NBER WORKING PAPER SERIES

ESTIMATING MATCHING GAMES WITH TRANSFERS

Jeremy T. Fox

Working Paper 14382
http://www.nber.org/papers/w14382

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2008

Thanks to SupplierBusiness as well as Thomas Klier for help with the empirical application. I thank the National Science Foundation, the NET Institute, the Olin Foundation, and the Stigler Center for generous funding. Thanks to helpful comments from colleagues, referees and workshop participants at Aarhus, Bar-Ilan, Caltech, Carlos III, Cornell, the Econometric Society World Congress, Haifa, Harvard, Illinois, Minnesota, North Carolina, Northwestern, Northwestern Kellogg, the NY Fed, NYU, Ohio State, Penn State, Pittsburgh, the Society of Economic Dynamics, the Stanford Institute for Theoretical Economics, Tel Aviv, Toronto, Toulouse, Vanderbilt, Virginia, Washington University and Wisconsin. Chenchuan Li, David Santiago and Louis Serranito provided excellent research assistance. Email: fox@uchicago.edu. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Estimating Matching Games with Transfers
Jeremy T. Fox
NBER Working Paper No. 14382
October 2008
JEL No. C1,C14,C71,D85,L22,L62

ABSTRACT

Economists wish to use data on matches to learn about the structural primitives that govern sorting. I show how to use equilibrium data on who matches with whom for nonparametric identification and semiparametric estimation of match production functions in many-to-many, two-sided matching games with transferable utility. Inequalities derived from equilibrium necessary conditions underlie a maximum score estimator of match production functions. The inequalities do not require data on transfers, 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. I present an empirical application to automotive suppliers and assemblers.

Jeremy T. Fox
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
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 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 and bidders to multiple objects for sale in an auction. Theoretical work is also ongoing on models of one and many-sided matching.

Matching games are distinguished from simpler models of markets because agents on all sides of the market make a limited number of matches. Either agents may be able to make an exogenously limited number of matches, or nonlinearities in match payoffs may endogenously limit the number of matches of any given party. Either way, agents on the same side of the market are rivals to match with agents on the other side. In marriage, each woman can have only one husband, so men compete to marry the most attractive women.

Matching games are inviting frameworks for empirical work as the models apply to a finite number of agents with flexible specifications for the production functions generating match output. A typical dataset for a matching market lists a series of observed matches and some characteristics about the parties in each match. Economists assume the data come from a market in equilibrium and want to estimate the production function generating match output for observed and counterfactual matches.

This paper provides a structural estimator for the production function that gives the total output of a match as a function of observable agent characteristics. This production function subsumes individual agent preferences in a transferable utility matching game, a game where matched agents exchange monetary transfers as part of a price-taking matching equilibrium. The match production function governs who matches with whom, the dependent variable data I use for estimation. I do not use data on the equilibrium transfers, although matched agents exchange such monies in the economic model. Using data on only observed matches is helpful for studying markets such as marriage, where the idea of exchanging money in a market setting is an approximation to how resources are allocated in a household, as well as for studying relationships between firms, where the monies exchanged are often private contractual details.

I present an empirical example from industrial organization. I use data on the identities of the suppliers of individual car parts for particular car models. In this upstream-downstream market, a match is a car part for a specific car model, and the two sides of the market are car part suppliers, like Bosch and Delphi, and the final assemblers of cars, like General Motors and Toyota. Suppliers typically produce many different car parts. I focus on two related empirical questions. First, I estimate the returns to specialization from the viewpoint of the supplier. I estimate how these returns from specialization vary at the levels of producing car parts for a particular car, for a particular brand of car, for a particular assembler (parent company) and for a particular home region for a brand (Europe, North America, Asia). Second, I examine whether suppliers that can meet the quality levels of Asian assemblers (Honda and Toyota receive the highest quality ratings from sources such as *Consumer Reports*) are better able to compete and win contracts from non-Asian suppliers. If so, this pattern of sorting is consistent with an explanation where matching with Toyota makes a supplier higher quality or with an explanation where Toyota only matches with high-quality suppliers. Either way, matching with Toyota coincides with a competitive edge that helps win business from non-Asian assemblers. In other words, the empirical evidence is compatible with a story where matching with one type of firm (Asian assemblers) may be complementary with matching with other firms (non-Asian assemblers).

The methodological contributions are split into identification and estimation. Match data comes from the outcome to a market, which intermingles the preferences of all participating agents and finds an equilibrium. An agent may not match with its most preferred partner because that partner is taken. Given this rivalry for partners, it is not obvious what types of economic parameters are identified from having equilibrium outcome data from matching markets. Identification asks the question of just what economic parameters can be learned from data on who matches with whom? Aspects of match production functions can be identified in a transferable utility setting. I explore the nonparametric identification of match production functions using data on only equilibrium matches. Identification relies on inequalities implied by the equilibrium concept known as pairwise stability.[1] The results are split into results on cardinal and ordinal identification. Cardinal identification is probably more important in many empirical applications. Cardinal identification using qualitative match data arises because cardinal properties of match production functions govern sorting patterns in transferable utility matching games. I extend the results of Becker (1973) in several dimensions. For example, I show how to identify the relative importance of the components of a vector of agent characteristics in a match production function. This allows a researcher to measure which characteristics drive the sorting patterns seen in the data.

Ordinal identification asks whether a researcher can identify the relative ordering of match production for different types of matches. Here, I extend identification results from the semiparametric multinomial choice (maximum score) literature by Manski (1975) and especially Matzkin (1993). The extension is non-trivial because one cannot freely vary the choice set of a single agent when using data that are the equilibria to matching games. In matching, one can only choose a partner if the partner agrees to match with you. I prove the ordinal identification of match production functions by varying the exogenous distribution of the types of agents in a matching market.

My identification arguments do not require data on objects that are not found in many datasets but are important in matching models: the endogenous prices, the number of physical matches that an agent can make (quotas), or continuous outcomes such as production levels, revenues and profits. Quotas are often a modeling abstraction in many-to-many matching; not requiring data on such an abstraction is an advantage. The prices transferred between upstream and downstream firms are often private contractual details and are not observed. Estimation without data on this endogenous variable will be needed for many empirical applications.

For estimation, I provide a computationally simple maximum score estimator for match production functions. The estimator uses inequalities derived from necessary conditions for pairwise stability.[2] Evaluating the statistical objective function is computationally simple: checking whether an inequality is satisfied requires only evaluating match production functions and pairwise comparisons. Numerically computing the global maximum of the objective function is somewhat harder (it requires a global optimization routine) although estimation is certainly doable with software built into commercial packages such as Matlab or Mathematica.

There are three conceptual insights that are important for practical matching estimation. First, using maximum score avoids the need to non-parametrically estimate conditional match probabilities in a first stage, as a

[1] The cooperative game theoretic equilibrium concept of pairwise stability is not a special case of the noncooperative Nash solution concept, which Bajari, Hong and Ryan (2007b), Ciliberto and Tamer (2007) and Pakes, Porter, Ho and Ishii (2006) among others provide estimators for. Here I exploit the pairwise stability solution concept to investigate nonparametric identification under partial observation of the equilibrium outcomes (matches, not prices). The noncooperative papers do not study nonparametric identification.

[2] There is tradition of using necessary conditions and inequalities to estimate complex games. See Haile and Tamer (2003) and Bajari, Benkard and Levin (2007a). No previous research has used maximum score, the original estimator based on inequalities, to estimate complex games.

finite-sample analog of the identification arguments would seem to require. The conditioning arguments are all the observable exogenous agent characteristics in a matching market and would be of extremely high dimension in any dataset of economic importance. Therefore, maximum score avoids a data curse of dimensionality in the size of the matching market, arising from needing to estimate match probabilities nonparametrically. The second insight is that maximum score avoids having to compute an equilibrium to a matching market as part of evaluating a statistical objective function. Only necessary conditions (inequalities) from equilibrium are imposed. Third, maximum score avoids having to manually compute integrals over econometric unobservables, although the underlying economic model allows these unobservables to affect the matches that are observed, to an extent that will be discussed in detail.

A traditional simulated method of moments or simulated maximum likelihood estimator interacts the last two issues: inside the statistical objective function, a solution to a matching game must be computed for each combination of a empirically distinct matching market and a vector of simulation draws for all the econometric unobservables for that market. This interaction of nested solutions and integrals over error terms creates a computational curse of dimensionality in the size of matching markets. To understand the logic behind the combinatorics, let there be 3 men and 3 women in a marriage market. None of the agents can be single (for simplicity only). Let the ordered pair $\langle 1,2\rangle$ refer to a marriage between man 1 and woman 2. It turns out that there are $3^2 = 9$ possible marriages that can happen, which are

$$\langle 1,1\rangle, \langle 1,2\rangle, \langle 1,3\rangle, \langle 2,1\rangle, \langle 2,2\rangle, \langle 2,3\rangle, \langle 3,1\rangle, \langle 3,2\rangle, \langle 3,3\rangle.$$

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

$$\{\langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle\}, \{\langle 1,1\rangle, \langle 2,3\rangle, \langle 3,2\rangle\}, \{\langle 1,2\rangle, \langle 2,1\rangle, \langle 3,3\rangle\}, \{\langle 1,2\rangle, \langle 2,3\rangle, \langle 3,1\rangle\}, \{\langle 1,3\rangle, \langle 2,1\rangle, \langle 3,2\rangle\}, \{\langle 1,3\rangle, \langle 2,2\rangle, \langle 3,1\rangle\}.$$

Now let there be 100 men and 100 women in a marriage market. There are $100^2 = 10,000$ matches and $100! = 9.33 \times 10^{157}$ market-wide assignments. The number of atoms in the universe is much lower, at around 10^{79}, than the number of possible assignments. Forming the probability that the observed assignment represents the market's equilibrium assignment will not be possible if a matching mechanism must be solved inside a $100^2 = 10,000$ dimensional numerical integral over match-specific unobservables, and this integral must be repeatedly evaluated at different structural parameter values in an outer optimization routine.

This paper relies on one non-primitive assumption about the stochastic structure of the model, which will be discussed in detail soon. I feel the downside of the non-primitive nature of this assumption is far outweighed by the benefits it gives. Throughout the paper, I will argue that, because of the above combinatorics, writing down a parametric distribution and computing a likelihood by numerical integration is computationally prohibitive. In particular, the combinatorics of checking whether an assignment satisfies known necessary conditions to be a stable assignment (with no transfer data) become much worse with many-to-many matching, which is the case I study in the automotive supplier empirical example and in a separate application to FCC spectrum auctions (Bajari and Fox, 2007). The maximum score estimator in this paper 1) naturally generalizes concepts such as assortative matching from the matching literature, 2) gives full support to the dependent variable data on match assignments and so cannot be easily rejected, 3) does not suffer from a computational curse of dimensionality in the size of the market, and 4) does not suffer from a data curse of dimensionality. As a bonus, the maximum score framework allows the study of nonparametric identification of matching games, which is important as it is not obvious what can learned about structural parameters from data on who matches with whom. This is the first paper to study nonparametric identification in any type of matching game.

