consistency, as there is no information on the revealed preferences of agents in those inequalities.

In many applications the number of exchanges of one downstream firm each between two upstream firms will be so large that the evaluation of all the inequalities from the simplest deterministic rule will be too computationally expensive. If there are 100 upstream firms and 1000 downstream firms and each upstream firm is the only supplier for 10 downstream firms in the data, the total number of inequalities that "turn on" for a given dataset (see below) is $\sum_{a=1}^{100} \sum_{b=a+1}^{100} \sum_{i=1}^{10} \sum_{j=1}^{10} 1 = 495{,}000$. If there are too many more inequalities than 495,000, the deterministic rule of using exchanges of one downstream firm per upstream firm will be computationally infeasible given current computer technology.

This paper considers only inequalities where each of two upstream firms exchanges one downstream firm with the other supplier. In markets where the equilibrium concept involves only pairwise deviations, inequalities with exchanges of more than two firms are not theoretically derivable from the assumption of single-agent best responses.

To apply a lemma from asymptotic theory,[17] we need the following assumption about the data generating process.

**Assumption 6.** *The number of possible matching quartets included in the maximum score objective function,*

[17] Lemma 2.4 from Newey and McFadden (1994), which appears in the proof of Theorem 2.

$\left|B^{\text{sub}}(X_h)\right|$, *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.*

I make the following covariate assumption that extends the identification covariate conditions to all included inequalities.

**Assumption 7.** *The underlying data generating process and the choice of inequalities to include in $B^{\text{sub}}(X_h)$ make the conditions of Assumption 4 hold for every inequality in the maximum score objective function.*

In the usual case when not all theoretically possible inequalities are included in the objective function, it is important that, asymptotically, all configurations of covariates appear. Otherwise, the production function is not identifiable at those points.

**Assumption 8.** *The choice of inequalities in $B^{\text{sub}}(X_h)$ induces a distribution of included characteristics with support equal to the sampling distribution of the characteristics in the true data generating process.*

## 4.2 The Matching Maximum Score Estimator

Define the matching maximum score estimator to be any production function $f \in \Theta$ that maximizes the objective function

$$Q_H(f) = \frac{1}{H}\sum_{h\in H}\sum_{z\in B^{\text{sub}}(X_h)} 1[E \text{ in } h] \times 1\left[f(a,M_a \mid E) + f(b,M_b \mid E) > f\left(a,(M_a\backslash\{i\})\cup\{j\} \mid \tilde{E}\right) + f\left(b,(M_b\backslash\{j\})\cup\{i\} \mid \tilde{E}\right)\right]$$
$$+ 1\left[\tilde{E} \text{ in } h\right] \times 1\left[f(a,M_a \mid E) + f(b,M_b \mid E) < f\left(a,(M_a\backslash\{i\})\cup\{j\} \mid \tilde{E}\right) + f\left(b,(M_b\backslash\{j\})\cup\{i\} \mid \tilde{E}\right)\right], \quad (8)$$

where for all inequalities the assignment $\tilde{E}$ is formed from $E$ by replacing match $ai$ with $aj$ and match $bj$ with $bi$.

Here $H$ is the number of markets observed by the econometrician. The terms $1[\cdot]$ are indicator functions equal to 1 when the condition in brackets is true, and 0 otherwise. The main dependent variable of interest from the two-sided game is $1[E \text{ in } h]$, which is equal to 1 if the assignment $E$ occurs in the data for market $h$. As part of $E$, upstream firm $a$ matches with the set of downstream firms in $M_a$ and likewise supplier $b$ matches with the set of retailers in $M_b$. Note that if the game does not have externalities, the dependent variable can be rewritten as $1[M_a, M_b \text{ in } h]$, which is 1 if upstream firm $a$ matches with the set of downstream firms $M_a$ (and only those) and upstream firm $b$ matches with the set $M_b$ in market $h$, which could happen under many different assignments $E$.

The above primitive definition for $Q_H(f)$ is written in a way that makes it easy to compute the probability limit of the maximum score objective function in the proof of the estimator's consistency. The dependent variable data are not known when taking a probability limit. However, programming the objective function for a given dataset is much simpler, as the dependent variable data are known, and there is no need to itemize over terms that are known be to be zero for all trial production functions $f$. For a given data set, a researcher only programs the inequalities that actually "turn on" because the dependent variable data in the relevant indicator functions are true. For a given dataset with dependent variable data, let $A_h$ be the quartets that are relevant given the set of physical matches seen in the data. Further, let $M_a^h$ be the downstream firms supplied by upstream firm $a$ in the data. The maximum score objective function one programs is

$$Q_H(f) = \frac{1}{H} \sum_{h \in H} \sum_{\{a,b,i,j\} \in A_h} 1 \left[ f\left(a, M_a^h \mid E_h\right) + f\left(b, M_b^h \mid E_h\right) > f\left(a, \left(M_a^h \backslash \{i\}\right) \cup \{j\} \mid \tilde{E}_h\right) + f\left(b, \left(M_b^h \backslash \{j\}\right) \cup \{i\} \mid \tilde{E}_h\right) \right],$$

where $\tilde{E}_h$ is the counterfactual assignment that is equal to $E_h$ except that $a$ and $b$ exchange the partners $i$ and $j$. For a finite sample, the objective function considers quartets of two upstream firms and one downstream firm being supplied by each upstream firm, but not the other. A non-stochastic notion of local production maximization implies that if supplier $a$ matches with retailer $i$ but not $j$, and supplier $b$ matches with retailer $j$ but not $i$, then the sum the production for the observed matches must be greater than the production from the suppliers exchanging retailers. Assumption 3, the rank order property, extends local production maximization to the stochastic case where there are error terms unobserved to the econometrician. If the local production maximization condition is met for an observed pair of supplier relationships at some trial production function $f$, the score of correct predictions within the quartet increases by 1. The matching maximum score estimator is any production function in the class $\Theta$ that receives the highest score for not violating predictions of Assumption 3's version of local production maximization for observed match quartets.

Note that the quotas, the numbers of maximum physical matches employers can make, do not enter $Q_H(f)$ explicitly when it is programmed for a given data set. Any matching situation that violates the quota of any agent will not appear in the data, so we know that all inequalities in the objective function will correspond to matching situations that do not violate quota restrictions. Therefore, by focusing on the data at hand, the econometrician is guaranteed to not violate quotas. Also, the estimator only considers deviations in the inequalities where the number of matches for each firm are kept the same as is seen in the data. The estimator does not consider any deviations that might break the quota of an agent, so the estimator does not require or use data on quotas.[18]

For games with multiple equilibrium assignments, equilibrium selection rules do not enter the objective function. Unlike some procedures for dealing with multiple equilibria, there is no need to estimate the equilibrium selection rule in order to estimate the match production functions.

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 $Q_H(f)$ produces a consistent estimator of the true population parameter vector $f^0 \in \Theta$. Numerically maximizing an objective function over the space of an unknown function $f$ that must be in some class $\Theta$ satisfying Assumption 2 is nonstandard. Matzkin (1990) provides an operational procedure for the case where $\Theta$ is defined to be the class of least-concave functions. Matzkin (1991) and Matzkin (1992) also discuss estimation of discrete choice payoff functions under nonparametric shape restrictions.

[18] 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, Lankford, Loeb and Wyckoff (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.

Correlation between quotas and other observable exogenous variables is not a problem for the maximum score estimator. Neither this paper nor any other considers the case where the quotas might be endogenous: there are firm-specific unobservable terms in the production functions correlated with the quotas. For some applications, it is best to assume that quotas are not binding, and let the number of vacant match slots arise endogenously in equilibrium.

A more standard computer programming approach is to define $\Theta$ to be a class of production functions defined to be a parametric function known up to a finite-dimensional vector of unknown parameters, $\beta$. When one specifies a parametric functional form for the production function, the maximum score estimator is labeled semiparametric, rather than nonparametric. In practice, one uses a numerical optimization package to compute a maximum of the objective function. The objective function is not differentiable in $\beta$, so local numerical optimization methods cannot be used.[19] The estimates in the application of Bajari and Fox (2005) use a global optimization routine, specifically differential evolution (Storn and Price, 1997).