The paper is organized as follows. Section 2 provides a brief overview of some results from matching theory and discusses the role of matching estimation in empirical work. Section 3 outlines a many-to-many matching game. Section 4 discusses how to add econometric error terms to a matching game. Section 5 discusses cardinal identification and Section 6 discusses ordinal identification. Section 7 discusses estimation with the maximum score estimator and relevant asymptotic results. Section 8 provides Monte Carlo evidence about the performance of the estimator. Section 9 is the empirical application to automotive suppliers and assemblers. Section 10 sits somewhat apart from the remainder of the paper; the section provides an alternative model and asymptotic theory for the practical situation of having one large matching market, rather than a number of separate markets. Having developed my approach to estimation and identification in detail, Section 11 compares the approach to some parametric estimators for various types of matching games that have been recently introduced in the literature.

2 One-to-one, two-sided matching games with transfers

Not all readers will be intimately familiar with matching theory. This section covers some classic results and then shows how identification, estimation and numerical analysis are needed to extend them.

2.1 The sorting characterization of Becker

In two-sided matching models, agents are rivals to match with the most attractive partners on the other side of the market. In a classic paper, Becker (1973) studies the matching of men and women in marriage. The model is static, there is perfect information, the agents are price takers and each agent can have at most one match, of an opposite gender. There are a finite set of men indexed by a scalar type x^m and a finite set of women indexed by a scalar type x^w.[3] Think of x^m as the years of schooling of a man and x^w as the years of schooling of a woman. Assume x^m and x^w have continuous support. A matching is the ordered pair $\langle x^m, x^w \rangle$.[4] Each person can have at most one match; $\langle 0, x^w \rangle$ represents a single woman and $\langle x^m, 0 \rangle$ is a single man.

If a couple x^m and x^w marries, their marriage produces total output $f(x^m, x^w)$. I call $f(x^m, x^w)$ the production function. Married men and women split the production from their marriage in a market equilibrium. If p_x^m is the equilibrium payoff to man x^m and p_x^w is the equilibrium payoff to woman x^w, it is a result that $p_x^m + p_x^w = f(x^m, x^w)$ if x^m and x^w are married and $p_x^m + p_x^w \leq f(x^m, x^w)$ if they are not married. Otherwise, the equilibrium is not pairwise stable as x^m and x^w would prefer to deviate and match with each other. We say men and women have transferable utility if they can express utility in terms of money.

Becker (1973) uses matching theory to predict the pattern of sorting between men x^m and women x^w. Assortative matching occurs when men with high levels of schooling match with women with high levels of schooling. Anti-assortative matching is when high-schooled men match with low-schooling women, and vice

[3] Sattinger (1979) studies a similar model with a continuum, rather than a finite number, of agents. Tinbergen (1947) did early work on a related model.

[4] For matches, I use $\langle \cdot, \cdot \rangle$ for an ordered pair instead of $(\cdot, \cdot)$, as matches are a special object in this paper. The convention is that the characteristics of the man appears in the first coordinate and the characteristic of the woman is the second coordinate.

versa. Assortative matching will occur if x^m and x^w are complements over the entire support, or

$$\frac{\partial f\left(x^m, x^w\right)}{\partial x^m \partial x^w} > 0 \; \forall x^m \in \mathbb{R}, x^w \in \mathbb{R}, \tag{1}$$

when f is twice differentiable. Assortative matching occurs because high-x^m men and high-x^w women produce incrementally more output together. Likewise, anti-assortative matching occurs when x^m and x^w are substitutes, or $\frac{\partial f(x^m, x^w)}{\partial x^m \partial x^w} < 0$. If $\frac{\partial f(x^m, x^w)}{\partial x^m \partial x^w} = 0$, say when $f\left(x^m, x^w\right) = x^m + x^w$, then any sorting pattern is compatible with equilibrium.[5]

The Becker result can be used in empirical work. Say the researcher has a dataset from an equilibrium to a one-to-one, two-sided matching market with scalar types. If high type agents match, the researcher can conclude x^m and x^w are (weakly) complements. If there is anti-assortative matching, one may conclude that x^m and x^w are substitutes. No information on p_x^m and p_x^w is needed for this inference, even though in the model matched agents exchange these monies in equilibrium. This as an advantage: we can learn about the marriage production function $f\left(x^m, x^w\right)$ with data on only who matches with whom.

2.2 Equilibrium computation

Koopmans and Beckmann (1957) and Shapley and Shubik (1972) show how to compute an equilibrium to a one-to-one matching model with a finite number of men and a finite number of women. The production of a match between man a and woman i is given as $f_{\langle a,i \rangle}$, which is unrestricted to be $f_{\langle a,i \rangle} = f\left(x_a^m, x_i^w\right)$, where x_a^m and x_i^w are scalar types, although the types of Becker are a special case. In this transferable utility world, the equilibrium matching maximizes the sum of economy-wide production, so the equilibrium marriages can be found using a social planner's problem, here a linear programming problem. Likewise, the equilibrium payoffs p_a^m for man a and p_i^w for woman i can be calculated by using the dual linear program to the social planner's problem. See Roth and Sotomayor (1990, Chapter 8) for details on these linear programming formulations.

2.3 Limitations of the Becker sorting characterization

The Becker (1973) sorting characterization is the first instance of using matching models in empirical work. However, the subsequent theoretical matching literature has not been able to analytically characterize the sorting between agents on two sides of a market under these relevant generalizations: 1) Each type x^m or x^w is a vector with unrestricted distributions of x^m and x^w; 2) the type $f_{\langle a,i \rangle}$ is match and not agent specific; 3) each agent can make more than one match, up to some agent-specific quota q_a^m or q_i^w; 4) payoffs are not additively separable across these multiple matches; and 5) the sign of $\frac{\partial f(x^m, x^w)}{\partial x^m \partial x^w}$ is not constant over the support of x^m and x^w. Consider vector types x^m and x^w: an agent may have a religion, income, physical appearance in addition to the earlier years of schooling. The theoretical literature cannot analytically characterize the equilibrium sorting using simple properties like complements and substitutes.

In the situations without analytical characterizations of sorting patterns, numerical analysis is needed to compute an equilibrium for a given production function $f\left(x^m, x^w\right)$ and distributions of agent characteristics $G^m\left(x^m\right)$ and $G^w\left(x^w\right)$. In some models, the theoretical literature shows it is possible to generalize the linear programming tools above to compute an equilibrium.

[5] The equilibrium assignment (matching) of men to women will be unique if there is not a non-generic case such as $f\left(x^m, x^w\right) = x^m + x^w$. The prices p_x^m and p_x^w will typically lie in intervals rather than being unique.

2.4 Uses of matching estimation

The production function $f(x^m, x^w)$ is typically not known in empirical applications. This paper shows how to use data on who matches with whom in equilibrium to estimate match production functions like $f(x^m, x^w)$. 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. For marriage, we can measure the relative importance of religion, income, schooling and physical appearance in the production of a match.

A related goal is to distinguish the contributions of match production functions and the distribution of exogenous agent characteristics, $G^m(x^m)$ and $G^w(x^w)$, in the equilibrium sorting. For example, Choo and Siow (2006) estimate changes in the sorting patterns between broad types of men and women across decades in the United States, and in part ask whether changes in match production functions are behind the differences in sorting patterns.

As with any structural paper, estimating a match production function is a key step towards being able to compute counterfactual equilibria. A researcher can use the computational tools discussed above to simulate market equilibria under the same $f(x^m, x^w)$ but different $G^m(x^m)$ and $G^w(x^w)$ or a different $f(x^m, x^w)$ and the same $G^m(x^m)$ and $G^w(x^w)$. For example, a researcher could estimate match production functions in a many-to-many matching market with scale economies, and then recompute the equilibrium to the market when the production function is changed to not allow scale economies.

3 Many-to-many matching games

I am the first empirical researcher to study many-to-many matching without additive separability in an upstream firm's payoffs across multiple downstream firm partners. These interactions in payoffs across partners are the key behind many empirical issues, as the empirical application to car parts suppliers and assemblers will illustrate. This section outlines a two-sided, many-to-many matching game without econometric errors. Simpler models such as marriage are special cases. In the next section, I discuss how to extend these models to introduce econometric error terms.

Some theoretical results on one-to-one, two-sided matching with transferable utility have been generalised by Kelso and Crawford (1982) for one-to-many matching, Leonard (1983) and Demange, Gale and Sotomayor (1986) for multiple-unit auctions, as well as Sotomayor (1992), Camiña (2006) and Jaume, Massó and Neme (2007) for many-to-many matching with additive separability in payoffs across multiple matches. These models are applications of general equilibrium theory to games with typically finite numbers of agents. Dagsvik (2000) and Choo and Siow (2006) pioneered the structural estimation of one-to-one (marriage), logit-based empirical matching models with discrete agent types. The identification strategy used in this paper can be extended to the cases studied by Kovalenkov and Wooders (2003) for one-sided matching, Ostrovsky (2004) for supply chain, multi-sided matching, and Garicano and Rossi-Hansberg (2006) for the one-sided matching of workers into coalitions known as firms with hierarchical production.[6] 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.

[6]Lucas (1978) and Rosen (1982) are predecessors to Garicano and Rossi-Hansberg (2006).

3.1 Matching markets

Consider an example where automobile assemblers (think General Motors and Toyota) match with automotive parts suppliers (think Bosch and Johnson Controls). Several exogenous objects define a matching market. Let U be a finite set of upstream firms, indexed by i. Let D be a finite of downstream firms, indexed by a. Let $Q : U \cup D \to \mathbb{N}_+$ be the set of quotas, where $q_a^u \in Q$ is the quota of an upstream firm a and $q_i^d \in Q$ is the quota of the downstream firm i. A quota represents the maximum number of physical matches that a firm can have. Let X be the collection of all payoff-relevant exogenous characteristics. I will be specific about the elements of X below. A matching market also has the exogenous preferences of agents, which I will also discuss below.

The space of matches is $(U \cup \{0\}) \times (D \cup \{0\})$. Let $\mu = \langle a, i \rangle$ be a match between downstream firm or automobile assembler a and upstream firm or car parts supplier i. As before, $\langle a, 0 \rangle$ refers to an unfilled quota slot for an assembler and $\langle 0, i \rangle$ refers to an unfilled quota slot for a supplier. A matching market outcome is a tuple (A, T). An assignment, or a finite collection of matches, is an element of the power set of $(U \cup \{0\}) \times (D \cup \{0\})$. For any assignment $A = \{\mu_1, \mu_2, \ldots\}$, $T = \{t_{\mu_1}, t_{\mu_2}, \ldots\}$ is a set of transfers for all matches in A. Each $t_{\mu_1} \in \mathbb{R}$ and represents a payment for a downstream firm to an upstream firm. In a marriage market with 100 matched couples, A is a finite set of 100 marriages and T is a finite set of 100 transfers between each of the married couples. A key concern that will come up at several points is that the number of matches in an assignment is often very large in a matching game. Altogether, the combination of the exogenous and endogenous elements of a matching market form the tuple (D, U, Q, X, A, T).

Given an outcome (A, T), the payoff of $i \in U$ is

$$r^u\left(\vec{x}\left(i, C_i^u(A)\right)\right) + \sum_{a \in C_i^u(A)} t_{\langle a, i \rangle}. \tag{2}$$

The payoff at (A, T) for $a \in D$ for the match $\langle a, i \rangle \in A$ is $r^d\left(\vec{x}\left(i, \{a\}\right)\right) - t_{\langle a, i \rangle}$. I use the convention that the car assembler is sending positive transfers to the car parts supplier, but the notation allows transfers to be negative. I will now explain each of the elements of the notation that enters these payoffs.