Researchers may in practice use a parametric linear-in-parameters specification for production functions. For identification, scale normalizations must be imposed on the linear-in-parameters production function. A convenient scale normalization to use during numerical optimization for an extremum estimator is $\beta_1 = \pm 1$, where $\beta_1$ is the coefficient on a continuous agent characteristic in production.[20]

### 4.3 Consistency

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

**Theorem 2.** *Under Assumptions 1, 2, 3, 6, 7, and 8, any production function $f_H \in \Theta$ that maximizes the matching maximum score objective function is a consistent estimator for $f^0 \in \Theta$, the true production function.*

The proof is in Appendix A.2. The most economically interesting part of the proof proves the true production function $f^0$ maximizes the probability limit of the objective function. The dependent variable data indicator functions of the form $\frac{1}{H}\sum_{h=1}^{H} 1[E \,\mathrm{in}\, h]$ converge to the expectation of matching probabilities $E_X\left\{P\left(E \mid X, f^0\right)\right\}$. Thus, asymptotically the estimator uses choice probabilities even though computationally estimation does not require a first-stage nonparametric estimation of $P\left(E \mid X, f^0\right)$ or the computation of $P(E \mid X, f)$ using a matching mechanism (linear program) for trial guesses of $f$.

Assumption 3, the rank order property, is used to show that the inequalities involving the production functions will multiply the higher of $P\left(E \mid X, f^0\right)$ and $P\left(\tilde{E} \mid X, f^0\right)$, where $\tilde{E}$ is $E$ with $aj$ replacing $ai$ and $bi$ replacing $bj$, when the trial production function is the truth, $f^0$.[21]

## 5 Games with the Core Solution Concept

The previous discussion emphasizes the weak conditions on the structure of a matching game (single-agent best responses under price taking behavior) required for consistency of the maximum score estimator. Using

[19] It is not clear that local optimization routines should be used for many smooth objective functions because even smooth objective functions may have many local optima.

[20] In estimation, if the sign of $\beta_1$ is not known from economic theory, it can be superconsistently estimated by estimating the model twice, once where $\beta_1$ is fixed at $-1$, and once where $\beta_1$ is fixed at 1. The final estimates for all parameters correspond to the sign of $\beta_1$ with the highest objective function value.

[21] Such an argument would not work if the objective function involved minimizing the number of incorrect predictions times a "penalty term" (other than the current 1s and 0s) reflecting the difference between the production levels of the matches in the data and some counterfactual matches, when evaluated at a hypothetical $f$.

weak conditions on equilibria is important for games with externalities and complementarities between the characteristics of agents on the same side of the market. However, some games, such as Becker's marriage model, produce equilibria that are in the core of the game. A core outcome is robust to deviations by coalitions of agents. As the entire market is one such coalition, an outcome in the core must maximize the sum of match productions for all matches in the assignment supporting the outcome.

The robustness of the core to deviation by coalitions of agents means that identification can include exchanges of two or more downstream firms per upstream firm. Compared to Assumption 2, it is now possible to consider global identification of production functions, meaning production functions in the identifiable set can now disagree about any arbitrary exchange of downstream firms. While in the discussion of Assumption 2 I emphasized that global nonparametric identification is not likely to be empirically relevant in most applications, it is of interest from the viewpoint of econometric theory.

When programming the estimator, a researcher can include exchanges of more than two downstream firms per upstream firm, as the core is robust to any deviation by a coalition of agents. In some instances, including more inequalities in estimation may increase the finite-sample precision of the estimator.

An assignment supporting the core is unique with probability 1 if the production levels have full support on the real line, as two arbitrary combinations of real numbers will sum to the same value with probability 0. This means that a matching probability is well-defined without the need to specify an unobserved equilibrium assignment selection rule. Therefore, this section uses the assumption that the observed assignment supports a core outcome to investigate sufficient conditions on within-nest error terms to satisfy Assumption 3, the rank order property. This section concludes with a Monte Carlo study that examines whether the estimator has a large finite-sample bias under model assumptions that are not enough for consistency. This discussion can be skipped by those readers willing to understand Assumption 3 as an intuitive assumption that can be motivated by several more primitive conditions, or an assumption that is of second-order importance in practice because most of the residual variation in the assignment data is captured with agent-specific nest fixed effects.

## 5.1 The Core

This section presents an alternative derivation of local production maximization using a cooperative game theory and general equilibrium solution concept: the core. The core can be used for games without externalities as well as for games with externalities that allow for side payments between unmatched partners. Most applications with externalities do not allow side payments, so for readability I drop externalities in this section.

Firms receive monetary payments that are their profits. If downstream firms $i$ and $j$ and upstream firm $a$ all match, $i$ receives profit $p_i^d$, $j$ receives profit $p_j^d$, and $a$ receives profit $p_a^u$. All firms prefer to receive higher profits. As firms only want to maximize profits, firms have transferable utility.

A matching game with endogenous prices produces as an outcome a set of physical pairings between firms in the market (an assignment) and a vector of profits, one for each firm. I will define the core of a matching game to be the set of profits that are both feasible and satisfy the property that no group of firms would prefer to deviate and match outside of the mechanism.

**Definition 6.** *1. Let an* ***outcome*** $\left\{ \{p_a^u\}_{a\in U}, \{p_i^d\}_{i\in D}, \{M_a\}_{a\in U} \right\}$ *be a vector of profits for all firms and an assignment.*

2. *Let a **feasible** outcome be an outcome that includes a vector of profits that is physically possible to produce given the assignment, or where*

$$\sum_{a\in U} p_a^u + \sum_{i\in D} p_i^d \leq \sum_{a\in U} f(a, M_a).$$

3. *Let a **contained coalition** be a set $\{C^u \subseteq U, C^d \subseteq D\}$ of upstream firms $C^u$ and downstream firms $C^d$, where all matches of upstream firms in $C^u$ are with downstream firms in $C^d$, and all matches of downstream firms in $C^d$ are with upstream firms in $C^u$.*

4. *Let an **a contained coalition with a feasible internal arrangement,** $C = \left\{C^u \subseteq U, C^d \subseteq D, \{M_a^C\}_{a\in C^u}\right\}$, be a contained coalition with a given set of matches of downstream to upstream firms for the firms in the coalition, where feasible means $|M_a^C| \leq q_a^u \forall a \in U$ and $\left|M_i^{C,d}\right| \leq q_i^d \forall i \in D$, where $M_i^{C,d}$ is the set of upstream firms matching with downstream firm $i$ in the contained coalition.*

5. *Let a feasible **core** outcome be a feasible outcome where each contained coalition with a feasible internal arrangement $C = \left\{C^u \subseteq U, C^d \subseteq D, \{M_a^C\}_{a\in C^u}\right\}$ receive greater profits than its production, or*

$$\sum_{a\in C^u} p_a^u + \sum_{i\in C^d} p_i^d \geq \sum_{a\in C^u} f\left(a, M_a^C\right),$$

*and that each firm receives nonnegative profits, or*

$$p_a^u \geq 0 \forall a \in U \text{ and } p_i^d \geq 0 \forall i \in D.$$

Agents in a match split the production from the match. In many-to-many matching, splitting output generalizes to the notion that any contained coalition splits the production from the set of matches in the coalition. Therefore, adding side payments to a game without externalities does not change the outcome.

**Lemma 1.** *For an outcomes $\left\{\{p_a^u\}_{a\in U}, \{p_i^d\}_{i\in D}, \{M_a\}_{a\in U}\right\}$ in the feasible core, the profits and the assignment generating production satisfy, for any contained coalition with a feasible internal arrangement $C$ that is part of the outcome,*

$$\sum_{a\in C^u} p_a^u + \sum_{i\in C^d} p_i^d = \sum_{a\in C^u} f\left(a, M_a^C\right).$$

*Proof.* Assume to the contrary for some contained coalition $C$ that is part of the core outcome. If $\sum_{a\in C^u} p_a^u + \sum_{i\in C^d} p_i^d < \sum_{a\in C^u} f\left(a, M_a^C\right)$, then the coalition $C$ would be better off by deviating from the core assignment, as all members could be paid more than their current profits. If $\sum_{a\in C^u} p_a^u + \sum_{i\in C^d} p_i^d > \sum_{a\in C^u} f\left(a, M_a^C\right)$, then feasibility means that the level of profits that the coalition $C$ earns is not all produced by the members $C$, and must come from some other contained coalition. Therefore, there exists at least one other contained coalition with a feasible internal arrangement $\tilde{C}$, $C \cap \tilde{C} = \emptyset$, such that $\sum_{a\in \tilde{C}^u} p_a^u + \sum_{i\in \tilde{C}^d} p_i^d < \sum_{a\in \tilde{C}^u} f\left(a, M_a^{\tilde{C}}\right)$, which means that the contained coalition $\tilde{C}$ would want to deviate from the core outcome. The deviation of $\tilde{C}$ violates the definition of the core, and is a contradiction.

If the contained coalition $C$ is the entire market, then $\sum_{a\in C^u} p_a^u + \sum_{i\in C^d} p_i^d > \sum_{a\in C^u} f\left(a, M_a^C\right)$ implies that the outcome is not feasible.

□

The key property for estimation of transferable utility two-sided matching games is that any outcome in the core must maximize the total marketwide output, $\sum_{a\in U} f(a, M_a)$.

**Lemma 2.** *Let the outcome* $\left\{\{p_a^u\}_{a\in U}, \{p_i^d\}_{i\in D}, \{M_a\}_{a\in U}\right\}$ *be in the feasible core. Then the assignment in the outcome maximizes total marketwide output* $\sum_{a\in U} f(a, M_a)$.

*Proof.* Output maximization follows from the definition of the core, as the entire market is a contained coalition with a feasible internal arrangement that maximizes total output. If output is not being maximized, the coalition of the entire market would deviate.

□

In a competitive economy with homogeneous goods for sale, the first welfare theorem states that any competitive equilibrium is Pareto optimal. With transferable utility, Pareto optimality is strengthened to production maximization in the sense that the decentralized equilibrium maximizes total marketwide production. Likewise, Lemma 2 states that any decentralized core outcome maximizes a social planning problem for the case of matching.[22]

## 5.2 Global Identification

The solution to the social planner's problem implies many restrictions for two upstream and two downstream firms at a time, in two-sided matching. Consider a hypothetical solution to the marketwide social planning problem with matches $ai$ and $bj$ but not $aj$ and $bi$. If the local production maximization inequality condition, Definition 2, is not satisfied, having matches $aj$ and $bi$ would improve total production 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 $j$) produces the definition of production maximization for the entire market, as long as remaining unmatched is considered a potential matching partner, where appropriate.

For many-to-one and many-to-many two-sided matching as well as coalition formation, the global production maximization property of the core allows estimation to include inequalities with exchanges of two or more downstream firms per upstream firm or coalition, for four or more firms being exchanged in total. Any deviation from global production maximization provides a valid set of inequalities unless the deviation violates quotas.

Including inequalities with exchanges of more than two downstream firms per upstream firm may not dramatically change the finite sample estimates, especially if a tight parametric specification for the class $\Theta$ of production functions is being used. In the multiple-unit auction application of Bajari and Fox (2005), we found that the magnitudes and signs of parameters were similar when we estimated using mainly exchanges of two items per bidder.

[22]The definition of the core involves group decision making. It is more traditional in two-sided matching theory (even without externalities) to define another solution concept. An example of a different solution concept is a stable match, which considers deviations by pairs of upstream and downstream firms. A theorist then proves that the other solution concept is equivalent to the core. Unfortunately, in many-to-one and many-to-many matching with general production functions that allow for complementarities across multiple downstream firms matching with the same upstream firm, the core is not equivalent to pairwise deviations.

Nevertheless, extending the discussion of local identification to global identification may be of some theoretical interest. In Section 3.5, Assumption 2 states that the local production maximization inequality allows the identification of production functions within a class $\Theta$ where local identification is sufficient for point identification. The single-agent best response motivation for local production maximization requires weak assumptions about the equilibrium of the matching game. By contrast, assuming the equilibrium is in the core is a strong assumption, but it will allow global identification.

Consider two production functions $f$ and $\tilde{f}$ in some class. It may be the case that $f$ and $\tilde{f}$ disagree only over exchanges of two downstream firms per upstream firm, as in

$$f(a,i,j)+f(b,k,l)>f(a,k,l)+f(b,i,j) \text{ and } \tilde{f}(a,i,j)+\tilde{f}(b,k,l)\leq\tilde{f}(a,k,l)+\tilde{f}(b,i,j).$$

If the equilibrium to a game is in the core, these inequalities can be used for identification and estimation. The version of the previous Assumption 2 for games where the equilibrium is in the core follows.

**Assumption 9.** *Let $f\in\Theta$, where $\Theta$ is a set of match production functions satisfying the following properties.*

1. *For each $f\in\Theta$, there is no $\tilde{f}\in\Theta$ such that for all two vectors of characteristics for upstream firms $x_a^u$ and $x_b^u$,*

   $$f(x_a^u,\vec{x}_1\mid E)+f(x_b^u,\vec{x}_2\mid E)\geq f\left(x_a^u,\vec{x}_3\mid\tilde{E}\right)+f\left(x_b^u,\vec{x}_4\mid\tilde{E}\right)\Longleftrightarrow\tilde{f}(x_a^u,\vec{x}_1\mid E)+\tilde{f}(x_b^u,\vec{x}_2\mid E)\geq\tilde{f}\left(x_a^u,\vec{x}_3\mid\tilde{E}\right)+\tilde{f}\left(x_b^u,\vec{x}_4\mid\tilde{E}\right),$$

   *where for feasible groups of downstream firms $\vec{x}_1$, $\vec{x}_2$, $\vec{x}_3$, and $\vec{x}_4$, $\vec{x}_3$ is formed from $\vec{x}_1$ by exchanging $N$ partners from $\vec{x}_2$, and $\vec{x}_4$ is formed from $\vec{x}_2$ by exchanging $N$ partners with $\vec{x}_1$. For games with externalities, $E$ is an assignment where matches of $a$ with the downstream firms with characteristics $\vec{x}_1$ and $b$ with the firms in $\vec{x}_2$ form, and $\tilde{E}$ is the same assignment except that matches of $a$ with the firms represented by $\vec{x}_3$ and $b$ with the firms represented by $\vec{x}_4$ form.*

2. *For each $f\in\Theta$, $f$ is continuous in all of its arguments.*

3. *$\Theta$ is compact.*

The difference from the case where equilibria are not in the core is that now exchanges of $N\geq 1$ downstream firms per upstream firm are allowed. Assumption 3, the within-nest rank order property, needs to be extended to allow for exchanges of $N\geq 1$ downstream firms per upstream firm as well. For conciseness, I do not repeat the assumption with this slight change in notation and wording.

With Assumption 9 and the above revision to Assumption 3, an analogous theorem to Theorem 1 can be easily be proved with a very similar argument. The maximum score estimator, equation (8), can be extended to allow for exchanges of $N$ or less downstream firms per upstream firm. Finally, consistency, Theorem 2, can be proved with almost the same argument.

## 5.3 Assortative Matching and the Supermodularity of Production

Lemma 2 shows that any core solution to a marriage market must maximize total marketwide output. Consider first the case where the production from a marriage is equal to the sum of the schooling of the man $a$, $x_a^u$, and the schooling of the woman $i$, $x_i^d$, with estimable weights $\beta_u$ and $\beta_d$: $f\left(x_a^u, x_i^d\right) = \beta_u x_a^u + \beta_d x_i^d$. Here, marriages with partners who have more schooling produce more output. However, in a market with at least one partner for every participant, it should be clear that any assignment of marriages where all agents marry produces the total marketwide output $\sum_{a \in U} \beta_u x_a^u + \sum_{d \in D} \beta_d x_i^d$. Because output is purely linear in characteristics, any exhaustive assignment will maximize total social surplus, and therefore any assignment can underly a core outcome.

The assignment, the set of physical matchings, only affects the total marketwide production if the characteristics of men and women interact in production. A condition that guarantees that the market will produce an assortative matching of highly-schooled men to women is supermodularity of production in schooling levels. Supermodularity means the cross-derivative of production with respect to male and female characteristics is positive, or $\frac{\partial f\left(x_a^u, x_i^d\right)}{\partial x_a^u \partial x_i^d} > 0$.