$C_i^u \subseteq D \cup \{0\}$ is a coalition of downstream firms that may match with upstream firm i. C^u with no arguments is an arbitrary collection of downstream firms. If the assignment A is an argument to the function $C_i^u(A)$, then

$$C_i^u(A) \equiv \begin{cases} \{a \in D \mid \langle a, i \rangle \in A\} & \text{if } \{a \in D \mid \langle a, i \rangle \in A\} \neq \emptyset \\ \{0\} & \text{if } \{a \in D \mid \langle a, i \rangle \in A\} = \emptyset \end{cases}$$

is the set of downstream firms matched to upstream firm i at the assignment A. $C^d \subseteq U \cup \{0\}$ and $C_a^d(A)$ have similar interpretations for downstream firms.

A feasible assignment A is one that is under quotas for all agents. This means $C_i^u(A) \leq q_i^u \; \forall i \in U$ and $C_a^d(A) \leq q_a^d \; \forall a \in D$.[7] Quotas ensure that firms are rivals to match with the most attractive partners on the other side, as opposed to all firms choosing the most attractive partner.[8]

This paper studies matching in characteristic space. Let $\vec{x}(i, C^u)$ be the vector of characteristics corresponding to the set of matches involving firm $i \in U$ and the set of potential downstream firm partners in C^u.[9] I will

[7] The maximum quota q_i^u of an upstream firm will determine the maximum number of arguments of $f(\vec{x})$, the primary object of estimation. When upstream firm i does not use all of its quota, null arguments can be included in the argument vector $\vec{x}$ of $f(\vec{x})$ to refer to the unfilled match slots.

[8] Quotas are not necessary; some other explanation such as decreasing returns to scale may explain why all matches do not occur.

[9] I use the vector notation $\vec{x}$ for characteristics only. Later I will refer to the individual, scalar elements of $\vec{x}$ as, say, x_k.

consider three types of characteristics. First consider the most standard case where each agent has a fixed type, a vector $\vec{x}_a^d$ for downstream firm a and a vector $\vec{x}_i^u$ for upstream firm i. For example, $\vec{x}_a^d$ could be the geographic location of a's assembly plant, information on the cars manufactured by a, the markets a sells to, etc. Likewise, $\vec{x}_i^u$ could be the geographic location of supplier i, the past experience of the supplier, etc. Allowing types to be vectors is an important extension of existing theoretical work. In this case, $\vec{x}(i,C^u) = \text{cat}\left(\vec{x}_i^u, x_{a_1}^d, \ldots, x_{a_n}^d\right)$, where $C^u = \{a_1, \ldots, a_n\}$.[10] X is the collection of characteristics for all potential matches in a market, whether the matches are part of the assignment A or not. Formally, $X = \left\{\vec{x}(i,C^u) \mid i \in U, C^u \subseteq D, |C^u| \leq q_i^u\right\}$.

I also consider cases where covariates vary directly at the match $\langle a,i\rangle$ or group-of-matches C^u levels. For match-specific characteristics, the long vector $\vec{x}(i,C^u) = \text{cat}\left(\vec{x}_{\langle a_1,i\rangle}^{u,d}, \ldots, \vec{x}_{\langle a_n,i\rangle}^{u,d}\right)$, where each $\vec{x}_{\langle a,i\rangle}^{u,d}$ is the vector of characteristics of the match $\langle a,i\rangle$. An example of an element of $\vec{x}_{\langle a,i\rangle}^{u,d}$ is a measure of whether two firms' inventory information systems are compatible. For group-specific characteristics, the vector $\vec{x}(i,C^u)$ is not a concatenation of shorter vectors for covariates that operate at the match $\langle a,i\rangle$ level. An example of group-specific characteristics is that $\vec{x}(i,C^u)$ might include the percentage of supplier i's assemblers that are located in countries with rigorous environmental regulations.

All upstream firms have the same revenue function, $r^u(\vec{x}(i,C^u))$, which gives the structural revenue of upstream firm i for the potential downstream firm partners in C^u. The potential nonlinearities in $r^u(\vec{x}(i,C^u))$ are key to this paper's empirical application to upstream-downstream markets. Downstream firm a has structural revenues $r^d(\vec{x}(i,\{a\}))$ from its potential match $\langle a,i\rangle$. I assume that the structural revenues of downstream firms are additively separable across multiple upstream firm partners: $r^d\left(\vec{x}\left(C^d,\{a\}\right)\right) = \sum_{i \in C^d} r^d(\vec{x}(i,\{a\}))$, where $C^d \subseteq U \cup \{0\}$ is a collection of upstream firms.

At a matching market outcome (A,T), the total profits of $i \in U$ are given by (2). The fact that transfers enter additively separably for both upstream and downstream firms allows us to focus on the following production function.

Definition 1. The production function for (i,C^u) for $i \in U$ and $C^u \subseteq D$ is

$$f(\vec{x}(i,C^u)) \equiv r^u(\vec{x}(i,C^u)) + \sum_{a \in C^u} r^d(\vec{x}(i,\{a\})).$$

For the marriage example of one-to-one matching with fixed types, $f(\vec{x}(i,\{a\})) \equiv r^u\left(\text{cat}\left(\vec{x}_i^u, \vec{x}_a^d\right)\right) + r^d\left(\text{cat}\left(\vec{x}_i^u, \vec{x}_a^d\right)\right)$.

For many-to-many matching, separability of $r^d(\vec{x}(i,\{a\}))$ across multiple upstream firm partners makes the set of arguments of f finite. Consider an example with matches $\langle a,i\rangle$, $\langle b,i\rangle$ and $\langle b,j\rangle$. If the model allowed arbitrary nonlinearities in both upstream and downstream firms' structural revenue functions, there would be a set of firms $\{a,b,i,j\}$ with production $f(\vec{x}(\langle\{i,j\},\{a,b\}\rangle))$, even though a and b, b and j as well as i and j have no direct links.[11] Requiring $r^d(\vec{x}(i,\{a\}))$ to be additively separable across multiple upstream firm matches makes the maximum number of matches described by $\vec{x}(i,C^u)$ equal to the maximum quota of any upstream firm, $\max\left\{q_i^u \mid i \in U, q_i^u \in Q\right\}$.

Sometimes I will view $f(\cdot)$ as an abstract function to be identified and estimated. In this case, I write $f(\vec{x})$, where the argument $\vec{x}$ is an arbitrary vector of characteristics.

[10] The concatenation operator makes one long vector out of a set of shorter vectors.

[11] In an empirical application, a researcher can choose a functional form for f so that nonlinearities in an upstream firm's profits across its downstream firm partners are distinguished from a downstream firm's nonlinearities across its upstream firm partners. Some additional assumption needs to be placed on production functions for nonparametric identification.

3.2 Pairwise stability

Because binding quotas prevent an agent from unilaterally adding a new partner without dropping an old one, the equilibrium concept in matching games allows an agent to consider exchanging a partner. I use the innocuous convention that suppliers pick assemblers.

Definition 2. A feasible outcome (A,T) is a pairwise stable equilibrium when:

1. For all $\langle a,i\rangle \in A$, $\langle b,j\rangle \in A$, $\langle b,i\rangle \notin A$, and $\langle a,j\rangle \notin A$,

$$r^u\left(\vec{x}(i,C_i^u(A))\right)+\sum_{c\in C_i^u(A)\setminus\{a\}} t_{\langle c,i\rangle}+t_{\langle a,i\rangle} \geq r^u\left(\vec{x}(i,(C_i^u(A)\setminus\{a\})\cup\{b\})\right)+\sum_{c\in C_i^u(A)\setminus\{a\}} t_{\langle c,i\rangle}+\tilde{t}_{\langle b,i\rangle}, \tag{3}$$

   where $\tilde{t}_{\langle b,i\rangle} \equiv r^d\left(\vec{x}\left(i,\left(C_i^u(A)\setminus\{a\}\right)\cup\{b\}\right)\right)-\left(r^d\left(\vec{x}\left(j,C_j^u(A)\right)\right)-t_{\langle b,j\rangle}\right)$.

2. For all $\langle a,i\rangle \in A$,

$$r^u\left(\vec{x}(i,C_i^u(A))\right)+\sum_{c\in C_i^u(A)\setminus\{a\}} t_{\langle c,i\rangle}+t_{\langle a,i\rangle} \geq r^u\left(\vec{x}(i,C_i^u(A\setminus\{a\}))\right)+\sum_{c\in C_i^u(A)\setminus\{a\}} t_{\langle c,i\rangle}.$$

3. For all $\langle a,i\rangle \in A$,

$$r^d\left(\vec{x}(i,C_i^u(A))\right)-t_{\langle a,i\rangle} \geq 0.$$

4. For all $\langle a,i\rangle \notin A$ where $\left|C_i^u(A)\right| < q_i^u$ and $\left|C_a^d(A)\right| < q_a^d$, there exists no $\tilde{t}_{\langle a,i\rangle} \in \mathbb{R}$ such that

$$r^u\left(\vec{x}(i,C_i^u(A))\right)+\sum_{c\in C_i^u(A)\setminus\{a\}} t_{\langle c,i\rangle} < r^u\left(\vec{x}(i,C_i^u(A)\cup\{a\})\right)+\sum_{c\in C_i^u(A)\setminus\{a\}} t_{\langle c,i\rangle}+\tilde{t}_{\langle a,i\rangle}$$

   and

$$r^d\left(\vec{x}(i,C_i^u(A)\cup\{a\})\right)-\tilde{t}_{\langle a,i\rangle} \geq 0.$$

Part 1 of the definition of pairwise stability says that upstream firm i would prefer its current assembler a instead of some alternative assembler b at the transfer $\tilde{t}_{\langle b,i\rangle}$ that would make assembler b switch to sourcing the part in question from i instead of its equilibrium supplier, j. Because of transferable utility, supplier i can always cut its price and attract b's business; at an equilibrium, it would lower its profit from doing so if the new business supplanted the relationship with a. This is the main component of the definition of pairwise stability that I will focus on in this paper.

Parts 2 and 3 deal with matched agents not profiting by unilaterally dropping a relationship and becoming unmatched. These are individual rationality conditions: all matches must give an incremental positive surplus. Finally, Part 4 involves two firms with free quota not wanting to form a new match. For the most part, I will not focus on these conditions in this particular paper because implementing them in identification and estimation would require more types of data. Parts 2–4 compare being matched to unmatched, and so implementing the restrictions from Parts 2–4 would require data on unmatched agents. A person being single or unmarried is often found in the data. The notion that a car parts supplier in an upstream–downstream market would have a free quota slot is a modeling abstraction. It is hard to find data on quotas.

I have not imposed sufficient conditions to ensure the existence of an equilibrium. In many-to-one, two-sided matching with complementarities across matches on the same side of the market, Hatfield and Milgrom

(2005), Pycia (2008) and Hatfield and Kojima (2008) demonstrate that preference profiles can be found for which there is no pairwise stable outcome.[12] The counterexamples mean that general existence theorems do not exist.[13]

Many interesting matching empirical applications require investigating possibilities outside of the scope of current existence theorems. I maintain the assumption that the data on an assignment represent part of an equilibrium for the game.[14] I discuss multiple equilibrium assignments below.