Becker (1973) uses these predictions about univariate production functions to explain stylized facts about the labor market. Becker's analysis is incomplete if $x_a^u$ and $x_i^d$ are vectors. For example, some of the inputs of men and women may be complements, and others may be substitutes, and these characteristics may be correlated in the cross section of men and the cross section of women. When the level of a dependent variable such as output is observed in data, controlling for multiple inputs, and particularly the correlation of inputs, lets multivariate ordinary least squares produce different coefficients than univariate least squares run for each input separately.[23]

When one extends the set of models considered to include one-to-many and many-to-many two-sided matching as well as coalition formation, researchers are not only interested in whether the inputs of agents on opposite sides of the market are complements or substitutes. With multiple partners, agents on the same side of the market may be complements or substitutes. For example, Bajari and Fox (2005) use the estimator in this paper on data from spectrum auctions to estimate whether geographically adjacent markets for mobile phone service are complements or substitutes.

To clarify, Becker's analysis and many of these follow-up points do not apply if the model's equilibrium is not in the core. Formal estimation is the only valid approach for analysis in a model with multiple equilibrium assignments and/or externalities.

### 5.4 Assignment Match Probabilities

All of the sufficient conditions for Assumption 3 rely on specifying some solution to the overall matching market as a function of the error terms, and then integrating out the errors $\psi_h$ to create a market assignment probability. Recall that an assignment is a set of physical pairings, but not the endogenous price vector. The following definition more formally defines the simultaneous match probabilities that appear in the statement of Assumption 3.

**Definition 7.** *In a matching market h with a matrix of characteristics $X_h$, consider the assignment $E$.*

$$P(E \mid X_h, f) = \text{Prob}_{\psi_h} 1\left[E \,\text{market}\,\text{assignment} \mid X, f\right], \tag{9}$$

[23] In linear regression, multivariate least squares applied produces the same slope coefficients as univariate least squares applied to each covariate separately when the covariance between the included characteristics is 0.

*where $\psi_h$ is a vector of error terms unobserved to the econometrician. This definition is well-defined if there is a unique assignment supporting the core with probability 1.*

Typically, there will be a unique assignment with probability 1 when the error terms and production functions take values on the real line. A core outcome must solve a social planning problem. The probability that any two combinations of an arbitrary collection of real numbers sum to the same value is 0.

### 5.5 The Social Planning Problem as a Single-Agent Discrete Choice Problem

In a matching game where Lemma 2 states that core outcomes are in part solutions to social planning problems, the "single agent" from the random utility model literature is the social planner. The social planner acts like a single agent making a discrete choice from the large but finite number of marketwide assignments. Let the massive set of all feasible assignments be $\mathcal{Z}_h$. The unique socially optimal assignment $E_h$ satisfies,

$$\sum_{a\in U} f\left(a, M_a^{E_h}\right) + \psi_{E_h} \geq \sum_{a\in U} f\left(a, M_a^{Z_h}\right) + \psi_{Z_h} \ \forall \text{ allocations } Z_h \neq E_h, Z_h \in \mathcal{Z}_h, \quad (10)$$

where $\psi_{Z_h}$ is a composite error term reflecting the sum of the unobserved portions of production in the assignment $Z_h$. The above single-agent decision rule states that the sum of the payoffs of a socially optimal assignment is greater than other feasible assignments, where feasible assignments enforce the quotas of agents. The inequality in equation (10) transforms the computation of the social optimum into a single-agent discrete choice problem with extra additive errors.

A sufficient condition for the rank order property involves placing the error terms at the marketwide assignment level. The final perceived marketwide payoff of assignment $E$ is $\sum_{a\in U} f\left(a, M_a^E\right) + \psi_E$, where $\psi_E$ is an error term corresponding to the marketwide assignment $E$. This transforms a complex 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, $\sum_{a\in U} f\left(a, M_a^E\right)$, and adds a random error term to the final payoff.

The interpretation is that the social planner tries to maximize total output, but is unable to do so because of random disturbances. Equivalently, these marketwide errors represent inefficiencies in finding a core outcome in the decentralized market. Marketwide errors are similar to the quantal response equilibrium concept in Goeree, Holt and Palfrey (2004). In a quantal response equilibrium, agents choose a best response subject to some noise. Here, the "agent" is the social planner, or the unmodelled decentralized process of finding a core outcome.

Manski (1975) proves that if a single agent decides between $J$ choices, with each choice $j$ giving payoff $u(x_j) + \varepsilon_j$, then the probability of picking choice $j$ exceeds the probability of picking an alternative $k$ if and only if $u(x_j) > u(x_k)$ when the error terms $\varepsilon_j$ are i.i.d or exchangeable. In words, the choice probabilities are rank-ordered by the deterministic payoffs $u(x_j)$. 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 joint density of the error terms is exchangeable, for a given agent.[24]

**Assumption 10.** *For all marketwide physical assignments $E$ of upstream to downstream firms in market $h$, and including the option of remaining unmatched where appropriate, let there be random variable $\psi_{Eh}$. Let the joint density of $\psi_{Eh}$ be $\omega(\psi_h \mid X_h)$, where $\psi_h$ is the vector of all $\psi_{Eh}$'s.*

1. *Let $\omega(\psi_h \mid X_h)$ be exchangeable across assignments $E$, and statistically independent across markets $h$, conditional on the matrix of potentially observable exogenous market characteristics $X_h$.*
2. *Let $\omega(\psi_h \mid X_h)$ have full support on $\mathbb{R}^{N_h}$, where $N_h = \dim \psi_h$.*

The density function of the error terms can vary across markets with potential market observables $X_h$. While exchangeable stochastic error terms is a restrictive assumption if the observable covariates have low explanatory power for predicting matches, Section 3.7 discusses how to relax Assumption 10 by allowing for firm-specific fixed effects over pre-specified nests of match partners.

The following lemma indicates that Assumption 10 is a sufficient for the rank-order property, Assumption 3.

**Lemma 3.** *Under Assumption 10, the rank order property, Assumption 3, holds.*

The proof of the result is in Appendix A.[25]

An exchangeable density is not the only condition under which the rank order property will hold. For quantal response equilibria and single-agent discrete models, Haile, Hortaçsu and Kosenok (2004) show that any set of choice probabilities can be generated by a member of the class of joint distributions for random variables that are independent but do not have identical marginal distributions, and, alternatively, a member of the class of joint distributions for random variables with identical marginal distributions but that are not independent across choices. As there are many realizations of matching probabilities consistent with the rank order property, there are many joint distribution for the social planner's errors that are consistent with the rank order property for a production function and a given realization of the potentially observable characteristics $X_h$ of a market.[26]

## 5.6 Match-Specific Errors

The most natural extension of the single-agent discrete choice random utility model formulation is to assume that the total production from upstream firm $a$ matching with the set of downstream firms $M_a$ is (ignoring fixed effects) $f\left(a, M_a^{E_h}\right) + \sum_{i \in M_a} \psi_{ai}$, where $\psi_{ai}$ reflects the idiosyncratic production of upstream firm $a$ matching with downstream firm $i$. The total marketwide production from an assignment $E_h$ of downstream firms to upstream firms is (again ignoring fixed effects)

$$\sum_{a \in U} f\left(a, M_a^{E_h}\right) + \psi_{E_h} = \sum_{a \in U} f\left(a, M_a^{E_h}\right) + \sum_{a \in U} \sum_{i \in M_a} \psi_{ai}.$$

[24]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.

[25]I prove that Assumption 3 holds as written, rather than any extension to exchanges of $N \geq 1$ downstream firms per upstream firm, although the same argument will establish the $N \geq 1$ case.

[26]Consistency of the maximum score estimator requires the rank order property to hold for all markets. It does not require that the same distribution of errors generate the rank order property in each market.

It is not a theorem that Assumption 3 holds if the $\psi_{ai}$ are independent and identically distributed with full support. The fact that a match appears in multiple assignments $E_h$ makes the assignment-level composite error terms statistically correlated. Eliminating non-exchangeable correlation across market outcomes is critical for the single-agent rank order property of Manski (1975) to hold for all possible deterministic payoffs and functional forms for the error terms. Of course, for any given market the error terms could have a joint distribution such that the rank order property holds. In this case, the rank order condition is a primitive assumption based upon economic intuition.

One approach to arguing for some notion of asymptotic good behavior in the number of agents in a market is to have the error terms die out as the market gets bigger.[27] A literature in empirical industrial organization, such as Ackerberg and Rysman (2006) and Bajari and Benkard (2003), argues that the variance of error terms should be decreased as the number of products increases, as each new product adds another error term and in a sense is an agent-specific product characteristic. In a matching game, each agent in a market adds as many new match-specific errors as the number of other agents, so the total number of match-specific characteristics quickly explodes.

If the number of agents in a market is $U+D$, then one way of modeling error terms (inspired by Ackerberg and Rysman (2006)) is to write $\psi_{ai} = \frac{\eta_{ai}}{U+D}$, where $\eta_{ai}$ in some base error term that has its magnitude decreased as the number of agents, $U+D$, goes to infinity. Then as the number of agents increases, the model converges to a matching game in only observable characteristics. Note that, in matching theory, adding i.i.d. match-specific errors to the true payoff of each match removes any role for the observed characteristics (types in theory) in theoretically computing the optimal assignment. Arbitrary match production levels are used in matching games with finite numbers of agents, for example Koopmans and Beckmann (1957). On the other hand and for good practical reasons, matching games with a continuum of agents almost always restrict the final production function to be a function of only agents' types, and not match-specific error terms. For example, see Shimer and Smith (2000). So the Ackerberg and Rysman suggestion of reducing the importance error terms as the number of agents increases corresponds to moving from the flexible production specifications used in the literature on games with finite numbers of agents to the type-specific production levels used in games with continua of agents.

I explore the robustness of the estimator to the presence of match-specific i.i.d. errors in a finite sample in a Monte Carlo study below.

### 5.7 Search Costs

Many matching markets have a large number of agents. In such markets, not all agents may be aware of all other agents. For example, in a marriage market, a man may meet only some subset of women in his dating life. Let the random variable "error term" $\psi_{aih}$ for market $h$ be equal to 1 if upstream firm $a$ is aware of downstream $i$ and, mutually, downstream $i$ is aware of upstream firm $a$, and 0 otherwise. Let search be costless and undirected, so that each pair of firms is mutually aware of the other with equal probability.

[27] I will not formally argue that this type of argument will give consistency in the size of a market, as the set of matches for all agents will change as new potential match partners are added to the market, and this makes proving uniform convergence of the maximum score objective function nonstandard.

One can redefine the matching jargon terms in Definition 6 so that an assignment is only physically possible if, in addition to satisfying quotas, all matched pairs of an upstream and an downstream firm are aware of each other. Let $\tilde{z}_h$ be the set of physically possible assignments given a realization of random error terms $\psi_{aih}$. Lemma 2, the equivalence of the decentralized core outcome with a social planning problem, naturally extends, except that production maximization and the core are defined to consider only set the set of assignments $\tilde{z}_h$, and deviations of coalitions of mutually aware agents.

For one-to-one matching with costless and undirected search, it can be shown that Assumption 3 holds when there is only one characteristic (say schooling) for each man and women entering production functions, and those production functions are either submodular or supermodular at all characteristic levels. Under supermodularity, core outcomes give an assortative ordering of matches. If the most highly schooled man is not aware of the most highly schooled woman, in a core assignment the man will match with the next most highly schooled woman. The assortative matching logic rules out a counterexample to Assumption 3 that can hold with non-sub or supermodular production. In the counterexample, men $a$ and $b$ match with women $j$ and $i$ respectively when $b$ is not aware of $j$, but when $b$ becomes aware of $j$, $b$ matches with $j$ and $a$ matches with some other woman $k$, instead of $i$, and even if the sum of production from man $a$ marrying woman $i$ and man $b$ marrying woman $j$ exceeds the production from man $a$ marrying $j$ and $b$ marrying $i$, which is the hypothesis from Assumption 3. The counterexample arises because production is not assortative: apparently $bi$ is a productive match, but $ai$ is not so productive, as $a$ does not match with $i$ when the very attractive option of $bj$ becomes possible and frees up $i$. Under assortative matching, $a$ would always match with $i$ when $b$ is aware of $j$ if revealed preference shows the production of $bj$ is very high, and $a$ was the match partner of $j$ when $bj$ was not possible.

The supermodularity result can be embedded in a search model with forward-looking agents. In a search model with explicit search costs, Atakan (2004) shows that supermodularity of production drives assortative matching. In a search model where costs are driven by time discounting, Shimer and Smith (2000) show that supermodularity plus some other conditions are sufficient for assortative matching. In these models, supermodularity plays two roles. First, supermodularity implies that the perfect information competitive benchmark involves assortative matching, which was shown by Becker (1973). Second, supermodularity ensures that all agents on one side of the market agree on the ordering of agents on the other side of the market, so that agents will predictably pick the best available partner from the ranked list of mutually aware partners.

Using only a univariate characteristic for each agent and assuming that the production function is either supermodular or submodular in the characteristic does not make for a rich empirical investigation. The matching estimator in this paper is most useful when there are multiple characteristics entering production and the economist does not make a priori assumptions about super or submodularity. Unfortunately, an i.i.d. search technology is not sufficient to generate the rank order property when there is not assortative matching in univariate inputs.

## 5.8 The ε-Core: Switching Costs for Deviating

Some games have empty cores, so theorists have introduced related solution concepts that are more likely to be nonempty. Kovalenkov and Wooders (2003) discuss one such solution concept, the ε-core. In an ε-core equilibrium, a coalition that wants to deviate must pay a switching cost, equal to $B\varepsilon$, where $\varepsilon > 0$ is the switching cost and $B$ is the number of firms in the coalition.

In a more general formulation, the switching cost might be coalition specific. Will match-specific switching costs provide a sufficient condition for the rank order property? There are two points to consider. First, the introduction of switching costs implies that there may be multiple equilibrium assignments. Therefore, the rank order property will be based on the unmodelled equilibrium assignment selection rule, and a formal analysis of sufficient conditions for the rank order property will not be practical. Second, match-specific switching costs induce correlation in the total switching costs for counterfactual assignments, just as match-specific production shocks induce correlation in the total unobservable production for assignments.

## 5.9 Monte Carlo Evidence on the Maximum Score Estimator

This section presents brief evidence that the maximum score estimator works well in finite samples. The goal is not to document the asymptotic argument of increasing the number of markets to infinity, but to consider variation that might be common in practice. I am especially concerned with several aspects of data that do not match the precise asymptotic consistency arguments: having data on a large number of agents in a single market, and having data that was generated with match-specific error terms.

Most commonly, researchers will use the semiparametric version of the estimator where production functions are parameterized by a finite-dimensional parameter vector $\beta$. For simplicity, the Monte Carlo study examines games of one-to-one two-sided matching when each agent is distinguished by two observables characteristics, for men, $x_{1,m}$ and $x_{2,m}$, and for women, $x_{1,w}$ and $x_{2,w}$. This game's equilibrium is in the core, as shown by Shapley and Shubik (1972), among others. The Monte Carlo study uses the parametric production function

$$f(x_m, w_w \mid \beta_1, \beta_2) = \beta_1 \times x_{1,m} \times x_{1,w} + \beta_2 \times x_{2,m} \times x_{2,w}.$$

As is standard in semiparametric discrete choice models, I impose the scale normalization that $\beta_1 = \pm 1$. The sign of $\beta_1$ is superconsistently estimable, so I set it to the true value of $+1$ throughout the study. For each gender and men as an example,

$$\begin{bmatrix} x_{1,m} \\ x_{2,m} \end{bmatrix} \sim N\left( \begin{bmatrix} 10 \\ 10 \end{bmatrix}, \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix} \right).$$

I choose high means to ensure that the values of the characteristics are almost always positive. The positive covariance between the observables suggests that using a multivariate estimator might be important for inference. I set $\beta_2 = 1.5$, so that the second observable characteristic is more important in sorting.[28]

I investigate two different specifications for the true model's error terms. The first specification is when the social planner has i.i.d. errors over marketwide assignments. In this case, each assignment $E_h$ has total production

$$\sum_m f(x_m, w_w^m \mid \beta_1, \beta_2) + \psi_{Eh},$$

where $\psi_{Eh}$ has a normal distribution with a standard deviation that is either 1 or 5. The rank order property holds under this specification. The main trouble with implementing the Monte Carlo study is that there are so many errors, $m!$ for each market, that generating the fake data is computationally much more burdensome than estimating the model.