3.3 Using matches only: local production maximization

A matching game outcome (A,T) has two components: the assignment, sorting or matching A and the equilibrium transfers T. I consider using data on only A. This is because researchers often lack data on transfers, even when the agents use transfers. Car parts suppliers and automobile assemblers exchange money, but the transfer values are private, contractual details that are not released to researchers.

I will exploit the transferable utility structure of the game to derive an inequality that involves A but not T. Consider an example where $C_i^u(A) = \{a\}$ and $C_j^u(A) = \{b\}$. The inequality (3) becomes

$$r^u(\vec{x}(i,\{a\})) + t_{\langle a,i\rangle} \geq r^u(\vec{x}(i,\{b\})) + r^d(\vec{x}(i,\{b\})) - \left(r^d(j,\{b\}) - t_{\langle b,j\rangle}\right), \tag{4}$$

after substituting the definition of $\tilde{t}_{\langle b,i\rangle}$. Likewise, there is another inequality for upstream firm j's deviation to match with a instead of b:

$$r^u(\vec{x}(j,\{b\})) + t_{\langle b,j\rangle} \geq r^u(\vec{x}(j,\{a\})) + r^d(\vec{x}(j,\{a\})) - \left(r^d(i,\{a\}) - t_{\langle a,i\rangle}\right). \tag{5}$$

Adding (4) and (5), cancelling the transfers $t_{\langle a,i\rangle}$ and $t_{\langle b,j\rangle}$ that now are the same on both sides of the inequality, and substituting the definition of a production function, Definition 1, creates the new inequality

$$f(\vec{x}(i,\{a\})) + f(\vec{x}(j,\{b\})) \geq f(\vec{x}(i,\{b\})) + f(\vec{x}(j,\{a\})).$$

I call this a local production maximization inequality: "local" because only exchanges of one downstream firm per upstream firm are considered, and "production maximization" because the implication of pairwise stability says that the total output from two matches must exceed the output from two matches formed from an exchange of partners.

The local production maximization inequality suggests that interactions between the characteristics of agents in production functions drive the equilibrium pattern of sorting in a market. As the same set of firms

[12] Pycia (2007) has both existence and nonexistence results for matching markets without endogenous prices (Gale and Shapley, 1962).

[13] The fact that a pairwise stable equilibrium does not exist does not mean a decentralized matching market will unravel. Kovalenkov and Wooders (2003) and others study relaxed equilibrium concepts where it is easier to show existence, such as, for example, imposing a switching cost to deviate from the proposed assignment.

[14] In the non-nested-with-matching literature on estimating normal form Nash games, Ciliberto and Tamer (2007) throw out a particular realization of the error term's contribution to the likelihood if no pure-strategy equilibrium exists. Bajari, Hong and Ryan (2007b) compute all equilibria including mixed strategy equilibria, as a mixed strategy equilibrium is guaranteed to exist in a normal form Nash game. In matching, there is no notion of a mixed strategy equilibrium, as quotas are binding for every realization of the game. In a mixed strategy, players' actions are random, so a woman in a marriage market with quota 1 could find herself married to two men because of a random realization in a mixed strategy equilibrium.

More technically, Nash's existence theorem relies on a fixed-point argument requiring continuous strategies, like mixed strategies. Existence theorems in matching games rely on Tarski's fixed point theorem, which relies on monotonic operators and hence requires structure on preferences to ensure this monotonicity.

appears on both sides of the inequality, terms that do not involve interactions between the characteristics of firms difference out. In a one-to-one matching game, if $f(\vec{x}(i,\{a\})) = \beta_u' \vec{x}_i^u + \beta_d' \vec{x}_a^d$, then a local production maximization inequality is

$$\beta_u' \vec{x}_i^u + \beta_d' \vec{x}_a^d + \beta_u' \vec{x}_j^u + \beta_d' \vec{x}_b^d \geq \beta_u' \vec{x}_i^u + \beta_d' \vec{x}_b^d + \beta_u' \vec{x}_j^u + \beta_d' \vec{x}_a^d, \tag{6}$$

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.

For some policy questions, the cancellation of characteristics that are not interactions between the characteristics of multiple firms is an empirical advantage. Many datasets lack data on all important characteristics of firms. If some of these characteristics affect the production of all matches equally, the characteristics difference out and do not affect the assignment of upstream to downstream firms. If the policy questions of interest are not functions of these unobserved characteristics, then differencing them out leads to empirical robustness to missing data problems.[15]

More generally, the equilibrium concept of pairwise stability can be used to form a local production maximization inequality.

Lemma 1. Given a pairwise stable outcome (A,T), let $B_1 \subseteq A$, let π be a permutation of the downstream firm partners in B_1, and let

$$B_2 = \{\langle \pi \langle a,i \rangle, i \rangle \mid \langle a,i \rangle \in B_1\}.$$

Then the inequality

$$\sum_{\langle a,i \rangle \in B_1} f(\vec{x}(i, C_i^u(A))) \geq \sum_{\langle a,i \rangle \in B_2} f(\vec{x}(i, C_i^u((A \backslash B_1) \cup B_2))) \tag{7}$$

holds.

All proofs are found in the appendix.[16] The definition of a local production maximization inequality is similar to (7), except that no particular outcome (A,T) needs to be stated.

Definition 3. Let there be a set of matches B_1 and let B_2 be a permutation π of B_1, $B_2 = \{\langle \pi \langle a,i \rangle, i \rangle \mid \langle a,i \rangle \in B_1\}$. For each i where $\langle a,i \rangle \in B_1$, let there be a set of downstream firms C_i^u such that $\langle a,i \rangle \in B_1$ implies $a \in C_i^u$. Call

$$\sum_{\langle a,i \rangle \in B_1} f(\vec{x}(i, C_i^u)) \geq \sum_{\langle a,i \rangle \in B_1} f(\vec{x}(i, (C_i^u \backslash \{a\}) \cup \{\pi \langle a,i \rangle\}))$$

a local production maximization inequality.[17]

The local production maximization inequality lemma is both good and bad news. The good news is that the definition of pairwise stability is powerful: the condition that no upstream firm wants to swap a single downstream firm partner for a single new partner at the equilibrium transfers implies local production maximization

[15] 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, other than price, are independent of the unobserved characteristics. By contrast, Lemma 1 shows that transfers difference out. The advantage of differencing out further unobserved product characteristics involves a concern that the unobserved characteristics are correlated with the observed characteristics.

[16] A permutation π of the downstream firm partners applied to a set of matches $\{\langle a,i \rangle, \langle b,j \rangle, \langle c,k \rangle\}$ gives each upstream firm a new downstream firm partner. An example of a permutation is $\{\langle c,i \rangle, \langle a,j \rangle, \langle b,k \rangle\}$. For simplicity of notation, I let $\pi \langle a,i \rangle = c$ give the index of the new downstream firm partner c of the upstream firm i.

[17] The notation B_2 is not strictly speaking needed for Definition 3. Later I use B_1 and B_2 when showing that an inequality satisfies Definition 3.

inequalities involving large sets of downstream matches B_1 and corresponding permutations B_2. The potential large size of B_1 in the lemma will be important for some of the nonparametric identification theorems below.

The bad news is that it is computationally intensive to check that any assignment A satisfies the lemma's necessary conditions for pairwise stability without observing data on the equilibrium transfers $t_{\langle a,i \rangle}$.[18] In many-to-many matching games, checking the necessary conditions for pairwise stability means that (7) would have to be evaluated for all pairs (B_1, B_2) of a unique $B_1 \subseteq A$ and a valid permutation B_2.[19] For example, in the empirical application to automobile assemblers and car parts suppliers in this paper, there is one type of car component where suppliers compete to provide 1349 parts on individual cars. Given an outcome assignment A, there are

$$\sum_{k=1}^{1349} \binom{1349}{i} \cdot i! \approx 8.34 \times 10^{3638} \tag{8}$$

pairs (B_1, B_2) that distinguish unique inequalities (7) that would need to be checked.[20] Section 7 will propose an estimator that does not require itemizing all theoretically valid necessary conditions.[21]

4 Adding econometric error terms

4.1 Data on many independent matching markets

I will consider identification using data on the population of different matching markets. As before, a matching market is described by (D, U, Q, X, A, T). Like in the automotive assemblers and suppliers empirical example, data on the transfers T are often not available. Similarly, quotas, Q, are often an abstraction of the matching model and are not found in datasets like the one on automotive assemblers and suppliers. Therefore, I will explore identification using data on (D, U, X, A). From now on, I subsume D and U into X in order to use more concise notation. The researcher then observes (A, X) across markets.[22]

With data on the population of statistically independent and identically distributed as well as economically unrelated matching markets, the researcher is able to identify $\Pr(A \mid X)$, the probability of observing assignment A given that the market has characteristics $X = \{\vec{x}(i, C^u) \mid i \in U, C^u \subseteq D, |C^u| \leq q_i^u\}$. I use a probability mass

[18]I have no proof that satisfying (7) for all pairs (B_1, B_2) is a sufficient (as opposed to necessary) condition for A to be part of a pairwise stable equilibrium (A, T) in many-to-many matching games. Sotomayor (1999) implies that in a game where each firm's payoffs are additively separable across multiple matches, $f(\vec{x}(i, \{a,b\})) = f(\vec{x}(i, \{a\})) + f(\vec{x}(i, \{b\}))$, then checking all sets with two matches and their permutations, such as $B_1 = \{\langle a,i \rangle, \langle c,j \rangle\}$ and $B_2 = \{\langle a,j \rangle, \langle c,i \rangle\}$, should be enough. Additively separability across multiple matches rules out many interesting empirical applications, such as studying automotive supplier specialization in the empirical example in this paper. In the application, each supplier can supply many parts and there may be specialization gains to supplying more than one part to a downstream customer.

[19]If the pairwise stable assignment was unique, checking all necessary conditions might lead to a likelihood estimator, as used by Sørensen (2007) for a matching game where agents are not allowed to exchange money. This type of estimator will be infeasible in many-to-many matching games with transfers, when the transfers are not observed. If the transfers were observed by the researcher, the likelihood estimator would have to recognize the endogenous formation of the transfers as part of the equilibrium.

[20]This simple calculation overstates matters for the production function I use in the empirical work. If supplier i supplies two parts for the same car, the inequalities that switch those two car parts would not need to be checked. In this case, the set of matches for each supplier would appear in (8).

[21]Some inequalities will be implied by others. For example, adding two inequalities of the form (7) gives another valid inequality. However, there is not an underlying structure where checking all inequalities where B_1 contains two matches will imply all inequalities where B_1 contains three matches in general many-to-many matching games.

[22]For sake of brevity in the discussion of nonparametric identification, I assume the researcher has data on all elements of X. By adding additional notation, one could extend the nonparametric identification results to the case where some elements of X are missing. For example, all of the agents in the market may not be observed. See Fox (2007) for a related discussion on estimating the single-agent multinomial choice model without data on all available choices.

function $\Pr(A \mid X)$ because the observed outcome, a finite set of matches A, is the realization of a discrete random variable.