[28] Also, I am concerned that 1.0 might be a default starting value in numerical optimization routines.

The second specification uses match-specific errors, as in Section 5.6. I sample match specific errors and solve for the optimal assignment using the linear programming problem that defines the social planner's problem. The linear programming formulation makes it much faster to generate the fake data. Like the market-specific errors case, the distribution of the errors is normal with a standard deviation of either 1 or 5.[29]

For all specifications, the maximum score objective function is numerically maximized using the global optimization routine known as differential optimization (Storn and Price, 1997).[30]

Table 1 reports estimates of the bias and root mean-squared error (RMSE) of the matching maximum score estimates under both specifications. The first two rows consider the specification with match-specific errors. For these specifications, generating fake data is problematic, so I consider twenty-five markets, each one with low numbers of agents. Both the bias and RMSE decrease when the number of agents per market increases.

The third through fifth rows consider estimation when there match-specific errors. The third and fourth rows consider a single market with a large number of agents. Again, the bias and RMSE decrease as the number of agents increases, which suggests that the matching estimator may perform well with data from a single, large market. The fifth row considers ten markets with thirty men and thirty women each, for a total of $30 \cdot 30 \cdot 10 = 9000$ possible matches. This compares to the third row, with $100 \cdot 100 = 10{,}000$ possible matches. The bias and RMSE using ten smaller markets is smaller than using one larger market, even though the number of inequalities involving two men and two women is larger in the single, larger market. Even though both estimators are misspecified, it appears averaging across markets improves the finite-sample performance of the estimator.

The last five rows repeat the earlier Monte Carlo studies with a larger standard deviation of the error term (5), to see how sensitive the above conclusions are to the relative explanatory power of the signal (observed covariates) and noise (the error terms). The RMSE's are indeed larger than the case with a standard deviation of 1, but the biases and RMSE's still decrease with the sample size. On the other hand, the benefit of having ten smaller markets seems to evaporate when the variance of the error term increases.

In these examples, the estimator does not have a large amount of bias when the precise conditions of the available consistency results are violated. The estimator seems reasonable to apply without making strong assumptions about the underlying model.

# 6 Subsampling and Smoothing for 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. 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

[29] In the first specification, for programming simplicity, I ignore the complication that individuals can remain unmatched. In the second specification, the linear programming formulation ensures that all consummated matches must provide nonzero surplus. Given the high means of both characteristics, very few of the agents are single in the fake data.

[30] For a finite sample, the objective function is a step function, and there is a continuum of global maxima, even if the parameter $\beta_2$ is point identified asymptotically. For each replication, the Monte Carlo study reports the maximum provided by the optimization routine, which is a consistent estimator under the conditions in this paper. If the maximum reported by the optimization package tends to always be near the lower bound of the set of finite-sample maxima, it could create an apparent downward, finite-sample bias. In practice, the presence of multiple global maxima in a finite sample often pales in importance when compared to the more serious concerns of inconsistency due to model misspecification.

Table 1: Monte Carlo Results, True Value $\beta_2 = 1.5$

| Errors | # Men | # Women | # Markets | Error Std. Dev. | Bias | RMSE |
|---|---|---|---|---|---|---|
| Assignment | 5 | 5 | 25 | 1 | 0.170 | 0.560 |
| Assignment | 8 | 8 | 25 | 1 | 0.073 | 0.302 |
| Match | 100 | 100 | 1 | 1 | 0.105 | 0.316 |
| Match | 200 | 200 | 1 | 1 | 0.044 | 0.116 |
| Match | 30 | 30 | 10 | 1 | 0.039 | 0.193 |
| Assignment | 5 | 5 | 25 | 5 | 0.194 | 1.51 |
| Assignment | 8 | 8 | 25 | 5 | 0.114 | 0.807 |
| Match | 100 | 100 | 1 | 5 | 0.234 | 1.09 |
| Match | 200 | 200 | 1 | 5 | 0.173 | 0.662 |
| Match | 30 | 30 | 10 | 5 | 0.252 | 0.990 |

$\sqrt{n}$) and that its limiting distribution is too complex for use in inference. However, there are several constructive suggestions for conducting asymptotic inference.

Delgado, Rodríguez-Poo and Wolf (2001) show that a resampling procedure, subsampling, consistently estimates the asymptotic distribution of test statistics (including the usual 95% confidence intervals) 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. The book Politis, Romano and Wolf (1999) provides a detailed overview of subsampling.

An alternative procedure to subsampling is to estimate a smoothed version of the maximum score estimator. Smoothing makes the objective function differential in $\beta$ for finite samples. For the single-agent binary-choice maximum score estimator, Horowitz (1992) proves that, under additional 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 model assumptions) 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 of the bootstrap-refined asymptotic distribution approximates the finite-sample distribution's coverage well. I conjecture Horowitz's results could be extended to the current matching estimators.[31]

# 7 Calibrating Using Profit and Transfer Data?

In some cases, researchers have data on the transfers ($t_{ai}$) between firms or the profits ($p_a^u$) firms receive. Many researchers have the intuition that price data can be used to calibrate the scale (cardinality) of production, as well as to estimate non-interacted production terms such as $\beta_u x_a^u + \beta_d x_i^d$ that do not contribute (if no agents can be

[31] 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.

single) to the assignments in the data. Further, transfer and profit data hold out the possibility of distinguishing the payoffs of upstream and downstream firms.

Unfortunately, error term assumptions become critical when working with price data. The Monte Carlo study shows that the maximum score estimator often performs reasonably even if the error term assumption does not fit precisely into a case where we have a consistency theorem. The reason is that consistency requires Assumption 3 to hold within a nest, and even if the rank order property is violated, the property is quite intuitive and in many cases is not likely to be violated by a large amount.

The same is not true of price data. Let the true model be match specific errors, as in Section 5.6. A selection problem will arise: the matches where we have observed profit or transfer data will tend to have attractive observables and error terms, so those matches with unattractive observables are likely to have especially high unobservable error terms. This negative correlation of $x$'s and errors for the selected sample of observed matches will make most estimates from including transfer and profit data as regressors or dependent variables inconsistent, and the finite-sample bias may be large.

For an example, say in a one-to-one matching market a researcher has data on the profits of upstream and downstream firms. The profits of upstream firm $a$ are $p_a^u$ and the profits of downstream firm $i$ are $p_i^d$. A researcher first uses the discrete assignment data and the maximum score estimator and estimates a linear-in-parameters production function $f\left(x_a^u, x_i^d\right) = \beta_{ud} x_a^u \star x_i^d$, where here $\star$ means that the researcher forms all cross products of inputs. The researcher must make a scale normalization in the maximum score stage, and wants to use the profit data to identify the production function's scale $\gamma$ in monetary units, as well as the parameters $\beta_u$ and $\beta_d$ on the uninteracted terms. Assume that there are no agent-specific nest fixed effects. Then the agent runs a profit regression to estimate $\gamma$, $\beta_u$, $\beta_d$ and the constant $\alpha$ in the model

$$p_a^u + p_i^d = \alpha + \beta_u x_a^u + \beta_d x_i^d + \gamma \beta_{ud} x_a^u \star x_i^d + \psi_{ai}.$$

If $\psi_{ai}$ is just measurement error in profits, or a expectational error in profits, then $\psi_{ai}$ is likely uncorrelated with the included characteristics of the firms $a$ and $i$. Indeed, if $\psi_{ai}$ is measurement error, one should just estimate the parameters from this regression and forget the matching data.

On the other hand, if $\psi_{ai}$ is a match-specific error observed to the firms during the matching process, $\psi_{ai}$ will likely be correlated with the $x$'s, even if in the population of the matching game all hypothetical match-specific errors are uncorrelated with agent characteristics. The reason is that this regression is being run on only observed matches: the matches that are part of an equilibrium. As $\beta_{ud} x_a^u \star x_i^d$ and $\psi_{ai}$ are substitutes in the profitability of a match, they are likely to be negatively correlated. Fixing this selection problem may require joint estimation of the profit and matching problems, and is outside the scope of this paper.[32]

# 8 Conclusions

This paper's main purpose is to prove the identification of and introduce a new nonparametric maximum score estimator for generalized versions of the matching games first studied by Koopmans and Beckmann (1957),

[32] Sørensen (2004) implements a parametric version of joint estimation for selection correction in a Gale and Shapley (1962) matching game without endogenous transfers.

Shapley and Shubik (1972) and Becker (1973), and extended by Kelso and Crawford (1982), Leonard (1983), Demange, Gale and Sotomayor (1986), Sotomayor (1992), Kovalenkov and Wooders (2003) and Ostrovsky (2004), among others. The main assumptions for these matching games are the presence of endogenous prices and additive separability between transfers and other parts of payoffs. If price-taking agents make single-agent best responses, productions functions must satisfy inequalities I call local production maximization: if an exchange of partners produces a higher production level, than it cannot be individually rational for some agent.

The matching maximum score estimator has many practical advantages over possible alternative estimation methods. First, the maximum score estimator is nonparametric, meaning that functional forms for the production function and a parametric distribution for the errors do not need to specified. Second, the matching maximum score estimator uses data on only observed matches and agent characteristics. It does not require the often unavailable data on endogenous prices, quotas and production levels. Third, the estimator allows externalities based upon the matches of other agents. Fourth, the estimator can handle multiple equilibrium assignments without exclusion restrictions, estimating equilibrium selection rules, or computing all equilibrium assignments as part of estimation.

Fifth and finally, the matching maximum score estimator is reasonably easy to compute. Evaluating the objective function involves only calculating match production levels and checking the local production maximization inequality. No nested matching mechanism needs to be solved. Also, first stage estimates of match probabilities are not needed. Most importantly, the maximum score estimator does not suffer from a curse of dimensionality in the number of agents in a market, as the estimator is consistent when a subset of matching inequalities are entered into the objective function.

# A Proofs

## A.1 Theorem 1 (Identification)

We want to show that the identified set $\mathcal{F} \subseteq \Theta$ is a singleton production function $f^0$, using the data on the observed match probabilities $P\left(E \mid X, f^0\right)$ for different markets characterized by the matrix $X$. Assume to the contrary. Then there is a $\tilde{f} \in \Theta$ such that $\tilde{f} \neq f^0$, where $P\left(E \mid X, f^0\right) = P\left(E \mid X, \tilde{f}\right)$ for all markets, except possibly for a set of markets with zero measure.

As $\tilde{f}$ is a different function than $f^0$, by Assumption 2 there exist markets where, focusing on the firms $a$, $b$, $i$ and $j$ mentioned in the statement of the theorem, as well as the definition of $\tilde{E}$,

$$f^0(a, M_a \mid E) + f^0(b, M_b \mid E) > f^0\left(a, (M_a \backslash \{i\}) \cup \{j\} \mid \tilde{E}\right) + f^0\left(b, (M_b \backslash \{j\}) \cup \{i\} \mid \tilde{E}\right)$$

$$\tilde{f}(a, M_a \mid E) + \tilde{f}(b, M_b \mid E) < \tilde{f}\left(a, (M_a \backslash \{i\}) \cup \{j\} \mid \tilde{E}\right) + \tilde{f}\left(b, (M_b \backslash \{j\}) \cup \{i\} \mid \tilde{E}\right),$$

or the reverse inequalities ($<$ and then $>$). Part 1 of Assumption 2 rules out that evaluating a local social maximization inequality at different sets of firms always produces the same value for $f^0$ and $\tilde{f}$.

The inequalities are strict because of the continuous covariate, Assumption 4, and the continuity of production functions, Assumption 2. Focus on the direction of the inequalities in the displayed equations. By Assumption

3, the rank order property, at this market, the production function $f^0$ predicts that

$$P\left(E \mid X, f^0\right) > P\left(\tilde{E} \mid X, f^0\right)$$

while $\tilde{f}$ predicts that

$$P\left(E \mid X, \tilde{f}\right) < P\left(\tilde{E} \mid X_h, \tilde{f}\right).$$

This is a contradiction if there exists a positive measure of such markets. This is immediate because Assumption 2 states that production functions are continuous, and Assumption 4 states that at least one covariate for upstream firms has full support and is continuously varying.

## A.2 Theorem 2 (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 applies to general maximization problems and does not require that an element of the parameter space be a finite vector. Optimization over function space is allowed. The theorem has four conditions:

1. The probability limit of the subset maximum score objective function, $Q_\infty(f)$, has a unique global maximum at the true production function, $f^0$ (constructive identification).
2. The parameter space $\Theta$ is compact.
3. The probability limit of the objective function, $Q_\infty(f)$, is continuous in $f$.
4. The objective function converges uniformly in probability to its limit.

Condition 2, compactness, is satisfied by Assumption 2.

### A.2.1 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 observe infinitely many markets with the same number of agents and identical characteristics, all captured by $X_h$. By a law of large numbers and the law of iterated expectations,

$$\text{plim}_{H\to\infty}\left(\frac{1}{H}\sum_{h=1}^{H} 1\left[E \,\text{in}\, h\right]\right) = E_{X,\psi}\left\{1\left[E\right]\right\} = E_X E_\psi \left\{1\left[E\right] \mid X\right\} = E_X\left\{P\left(E \mid X, f^0\right)\right\},$$

where $\psi_h$ is the vector of all stochastic terms in the market, and the true production function $f^0$ 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. A similar argument shows that the limit of $Q_H(f)$ as the number of markets, $H$, goes to infinity is

$$Q_\infty(f) = E_X \sum_{z \in B^{\text{sub}}(X)} P\left(E \mid X, f^0\right) \times 1\left[f(a, M_a \mid E) + f(b, M_b \mid E) > f\left(a, (M_a \setminus \{i\}) \cup \{j\} \mid \tilde{E}\right) + f\left(b, (M_b \setminus \{j\}) \cup \{i\} \mid \tilde{E}\right)\right]$$
$$+ P\left(\tilde{E} \mid X, f^0\right) \times 1\left[f(a, M_a \mid E) + f(b, M_b \mid E) < f\left(a, (M_a \setminus \{i\}) \cup \{j\} \mid \tilde{E}\right) + f\left(b, (M_b \setminus \{j\}) \cup \{i\} \mid \tilde{E}\right)\right],$$

where the derivation uses a law of large numbers and the law of iterated expectations as well as the fact that the local production maximization inequality does not depend on $\psi$ and can be factored out of the expectation with respect to $\psi$, conditional on $X$.

I prove that $Q_\infty(f)$ has a global maximum at the true production function $f^0$ by first proving the integrand evaluated at particular set of the characteristics of all agents in a market, $X$, is globally maximized at $f^0$. If the integrand is indeed maximized for all $X$, except for a set with probability 0, then when $Q_\infty(f)$ is computed by integrating out $X$, the value of the integral will be maximized at $f^0$.

Therefore, fix $X$. For each $z \in B^{\text{sub}}(X)$, two additively separable terms appear in $Q_\infty(f)$: once where $P\left(E \mid X, f^0\right)$ multiplies an inequality involving production functions, and once where $P\left(\tilde{E} \mid X, f^0\right)$ multiplies the opposite inequality. First, under the covariate Assumption 4,

$$f(a, M_a \mid E) + f(b, M_b \mid E) = f\left(a, (M_a \setminus \{i\}) \cup \{j\} \mid \tilde{E}\right) + f\left(b, (M_b \setminus \{j\}) \cup \{i\} \mid \tilde{E}\right)$$

with probability 0, as each match upstream firm's characteristics has a freely varying characteristic conditional on the characteristics of the other firms. As the inequalities in $Q_\infty(f)$ are strict, such points do not contribute to the objective function, but as they occur with probability 0, choosing an alternative parameter vector $\tilde{f}$ to make one or the other arrangement for $i$ and $j$ have a greater sum of production levels will not increase the value of $Q_\infty(f)$.