Why do we need $\Pr(A \mid X)$? Consider a hypothetical situation where the same suppliers sell both car seats and car tires. Each car has only one seat supplier and one tire supplier. Assume that the matching market for car seats in statistically and economically independent of the market for car tires. It could be that, even if the X variables are identical because the suppliers and assemblers are identical in both markets, for seats supplier 1 is the dominant firm and for tires supplier 2 is the dominant firm. If there were no error terms in the matching model and each market had a unique equilibrium assignment, then $\Pr(A \mid X) = 1$ for the A that is the pairwise stable equilibrium assignment, and $\Pr(A \mid X) = 0$ for any other A. The seats and tires dataset would statistically reject this deterministic matching model. For another example, if agents in a marriage market have the scalar type years of schooling and the schooling levels of men and women are complements in match production, then men and women will assortatively match in equilibrium. However, when a real dataset is examined, the sorting will not be so perfect. Even if most agents match assortatively, there will be some instances in the data where men with high schooling levels match with women with low schooling levels. If the model is not augmented with econometric error terms, the model will be inconsistent with this data and the model will be rejected by the data, in a statistical sense. Therefore, a properly specified econometric model will have to add error terms in order to rationalize the data.

To ensure that the model gives full support to the data, I wish that $\Pr(A \mid X) > 0$ for any physically feasible (matches of each agent under that agent's quota) assignment A.[23] The probability $\Pr(A \mid X)$ will be induced by a stochastic structure S. In a model with match-$\langle a,i \rangle$-specific error terms, $S \in \mathscr{S}$ will represent the distribution of the error terms. Then

$$\Pr(A \mid X) \equiv \Pr\left(A \mid X; f^0, S^0\right) \equiv E_{Q|X}\left[\Pr\left(A \mid X, Q; f^0, S^0\right)\right],$$

where $\Pr\left(A \mid X, Q; f^0, S^0\right)$ is the probability of an assignment A being observed given the exogenous characteristics X, the exogenous quotas Q, the true match production function f^0, and the true distribution of the error terms S^0. The functions f^0 and S^0 are unknown to the econometrician and are arguments to the endogenous variable data generating process $\Pr\left(A \mid X, Q; f^0, S^0\right)$, but they are fixed across markets and are not random variables. The matching model and any equilibrium assignment selection rule together induce the distribution $\Pr\left(A \mid X, Q; f^0, S^0\right)$. I will discuss primitive formulations of error terms in detail below. The quotas in Q are unmeasured, so the econometrician observes data on $\Pr\left(A \mid X; f^0, S^0\right) \equiv E_{Q|X}\left[\Pr\left(A \mid X, Q; f^0, S^0\right)\right]$, where the expectation over Q is taken with respect to its distribution conditional on X.[24]

4.2 The rank order property

I wish to identify both cardinal and ordinal features of f^0, the match production function that generates the data. I will rely on some ideas from the maximum score literature to add econometric randomness to the matching outcomes.[25] Identification of single-agent discrete choice models in the maximum score literature relies on an

[23] This focus on allowing errors to affect the realization of A distinguishes this paper's approach to matching games from the work on estimating Nash games by Pakes, Porter, Ho and Ishii (2006), which does not allow for these errors in general normal form Nash games and so, except in a few cases such as ordered choice, does imply the analog to $\Pr(A \mid X) = 0$ for some physically possible A's.

[24] The transfers T do not need to be integrated out because T is a separate endogenous outcome from A.

[25] Maximum score is a partial identification approach. In single-agent discrete choice problems with the payoff structure $x'_{a,i}\beta + \varepsilon_{a,i}$ for agent i and choice a, S represents the distribution of error terms $\varepsilon_{a,i}$. S is not identified under the standard (rank order property) conditions

intermediate (non-primitive) property that I call the rank-order property. For example, Manski (1975) studies the identification of both binary and multinomial choice models using a rank-order property. Different sufficient conditions on S, in single-agent discrete choice the distribution of choice-specific error terms, imply the rank order property for binary and for multinomial choice.[26] Sufficient conditions for the rank order property in matching models require more exposition. Therefore, I first describe a non-primitive rank order property for matching games and discuss how it relates to the deterministic analysis of Becker (1973). Later I will discuss sufficient primitive conditions.

First consider one-to-one matching. Say each agent has a one-dimensional type, schooling. The econometrician observes $\Pr(A \mid X) = \Pr\left(A \mid X; f^0, S^0\right)$, which here is the likelihood of a certain sorting pattern given a distribution of schooling levels among men and among women. Say also that the schooling levels of men and women are everywhere complements in production. Then men and women would assortatively match in a deterministic model. In a stochastic econometric model, the rank order property will in effect say, for assignments A_1 and A_2, $\Pr(A_1 \mid X) > \Pr(A_2 \mid X)$ whenever A_2 can be formed from A_1 by having less assortative matching: some couples with higher schooling levels will exchange partners with couples that have lower schooling levels. The intuition for why $\Pr(A_1 \mid X) > \Pr(A_2 \mid X)$ traces back to production functions: the sum of production under A_1 will exceed that of A_2 if schooling levels are complements. Even though A_1 and A_2 can both occur with positive probability because of econometric unobservables (unlike in Becker), here Becker's idea that assortative matching occurs when types are complements can be made stochastic by rank ordering the probabilities of assignments by how much assortative matching occurs.

The general model allows many-to-many matching, where pairwise stability does not give a link to economy-wide production ("efficiency"), as well as a general space of observable characteristics X, not just scalar types. The rank order property is stated as an assumption and can be seen as a stochastic version of local production maximization.

Assumption 1. Let A_1 be a feasible assignment for a market with characteristics X. Let $B_1 \subseteq A_1$ and let π be a permutation of the downstream firm partners in B_1, giving

$$B_2 = \{\langle \pi \langle a,i \rangle, i \rangle \mid \langle a,i \rangle \in B_1\}.$$

Let $A_2 = (A_1 \backslash B_1) \cup B_2$. Let $S \in \mathscr{S}$ be any distribution of the error terms and let $f \in \mathscr{F}$ be any production function. Assume that

$$\sum_{\langle a,i \rangle \in B_1} f\left(\vec{x}\left(i, C_i^u(A_1)\right)\right) > \sum_{\langle a,i \rangle \in B_2} f\left(\vec{x}\left(i, C_i^u(A_2)\right)\right) \tag{9}$$

if and only if

$$\Pr(A_1 \mid X; f, S) > \Pr(A_2 \mid X; f, S).$$

Keep in mind that X, f and S are held fixed: the rank-order property is an assumption about the stochastic structure of the model.

To understand the rank order property, consider a situation where A_1 contains thousands of matches and $B_1 = \{\langle a,i \rangle, \langle b,j \rangle\}$ contains only two matches. Then $A_2 = (A_1 \backslash B_1) \cup B_2$ is equal to A_1 except that the matches

for the identification of β in either binary or multinomial choice problems (Manski, 1975, 1988). In matching, I will focus on identifying f and not S.

[26]For example, the binary choice model is identified under median independence and the multinomial choice model is identified under exchangeability of the joint distribution of the choice-specific errors (Fox, 2007).

$B_2 = \{\langle a,j\rangle, \langle b,i\rangle\}$ form. Given X and Q, neither A_1 or A_2 may be a stable assignment to the matching model without error terms. But A_1 might dominate A_2 in the deterministic model in that at least two agents in B_2 would prefer to match with each other instead of their assigned partners, leading to A_1. More generally, if the local production maximization inequality (9) is satisfied, then some agents in B_1 want to deviate in the deterministic matching model. In a model with error terms, both A_1 and A_2 could be pairwise stable assignments to some realizations of the unobserved components in the matching model. The assumption says that A_1 will be more likely to be a pairwise stable assignment to some realized model than A_2.

The assumption is that among two very similar assignments, A_1 and A_2, which differ in at most one downstream firm match per upstream firm, the assignment that cannot be pairwise stable in the deterministic game is less likely to occur. This is certainly a natural assumption that a researcher could use when interpreting descriptive empirical results on sorting patterns through the lens of matching theory.

As the quotas in Q are not observed in many empirical applications, a slightly more primitive version of Assumption 1 is that (9) holds if and only if $\Pr(A_1 \mid X, Q; f, S) > \Pr(A_2 \mid X, Q; f, S)$, for any valid Q. Then taking expectations with respect to $Q \mid X$ gives Assumption 1. Even if Q is unobserved, for the most part I have only considered inequalities where the total number of matches of each agent in A_1 and A_2 is kept the same.[27] If unmatched agents are not considered in B_1 and B_2 and if A_1 is a feasible assignment for Q, A_2 is also a feasible assignment for that Q.

4.3 Sufficient condition for the rank order property

This subsection explores a sufficient condition for the rank order property, Assumption 1, in the context of models where assignments have unobserved components in production. In this subsection only, I assume that the outcome (A,T) is in the core. The core is an equilibrium concept. A core outcome is robust to deviations by any group of firms. If the group of all firms cannot improve its joint payoff, a core assignment must maximize the sum of production for the entire matching economy. Consequently, the decentralized matching market assignment can be restated as a social planning problem.[28]

There is a finite, although potentially large, number of assignments. The social planning problem is a single-agent, unordered, discrete choice problem of Manski (1975), where the single agent is the social planner. From Manski's work, we know the sufficient condition that will arise. For an assignment A, let its total production be $\sum_{\langle a,i\rangle \in A} f\left(\vec{x}\left(i, C_i^u(A)\right)\right) + \psi_A$, where ψ_A is an unobserved component of the production of assignment A. Let ψ be the vector of all ψ_A's. Let ψ have the density S and let ψ be independent of Q and X.[29] Let $\Pr(A \mid Q, X; f, S)$ be the probability A is the core assignment.

Lemma 2. Let the equilibrium concept be the core and let the density S conditional on X exist, have full support and be exchangeable in the elements of ψ. Then the rank order property, Assumption 1, holds.

[27]Unmatched agents 0 could be included in matches in B_1. For nonparametric identification, data on unmatched agents will not be needed, except for one theorem.

[28]The necessary conditions from pairwise stability are enough to nonparametrically identify production functions. Extra inequalities from the stronger equilibrium concept the core will not be required. However, the social planning property ensures a unique assignment with probability 1. If some observables or unobservables in production have a continuous distribution, then the probability that any two assignments both solve the social planning problem is 0. To eliminate the role of the equilibrium assignment selection rule, this section considers only games where the outcome is in the core.

[29]More generally, there is no need for the density S to be the same for all markets X. See Fox (2007) for more discussion of letting X be a conditioning argument in S, for the single agent, multinomial choice case.

This lemma was proved in Goeree, Holt and Palfrey (2005) and is a slight generalization of a result in Manski (1975). See the appendix for another proof.[30]

The social planner errors can be interpreted as errors in the deterministic model from finding the true core solution. One could then view exchangeability of the joint density as a structural assumption on the equilibrium assignment selection process. Adding errors to a deterministic model is similar to the quantile response equilibrium method of perturbing behavior (Goeree et al.). The social planning problem is a structural assumption that does exactly generalize the intuition from the empirical matching literature (without error terms) that assignments with, say, more assortative matching are more likely to occur.[31]

4.4 Match-specific error terms

In the literature on estimating perfect-information Nash games Andrews, Berry and Jia (2004); Bajari, Hong and Ryan (2007b); Beresteanu, Molchanov and Molinari (2008); Berry (1992); Bresnahan and Reiss (1991); Ciliberto and Tamer (2007); Jia (2008); Mazzeo (2002); Tamer (2003), a typical assumption is that the payoff to an agent a from taking action i is $r\left(\vec{x}_{a,i}\right)+\varepsilon_{a,i}$, where $\vec{x}_{a,i}$ are observables covariates and $\varepsilon_{a,i}$ is a player-a-and-action-i-specific shock. This payoff structure is borrowed from the literature on single-agent discrete choice models. In a perfect information Nash game, the $\varepsilon_{a,i}$'s are assumed to be common knowledge for all players; only the econometrician lacks data on them. These errors are added to the model to be able to fit any observed equilibrium.[32]

The analog of the practice for perfect information Nash games in matching games (a non-nested class of games) is match-specific error terms $\varepsilon_{\langle a,i\rangle}$. The equilibrium assignment to a transferable utility matching game whose outcome is in the core of the game can be computed with a linear programming problem.[33] Let the total output of a set of downstream firm partners C^u for upstream firm i be

$$f\left(\vec{x}\left(i,C^u\right)\right)+\sum_{a\in C^u}\varepsilon_{\langle a,i\rangle}, \tag{10}$$

where $\varepsilon_{\langle a,i\rangle}$ is the match-$\varepsilon_{\langle a,i\rangle}$-specific error term, which is independent of all components of X and Q. Let the

[30] An exchangeable joint density satisfies $g\left(y_1,y_2,\ldots,y_n\right)=g\left(\pi y_1,\pi y_2,\ldots,\pi y_n\right)$ for any permutation π of any vector of arguments $\left(y_1,\ldots,y_n\right)$.

[31] An exchangeable joint density for assignment level errors is a sufficient but not necessary condition for the rank order property. Consider the comparison of an assignment $A_1=\{\langle 1,1\rangle,\langle 2,2\rangle,\langle 3,0\rangle,\langle 0,3\rangle\}$, where downstream and upstream firms 3 are both unmatched, to another assignment $A_2=\{\langle 1,2\rangle,\langle 2,3\rangle,\langle 3,1\rangle\}$ where all firms are matched. The local production maximization inequality in Assumption 1 does not allow comparing A_1 and A_2 because A_2 is not a rearrangement of downstream firm partners from A_1: A_2 does not reallocate the former states of being unmatched. Therefore, the comparison of A_1 and A_2 is not relevant for the rank order property. However, the social planner with exchangeable errors model makes predictions about the relative frequencies of A_1 and A_2 based on their relative sums of production. For example, Lemma 2 says $\sum_{\langle a,i\rangle\in A_1} f\left(\vec{x}\left(i,C_i^u(A)\right)\right)=\sum_{\langle a,i\rangle\in A_2} f\left(\vec{x}\left(i,C_i^u(A)\right)\right)$ if and only if $\Pr\left(A_1\mid X;f,S\right)=\Pr\left(A_2\mid X;f,S\right)$. This is despite the fact that in A_1 match $\langle 3,3\rangle$ will often form if it gives more production than being unmatched, so, in many realizations of uncertainty, A_1 will not be be a stable assignment. Therefore, the rank order property is a weaker assumption than a social planner with exchangeable assignment-level errors.

[32] Two other sets of assumptions have some drawbacks. First, in static, private information Nash games, agents do not know the error terms $\varepsilon_{a,i}$ of rival players and hence the equilibrium actions of rivals. Therefore, there could be ex-post regret: a firm entered the market predicting it would be a monopolist and found instead that it was one of several firms, a situation where it is not profitable. (In matching games, because of the physical property that the number of matches of each agent must be under its quota, not knowing whether your partner will agree to match with you when you take your action makes it impossible to ensure that the resulting equilibrium is pairwise stable. See Section 11.1 for more discussion.) Second, the assumption that $\varepsilon_{a,i}$ is pure measurement error or a shock to profits after matches are formed (expectational error) implies that the equilibrium assignment A is a deterministic function of X. Therefore, it will be easy to reject the model as it will assign probability 0 to many A's that may be found in the data.

[33] Pairwise stability is equivalent to the core for one-to-one matching games and many-to-many matching games where payoffs are additively separable across multiple matches. The core is a stronger solution concept for many-to-many matching games where payoffs are not additively separable across multiple matches.

stochastic structure S represent the distribution of $\varepsilon_{\langle a,i\rangle}$. Then, in the perfect information world where $\varepsilon_{\langle a,i\rangle}$ is observed by all agents in the model but is not in the data,

$$\Pr(A \mid X, Q; f, S) = \int_{\varepsilon} 1\left[A \text{ solves assignment linear program} \mid X, Q, \varepsilon\right] dS(\varepsilon), \tag{11}$$

where ε is the vector of error terms for all $U \cdot D$ possible matches as well as the option of being single for each agent. Under this model, S can be chosen so that each physically feasible A will always have positive probability.

Unfortunately, the integrand in (11) includes a linear program and the resulting $\Pr(A \mid X, Q; f, S)$ does not have the nice properties for the single-agent multinomial choice model that Manski (1975) found when, say, $\varepsilon_{a,i}$ was iid across choices for the same agent. For matching, unlike single-agent discrete choice, it is not a theorem that iid errors yield the rank order property for matching, Assumption 1. If the rank order property for matching is a natural, stochastic generalization of the core property found by Becker (1973) and others, then iid errors is not a primitive condition for the stochastic structure $\mathscr{S}$ that gives this natural generalization, exactly.

All models are approximations to reality. If the true production function is thought to include iid match-specific shocks as in (10) and therefore assignment probabilities are given by (11), then the rank order property may actually be a pretty close approximation. After all, the transferable utility and price-taking structure of the game does naturally imply that adding production functions is much more natural than in a noncooperative Nash game. I now present simulation results that examine how closely a perfect-information matching game with shocks as in (10) is approximated by the rank order property, a natural generalization of prior work on matching games without econometric errors.

Table 1 includes results from simulations that compute assignment probabilities for a one-to-one, two-sided matching game where match production is $f(\vec{x}(i,\{a\})) + \varepsilon_{\langle a,i\rangle}$.[34] The matching game has three upstream firms and three downstream firms. The deterministic payoffs of the game are chosen so that $\sum_{\langle a,i\rangle \in A_1} f(\vec{x}(i,\{a\})) = \sum_{\langle a,i\rangle \in A_2} f(\vec{x}(i,\{a\}))$ for two assignments, A_1 and A_2. As A_1 and A_2 differ by rearrangements of one downstream firm per upstream firm, the rank order property, Assumption 1, requires that $\Pr(A_1 \mid X; f, S) = \Pr(A_2 \mid X; f, S)$. Neither A_1 or A_2 is a deterministic stable assignment.[35] The deterministic payoffs $\sum_{\langle a,i\rangle \in A_1} f(\vec{x}(i,\{a\}))$ are constructed so that deviation by agents in A_2 is more attractive in an ease metric (two matched pairs could exchange partners, leaving the third pair alone) to provide a more compelling test against the idea that the rank order property is satisfied. The details of the game are in the notes to Table 1.

Table 1 considers six distributions S for iid match-specific unobservables. The table uses a simulation of the integral in (11) to compute $\Pr(A_1 \mid X; f, S) - \Pr(A_2 \mid X; f, S)$, a measure of how far off the rank order property is. The first line considers a standard normal distribution. As the variance is small and both A_1 and A_2 are not stable assignments in the deterministic game, the assignment probabilities are individually small. However, the difference $\Pr(A_1 \mid X; f, S) - \Pr(A_2 \mid X; f, S) = -0.01514$ is large relative to the magnitudes. The second line increases the normal standard deviation to 6. Both assignment probabilities increase to around 0.079, but the difference $\Pr(A_1 \mid X; f, S) - \Pr(A_2 \mid X; f, S)$ decreases in absolute value to 0.00009. The third line increases the standard deviation to 20; now the probabilities are around 0.065, although $\Pr(A_1 \mid X; f, S) - \Pr(A_2 \mid X; f, S)$ remains small, at 0.00011.

I also investigate to what degree the previous simulations relied on normality. The final three experiments

[34] The game is one-to-one matching, so Q is observed and the same for all markets.

[35] In other words, neither A_1 or A_2 would solve the linear programming assignment problem if all errors $\varepsilon_{\langle a,i\rangle}$ were 0.

Table 1: Assignment probabilities for two assignments with equal deterministic production with iid match-specific unobservables, by distribution

# Firms $U=D$	Distribution for iid errors, S	Error standard deviation	$\Pr(A_1 \mid X; f,S)$	$\Pr(A_2 \mid X; f,S)$	$\Pr(A_1 \mid X; f,S) -$ $\Pr(A_2 \mid X; f,S)$
3	$N(0,1)$	1	0.02126	0.03640	-0.01514
3	$N(0,36)$	6	0.07897	0.07888	0.00009
3	$N(0,400)$	20	0.06554	0.06543	0.00011
3	$0.33 \cdot N(0.5,0.04) + 0.67 \cdot N(-0.5,0.123)$	0.53	0.000098	0.001791	-0.00163
3	$0.33 \cdot N(2.5,0.04) + 0.67 \cdot N(-2.5,0.123)$	2.44	0.047894	0.045908	0.001986
3	$0.33 \cdot N(8.0,4.0) + 0.67 \cdot N(-6.0,6.25)$	6.98	0.033811	0.033967	-0.000156

The rank order property says $\Pr\left(A_1 \mid X; f^0, S\right) - \Pr\left(A_2 \mid X; f^0, S\right) = 0$ for any S. Total match production is $f(X(i,\{a\})) + \varepsilon_{\langle a,i\rangle}$, with the error's distribution given in the table. The assignment is calculated using linear programming (Roth and Sotomayor, 1990), as in (11). Each integral is simulated by using 1 million draws of the realizations for the collection of error terms for all matches and being single. Given the number of replications, the differences in the table probably do not reflect simulation error.

There are three upstream firms and three downstream firms in a one-to-one, two-sided matching game. The production of being unmatched is 0. The deterministic match production levels for matching with the three downstream firms are $\{3,1,2.8\}$ for upstream firm 1, $\{1,2.8,1\}$ for upstream firm 2, and $\{3,1,1\}$ for upstream 3. I compute the probabilities for the assignments $A_1 = \{\langle 1,2\rangle, \langle 2,3\rangle, \langle 3,1\rangle\}$, with production $1+1+3=5$, and $A_2 = \{\langle 1,1\rangle, \langle 2,3\rangle, \langle 3,2\rangle\}$, with production $3+1+1=5$. I chose the example so that assignment A_2 will be "more vulnerable" to a deviation to an assignment $A_3 = \{\langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle\}$ with deterministic production $3+2.8+2.8=8.6$, as only two matched pairs in A_2, rather than all three pairs in A_1, need to exchange partners to deviate to A_3.

in Table 1 consider asymmetric mixed normal distributions with two modes. Again, it appears that the absolute value of $\Pr(A_1 \mid X; f,S) - \Pr(A_2 \mid X; f,S)$ is smaller when the standard deviation of the errors is higher.