I can restrict attention to the cases where one of the sums of production levels is strictly greater than the sum of production levels with the exchange of partners for downstream firms $i$ and $j$. Notice that the two inequalities involving sums of production levels 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 $z \in B^{\text{sub}}(X)$ is a global maximum of the integrand evaluated at $X$. By Assumption 3, the rank order property, the true production function $f^0$ implements this assignment. As $X$ is arbitrary, the integrand for a given $z \in B^{\text{sub}}(X)$ is globally maximized at all points, other than a set of measure 0, by $f^0$. As $z$ is arbitrary, $Q_\infty(f)$ is globally maximized at $f^0$.

Note that there is an strong inequality in the indicator function in the objective function, so that $f = 0$ for all possible matches is a global minimum and not a global maximum.

The next step of the proof is to show that the global maximum of $Q_\infty(f)$, $f^0$, is unique. This argument is the same as the proof of Theorem 1, identification. For some possible other global maximum, $\tilde{f} \in \Theta$, the proof of Theorem 1 shows that there is a set of markets where $\tilde{f}$ gives inconsistent predictions according to the rank order property, Assumption 3. By Assumption 4, this set has positive measure. So $\tilde{f}$ implements a sub-optimal series of match probabilities to enter $Q_\infty(f)$, and thus cannot be a global maximum of $Q_\infty(f)$.

### A.2.2 Continuity of the Limiting Objective Function and Uniform Convergence

The two-sided matching maximum score objective function is not continuous in $f$. Condition 3 is that the probability limit of the objective function, $Q_\infty(f)$, is continuous in $f$. Lemma 2.4 from Newey and McFadden (1994) can be used to prove continuity of $Q_\infty(f)$ as well as uniform in probability convergence of $Q_H(f)$ to $Q_\infty(f)$, 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 even if we view the number of upstream and downstream firms as random; that the parameter space $\Theta$ is compact (part of Assumption 2), that the terms for each market are continuous with probability 1 in $f$; and that the terms for each market are bounded by a function whose mean is not infinite. While the terms for each market are not continuous in $f$ because of the indicator functions, they are continuous with probability 1 by the support condition on the covariates, Assumption 4. As the continuous covariate $x^u_{1,a}$ is freely varying conditional on the other covariates ties in the inequalities in the objective function happen with probability 0.

The other condition we need to verify to apply Lemma 2.4 is that the market-specific inequalities are bounded by a function with a non-infinite mean. The score of correct predictions for a market can be at most the number of inequalities included in $B^{\text{sub}}(X_h)$, which itself can be no larger than the number of combinations of sets of agents and two members from those sets, which is large but finite if the number of agents in a market is finite. Assumption 6 states that the mean number of such inequalities is not infinite.

## A.3 Lemma 3 (Exchangeability Sufficient for the Rank Order Property)

I first derive an explicit formulation for choice probabilities in terms of the density function for the exchangeable errors. Let $\psi$ be the vector of error terms for all marketwide physical assignments. The condition for an assignment $E_h$ to be optimal is seen in equation (10). Writing equation (10) out in more detail gives

$$P(E_h \mid X_h) = \int_{-\infty}^{\infty} \prod_{Z \neq E_h} \int_{-\infty}^{\sum_{a \in U} f\left(a, M_a^{E_h}\right) + \psi_{E_h} - \sum_{a \in U} f\left(a, M_a^{Z_h}\right)} f(\psi \mid X)\, d\psi,$$

or an integral with as many dimensions as marketwide assignments. The upper limit of the integrals is strictly increasing in the deterministic payoff for choosing marketwide assignment $E_h$, $\sum_{a \in U} f\left(a, M_a^{E_h}\right)$, and one upper limit is strictly decreasing in the marketwide assignment for all $Z \neq E_h$. Because the joint density $f(\psi \mid X)$ is exchangeable, the function $P(E_1 \mid X_h)$ is the same as $P(E_2 \mid X_h)$ for two different assignments $E_1$ and $E_2$, except where the payoffs of $E_1$ and $E_2$ enter. 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. The strictness of the inequalities come from the full support portion of Assumption 10, which states that the support of the error terms will always be larger than the support of the data, so that one does not "run out" of error terms.

The above argument is the same with agent-specific nest fixed effects included, because the only difference between $\tilde{E}$ and $E$ is the matches $ai$ and $bj$ in $E$ are reversed in $\tilde{E}$, and $a$ and $b$ are in the same nest, and $i$ and $j$ are in the same nest. The fixed effects are the same for the assignments $E$ and $\tilde{E}$.

The "only if" part of Assumption 3 follows easily by reversing the above arguments: if two functions are the same, except that the first function value is larger, the first function must have a larger argument if the functions are increasing in their arguments.

## 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.

**Ackerberg, Danial and Marc Rysman**, "Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects," *RAND Journal of Economics*, 2006.

**Atakan, Alp E.**, "Assortative matching with explicit search costs," October 2004. working paper.

**Bajari, Patrick and C. Lanier Benkard**, "Discrete Choice Models as Structural Models of Demand: Some Economic Implications of Common Approache," 2003. working paper.

**_ and Jeremy T. Fox**, "Collusion and Complementarities in an FCC Spectrum Auction," April 2005.

**_ , Han Hong, and Stephen Ryan**, "Estimation of Discrete Games of Complete Information," 2004. working paper.

**Bayer, Patrick, Robert McMillan, and Kim Reuben**, "An Equilibrium Model of Sorting in an Urban Housing Market," June 2004. working paper.

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

**Berry, Steven T., James Levinsohn, and Ariel Pakes**, "Automobile Prices in Market Equilibrium," *Econometrica*, July 1995, *63* (4), 841–890.

**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.

**Choo, Eugene and Aloysius Siow**, "Who Marries Whom and Why," November 2003. working paper.

**Ciliberto, Federico and Elie Tamer**, "Market Structure and Multiple Equilibria in Airline Markets," December 2003. working paper.

**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.

**Demange, Gabrielle, David Gale, and Marilda A. Oliveira Sotomayor**, "Multi-Item Auctions," *The Journal of Political Economy*, August 1986, *94* (4), 863–872.

**Fox, Jeremy T.**, "Semi and Nonparametric 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.*

**Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey**, "Regular Quantal Response Equilibrium," 2004. working paper.

**Haile, Phillip, Ali Hortaçsu, and Grigory Kosenok**, "On the Empirical Content of Quantal Response Equilibrium," December 2004. working paper.

**Hatfield, John William and Paul R. Milgrom**, "Matching with Contracts," *The American Economic Review*, September 2005, *95* (4), 913–935.

**Hitsch, Günter, Ali Hortaçsu, and Dan Ariely**, "What Makes You Click? An Empirical Analysis of Online Dating," 2005. working paper.

**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.

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

**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.

**Kovalenkov, Alexander and Myrna Holtz Wooders**, "Advances in the Theory of Large Cooperative Games and Application to Club Theory: the Side Payments Case," in Carlo Carraro, ed., *The Endogenous Formation of Economic Coalitions*, Edward Elgar: Chelenham, 2003.

**Leonard, Herman B.**, "Elicitation of Honest Preferences for the Assignment of Individuals to Positions," *The Journal of Political Economy*, June 1983, *91* (3), 461–479.

**Lewbel, Arthur**, "Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables," *Journal of Econometrics*, 2000, *97*, 145–177.

**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.**, "Least Concavity and the Distribution-Free Estimation of Nonparametric Concave Functions," September 1990. Cowles Foundation Discussion Paper 958.

—, "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," *Econometrica*, September 1991, *59* (5), 1315–1327.

—, "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica*, March 1992, *60* (2), 239–270.

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

**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.

**Pakes, Ariel, Jack Porter, Katherine Ho, and Joy Ishii**, "Moment Inequalities and Their Application," April 2005. working paper.

**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.

**—, —, and Michael Wolf**, *Subsampling*, Springer, 1999.

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

**Shimer, Robert and Lones Smith**, "Assortative Matching and Search," *Econometrica*, 2000, *68* (2), 343–369.

**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.

**Storn, Rainer and Kenneth Price**, "Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," *Journal of Global Optimization*, 1997, *115*, 341–359.