Table 1 implies that assignment probabilities that differ by exchanges of only one downstream firm for each upstream firm are nearly rank-ordered by their deterministic payoffs, when true payoffs include match-specific stochastic components. Based on these simulation results, I conjecture that maximum score will have a relatively small bias if the stochastic structure of the model involves iid match-specific unobservables in production. I will explore this conjecture with a Monte Carlo study later, in Section 8.

4.5 Multiple equilibrium assignments

Return to using pairwise stability as the only equilibrium concept. In a one-to-one matching game or a many-to-many matching game with additive separability across multiple matches (Sotomayor, 1999), generically there will be only one equilibrium assignment.[36] In more complex many-to-many matching games, multiple assignments may be pairwise stable. In a game with multiple equilibrium assignments, (11) becomes

$$\Pr(A \mid Q, X; f, S) = \int_{\varepsilon} 1[A \, \text{selected assignment} \mid A \, \text{stable}, X, Q, \varepsilon] \cdot 1[A \, \text{pairwise stable} \mid X, Q, \varepsilon] \, dS(\varepsilon). \qquad (12)$$

Let $\Upsilon(A \mid Q, X; f, S)$ be equal to $\int_{\varepsilon} 1[A \, \text{pairwise stable} \mid Q, X, \varepsilon] \, dS(\varepsilon)$. Define A_1 and A_2 as in Assumption 1. For a model with multiple equilibrium assignments, the rank order property, Assumption 1, will hold under the following conditions: 1) $\Upsilon(A_1 \mid Q, X; f, S) > \Upsilon(A_2 \mid Q, X; f, S)$ if and only if inequality (7) holds, and 2) $\Pr(A_1 \mid Q, X; f, S) >$

[36] Again, if the equilibrium is in the core of the matching game, the equilibrium assignment is the solution to a linear programming program. If the total payoff to a match $f(\vec{x}(i, C^u)) + \sum_{a \in C^u} \varepsilon_{\langle a,i\rangle}$ has a continuous distribution because of continuous elements of $\vec{x}(i, C^u)$ or $\varepsilon_{\langle a,i\rangle}$, then the event that two or more assignments solve the linear programming problem has probability zero. Here the probability is taken across markets.

$\Pr(A_2 \mid Q,X;f,S)$ if and only if $\Upsilon(A_1 \mid Q,X;f,S) > \Upsilon(A_2 \mid Q,X;f,S)$. Part 1 says A_1 will more likely to be stable than A_2 if A_1 has a higher production after an exchange of one downstream firm for each upstream firm in some set $B_1 \subseteq A_1$. Part 2 says the equilibrium assignment selection rule preserves the rank ordering of stability: assignments that are more likely to be stable are more likely to occur. These conditions together imply Assumption 1 and hence allow a unified framework to be used to study identification and estimation of matching games, regardless of the number of stable assignments for each ε and X combination. Assumptions about an equilibrium assignment selection rule may be strong, but for now they are currently the only feasible alternative for matching games with large numbers of agents and multiple equilibrium assignments. Under this assumption, the resulting maximum score estimator will be consistent without any change to deal with multiple equilibrium assignments.

The literature on Nash games, a non-nested class with matching games, presents some alternative strategies to dealing with multiple equilibria. For estimation, the method of Bajari, Hong and Ryan (2007b) and the lower bound to the likelihood of Ciliberto and Tamer (2007) require nesting calls to computer software that compute all equilibria to a game for each realization of error terms.[37] As the combinatorics in (8) state, even checking whether a single assignment satisfies all the known necessary conditions for a stable assignment is computationally infeasible in the empirical application in this paper. Further, the statistical objective function in Ciliberto and Tamer requires a first stage, nonparametic estimate of $\Pr(A \mid X)$ for all market characteristics X and assignments A observed in the data.[38] I will argue below that such a first-stage nonparametric estimator will suffer from a data curse of dimensionality because A and X are usually of very high dimension in matching games. Maximum score will eliminate this first-stage, nonparametric estimation and hence the data curse of dimensionality.

4.6 Do search and switching costs give the rank order property?

Search is another model for the stochastic structure $\mathscr{S}$ of a matching game. The perfect information benchmark is perfect assortative matching if scalar inputs are complements in production for one-to-one matches. The complements result can be embedded in a search model with forward-looking agents. Shimer and Smith (2000) show that complementarities in the production function plus complementarities in the first and cross-partial derivatives are sufficient for assortative matching in a search model with one-to-one matching and scalar agent types where costs are driven by time discounting. Atakan (2006) shows that complementarities in the production function alone drive assortative matching in a model with additive search costs.

Atakan (2007) shows that there exists a constrained (by the search technology) efficient equilibrium for general production functions. I conjecture that some variant of the maximum score estimator in this paper will be consistent in the constrained efficient world of Atakan (2007).

Kovalenkov and Wooders (2003) worry about the nonexistence of equilibria as defined by standard concepts and therefore discuss the ε-core solution concept. In an ε-core equilibrium, a coalition $B \subseteq U \cup D$ 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. The introduction of switching costs implies that there may be multiple equilibrium

[37] Andrews, Berry and Jia (2004) and Beresteanu, Molchanov and Molinari (2008) study the same model as Ciliberto and Tamer (2007).

[38] In private communications, Tamer suggests that a GMM estimator with moment inequalities in the observed and predicted game outcomes could be used to avoid this first stage. This approach would still require simulating the probability that an outcome is an equilibrium. The maximum score estimator that I will propose avoids simulating equilibrium probabilities by working with rank orders of probabilities rather than probabilities themselves.

assignments. Therefore, the rank order property will be based on the equilibrium assignment selection rule in (12).

5 Cardinal nonparametric identification

No previous paper has studied the nonparametric identification of matching games. It is not obvious what types of parameters one can learn about using data on equilibrium matches but not the choice set available to each agent in equilibrium. Because firms on the same side of the market are rivals to match with potential partners, the failure for a match to form does not mean that the match gives low production. In this section, I generalize the identification results in Becker (1973), described in Section 2.1, to the case of many-to-many matching with vectors of agent characteristics, among other extensions. As (1) shows, in Becker's scalar types world one can only identify whether the inputs of men and women are complements or substitutes in the production function. Likewise, in other settings the full cardinality of f cannot be identified from data on assignments alone.

As mentioned before, I will explore identification with market-level data on (A,X). Let (A,X) be statistically independent across matching markets. With this data, I can identify both $\Pr(A \mid X)$ and $G(X)$, the distribution of X across markets.

5.1 Cardinal identification preliminaries

I use an extension of a standard definition for point identification by Gourieroux and Monfort (1995, Section 3.4).

Definition 4. Let $\mathscr{F}$ be a class of production functions. Let $f^0 \in \mathscr{F}$ be the production function and let $S^0 \in \mathscr{S}$ be the stochastic structure in the data generating process.

- f^0 is identified within the class of production functions $\mathscr{F}$ if there does not exist $f^1 \neq f^0 \in \mathscr{F}$, stochastic structure $S^1 \in \mathscr{S}$, and some possibly empty set $\mathscr{Y}$ of market characteristics of probability 0 such that $\Pr\left(A \mid X; f^1, S^1\right) = \Pr\left(A \mid X; f^0, S^0\right)$ for all (A,X) with $X \notin \mathscr{Y}$.

- Let $c(\cdot)$ be a known function that produces either a scalar, vector, another function of $\vec{x}$ or a vector of functions of $\vec{x}$. A feature of f^0 $c\left(f^0\right)$ is identified within the class of production functions $\mathscr{F}$ if there does not exist $f^1 \in \mathscr{F}$ where $c\left(f^1\right) \neq c\left(f^0\right)$, stochastic structure $S^1 \in \mathscr{S}$, and some possibly empty set $\mathscr{Y}$ of market characteristics of probability 0 such that $\Pr\left(A \mid X; f^1, S^1\right) = \Pr\left(A \mid X; f^0, S^0\right)$ for all (A,X) with $X \notin \mathscr{Y}$.

The probability of $\mathscr{Y}$ is $\int_{\mathscr{Y}} dG(X)$. I maintain the following assumption for the discussion of cardinal identification.

Assumption 2.

1. Each $f \in \mathscr{F}$ is three times differentiable in all of its arguments.

2. The collection of market characteristics X has support equal to the product of the supports of each $\vec{x}(i,C^u)$ in X. Each $\vec{x}(i,C^u)$ has support on an open rectangle on $\mathbb{R}^K$, where $K = |\vec{x}(i,C^u)|$.

I make this assumption to focus on cross-partial derivatives, like Becker. These conditions can be relaxed.[39]

The features of the production functions that govern sorting depend on how the characteristics that enter $\vec{x}(i, C^u)$ vary. I will present results where characteristics vary at the levels of the individual firm i, the individual match $\langle a, i \rangle$, and the group C^u of downstream firms matching with one upstream firm. Keep in mind that a unit of observation is a market. When I informally speak of $\vec{x}(i, C^u)$ varying, what I mean is that there are two markets with very similar characteristics X, except for the values of $\vec{x}(i, C^u)$ for some assembler i and set of parts suppliers C^u.[40]

5.2 Cardinal identification theorems

5.2.1 Cardinal identification with firm-specific characteristics

First I consider firm-specific characteristics. Let there be K^u characteristics for each upstream firm and K^d characteristics for each downstream firm. In this case, the vector

$$\vec{x}(i, C^u) = \text{cat}\left(\left(x^u_{i,1}, \ldots, x^u_{i,K^u}\right), \left(x^d_{a_1,1}, \ldots, x^d_{a_1,K^d}\right), \ldots, \left(x^d_{a_n,1}, \ldots, x^d_{a_n,K^d}\right)\right),$$

where $C^u = \{a_1, \ldots, a_n\}$ is a finite set of n downstream firms and where $x^d_{a,e}$ is the eth out of the K^d characteristics of downstream firm a. Recall that $\vec{x}$ is an arbitrary characteristic vector of the form $\vec{x}(i, C^u)$.

Theorem 1. *Let* $\vec{x}(i, C^u) = \text{cat}\left(\left(x^u_{i,1}, \ldots, x^u_{i,K^u}\right), \left(x^d_{a_1,1}, \ldots, x^d_{a_1,K^d}\right), \ldots, \left(x^d_{a_n,1}, \ldots, x^d_{a_n,K^d}\right)\right)$*. Let* $\vec{x}$ *be a given point of evaluation of* f*.*

1. *Let* x_1 *and* x_2 *be scalar characteristics in* $\vec{x}$ *from two different firms, either one upstream firm and one downstream firm or two downstream firms. Assume* $\frac{\partial f^0(\vec{x})}{\partial x_1 \partial x_2} \neq 0$*. Then the sign of* $\frac{\partial f^0(\vec{x})}{\partial x_1 \partial x_2}$ *is identified.*
2. *Let* x_1 *and* x_2 *be scalar characteristics in* $\vec{x}$ *from two different firms, and let* x_3 *and* x_4 *be two scalar characteristics from two different firms as well. The identities of the firms in the two pairs* (x_1, x_2) *and* (x_3, x_4) *can be the same or not. Assume* $\frac{\partial f^0(\vec{x})}{\partial x_1 \partial x_2} \neq 0$ *and* $\frac{\partial f^0(\vec{x})}{\partial x_3 \partial x_4} \neq 0$*. Then the ratio* $\frac{\partial f^0(\vec{x})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{x})}{\partial x_3 \partial x_4}$ *is identified.*

Part 1 shows that Becker's result for the case of scalar types for men and women applies to each pair of scalar characteristics for distinct firms. Part 1 also extends Becker to a local notion of identification: f's inputs can be complements at some areas of support and substitutes at other areas. Note that Theorem 1 does not allow a researcher to tell whether a pair of inputs are "more" complementary at some $\vec{x}_a$ than some other point $\vec{x}_b$. For example, $\frac{\partial f(\vec{x}_a)}{\partial x_1 \partial x_2} = 5$ and $\frac{\partial f(\vec{x}_b)}{\partial x_1 \partial x_2} = 7$ cannot be distinguished from any other pair of positive values.

Becker studies only scalar types. Part 2 of the theorem shows the econometrician can go further and identify the relative importance of the complementarities for two pairs of characteristics. This is perhaps the most important result on identification in this paper. The econometrician can run a "horse race" where he or she tries to measure the relative importance of several pairs of characteristics. This is not just a ordinal

[39]While there are definitions such as increasing differences (Milgrom and Shannon, 1994) that encompass complementarities without relying on differentiable f's and continuous support for the x's, working with broader definitions makes the results harder to interpret and to compare to Becker's.

[40]The proofs of the identification theorems use data on different X's in part to satisfy the technical requirement in Definition 4 that f^0 and any alternative f^1 make different predictions for a set of X's with positive probability.

comparison, as in $\frac{\partial f^0(\vec{x})}{\partial x_1 \partial x_2} > \frac{\partial f^0(\vec{x})}{\partial x_1 \partial x_3}$. Rather, the cardinal ratio of the degree of complementarities, $\frac{\partial f^0(\vec{x})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{x})}{\partial x_3 \partial x_4}$, can be identified from qualitative data on who matches with whom.[41]

This proof of this theorem is the first result on the nonparametric identification of matching games and the most novel piece of original mathematics in the paper. The intuition behind the proof of Part 1 is that a market with characteristics X can be found where either assortative or anti-assortative matching on x_1 and x_2 is possible. Which occurs determines the sign of $\frac{\partial f^0(\vec{x})}{\partial x_1 \partial x_2}$. A technical complexity handled in the proof is that Definition 4 requires a continuum of such market characteristics X. The intuition for the proof of Part 2 is that the relative magnitudes of the cross-partials can be seen by whether assortative matching in one pair of characteristics dominates assortative matching in another pair when the distribution of firm characteristics prevents assortative matching in both pairs of characteristics simultaneously. In marriage, an example could be when agents with high levels of intelligence have poor physical appearances. If physically attractive people marry each other, this is evidence complementarities for physical appearance exceed those for intelligence.

5.2.2 Cardinal identification with match-specific characteristics

The characteristics in $\vec{x}(i, C^u)$ can be specific to the individual matches $\langle a, i\rangle$. Now let C^u be a set of n downstream firms and let

$$\vec{x}(i, C^u) = \text{cat}\left(\left(x^{u,d}_{\langle a_1,i\rangle,1}, \ldots, x^{u,d}_{\langle a_1,i\rangle,K}\right), \ldots, \left(x^{u,d}_{\langle a_n,i\rangle,1}, \ldots, x^{u,d}_{\langle a_n,i\rangle,K}\right)\right),$$

where there are K characteristics for each potential match $\langle a, i\rangle$, with $x^{u,d}_{\langle a,i\rangle,e}$ being the eth such scalar characteristic. Consider an application to international trade, where supplier i may be in a different country than assembler a. The match-specific characteristic $x^{u,d}_{\langle a,i\rangle,e}$ may be the tariff rate that a's country levies on i's exports. In this case, the feature of f that governs sorting is f's own-second derivatives.[42]

Theorem 2. *Let the two scalars x_1 and x_2 be distinct elements of $\vec{x}(i, C^u) = \text{cat}\left(\left(x^{u,d}_{\langle a_1,i\rangle,1}, \ldots, x^{u,d}_{\langle a_1,i\rangle,K}\right), \ldots, \left(x^{u,d}_{\langle a_n,i\rangle,1}, \ldots, x^{u,d}_{\langle a_n,i\rangle,K}\right)\right)$, corresponding to different matches. Let $\vec{x}$ be a given point of evaluation of f.*

1. *Assume $\frac{\partial^2 f^0(\vec{x})}{\partial^2 x_1} \neq 0$. The sign of $\frac{\partial^2 f^0(\vec{x})}{\partial^2 x_1}$ is identified.*
2. *Assume $\frac{\partial^2 f^0(\vec{x})}{\partial^2 x_1} \neq 0$ and $\frac{\partial^2 f^0(\vec{x})}{\partial^2 x_2} \neq 0$. The ratio $\frac{\partial^2 f^0(\vec{x})}{\partial^2 x_1} / \frac{\partial^2 f^0(\vec{x})}{\partial^2 x_2}$ is identified.*

Like with firm-specific characteristics, a researcher can measure the relative importance of sorting on various characteristics in the production function, $\frac{\partial^2 f^0(\vec{x})}{\partial^2 x_1} / \frac{\partial^2 f^0(\vec{x})}{\partial^2 x_2}$.[43]

Note that the proofs of the cardinal identification theorems are more general than the statements. The proofs do not require strong properties on the characteristics not given special attention in the statement of the theorems. For example, x_1 and x_2 could be match-specific, allowing Theorem 2 to be applied, while x_3–x_6 could be firm specific, requiring Theorem 1. The presence of the match-specific x_1 and x_2 does not invalidate applying Theorem 1 to x_3–x_6.

[41] The identification of $\frac{\partial f^0(\vec{x})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{x})}{\partial x_3 \partial x_4}$ seems parallel to the identification of marginal rates of substitution in single-agent choice. With this parallel, using the word "cardinal" may seem inappropriate as the ratio of marginal utilities is preserved under positive monotonic transformations of utility. However, the ratio of cross-partial derivatives is not preserved under positive monotonic transformations of utility. Therefore, there is no fundamental link between Theorem 1 and the ordinal identification results in Section 6.

[42] If f is differentiable, a function f is concave in x_1 at $\vec{x}$ when $\frac{\partial^2 f(\vec{x})}{\partial^2 x_1} > 0$ and convex when $\frac{\partial^2 f(\vec{x})}{\partial^2 x_1} < 0$.

[43] Some might think firm-specific characteristics are a special case of match-specific characteristics. Firm-specific characteristics increase the difficulty of showing the identification of features of f^0 because the same firm characteristics must appear on the left and the right sides of a local production maximization inequality. By contrast, hypothetical markets exist where the match-specific characteristics for the matches $\langle a, i\rangle$ and $\langle a, j\rangle$ may take on any pair values, under Assumption 2.

5.2.3 Cardinal identification with group-specific characteristics

Characteristics can be specific to the group of downstream firms and an upstream firm that the group matches with. Let $\vec{x}(i,C^u) = \left(x^{\text{group}}_{(i,C^u),1}, \ldots, x^{\text{group}}_{(i,C^u),K}\right)$, where there are K scalar group-(i,C^u)-specific characteristics, with the eth such scalar characteristic being $x^{\text{group}}_{(i,C^u),e}$. An example of using the estimator in this paper for the group-characteristic case is Bajari and Fox (2007), who model bidders matching to a package of geographic licenses in a spectrum auction. A characteristic of a package of licenses is the extent of the geographic complementarities among the licenses. Bajari and Fox use a measure like the gravity equation in international trade to create a proxy for these geographic complementarities.

Theorem 3. *Let the two scalars x_1 and x_2 be distinct elements of $\vec{x}(i,C^u) = \left(x^{\text{group}}_{(i,C^u),1}, \ldots, x^{\text{group}}_{(i,C^u),K}\right)$, corresponding to different matches. Let $\vec{x}$ be a given point of evaluation of f.*

1. *Assume $\frac{\partial^2 f^0(\vec{x})}{\partial^2 x_1} \neq 0$. The sign of $\frac{\partial^2 f^0(\vec{x})}{\partial^2 x_1}$ is identified.*
2. *Assume $\frac{\partial^2 f^0(\vec{x})}{\partial^2 x_1} \neq 0$ and $\frac{\partial^2 f^0(\vec{x})}{\partial^2 x_2} \neq 0$. The ratio $\frac{\partial^2 f^0(\vec{x})}{\partial^2 x_1} / \frac{\partial^2 f^0(\vec{x})}{\partial^2 x_2}$ is identified.*

The proof of this theorem is omitted because the mathematical argument is nearly identical to the proof of Theorem 2.

5.3 A parametric, cardinally identifiable class

For practical estimation, empirical researchers should work with a production function f that is uniquely determined given the features that the previous theorems show are identified. Consider a parametric example of specifying a production function class $\mathscr{F}$ when only cross-partial derivative information is identifiable. Consider the case where all inputs are continuous and firm-specific. The production function

$$f_\beta\left(\vec{x}(i,C^u)\right) = \sum_{c=1}^{K^u} \sum_{e=1}^{K^d} \left(\beta^{u,d}_{c,e} \sum_{a \in C^u} x^u_{i,c} \cdot x^d_{a,e} \right),$$

forms all pairwise multiplicative interactions between individual characteristics of the upstream firm, i, and the individual characteristics of each downstream firm. Each upstream firm characteristic is indexed by c and each downstream firm characteristic is indexed by e. Each parameter $\beta^{u,d}_{c,e}$ is a positive cross-derivative that remains constant over the entire support of the production function, which is unnecessarily restrictive. The production function is exchangeable in the identities of downstream firms: $\beta^{u,d}_{c,e}$ is the same for all downstream firms. A matching is a qualitative outcome, so a scale normalization is needed. One normalization is $\beta^{u,d}_{1,1} = \pm 1$, for the first characteristic of each upstream firm and each downstream firm.[44]

This parametric example is additively separable across downstream firms. Many empirical applications, such as the work on automotive supplier specialization in this paper, will involve interactions between multiple downstream firms matched to the same upstream firm. Interactions between the characteristics of downstream firms could be added, as in

$$f_\beta\left(\vec{x}(i,C^u)\right) = \sum_{c=1}^{K^u} \sum_{e=1}^{K^d} \left(\beta^{u,d}_{c,e} \sum_{a=1}^{n} x^u_{i,c} \cdot x^d_{a,e} \right) + \sum_{e=1}^{K^d} \sum_{g=1}^{K^d} \left(\beta^{d,d}_{e,g} \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} x^d_{a,e} \cdot x^d_{b,g} \right),$$

[44] If a researcher uses an extremum estimator, he or she can estimate the model once for $\beta^{u,d}_{1,1} = +1$ and another time for $\beta^{u,d}_{1,1} = -1$ and choose as the final estimates the parameter vector with the highest objective function values.

for $C^u = \{1, \ldots, n\}$, where the new parameters $\beta_{e,g}^{d,d}$ reflect the interaction of individual characteristics between two downstream firms, a and b. If the production function is to be exchangeable in the identities of downstream firms, $\beta_{e,g}^{d,d} = \beta_{g,e}^{d,d}$.

6 Ordinal nonparametric identification

The cardinal identification analysis is sufficient for many matching empirical applications. However, it is impossible to use the cardinal analysis to tell apart two production functions with, for firm-specific characteristics, the same cross-partial derivatives. For example, $f^1((x_1,x_2)) = -(x_1 - x_2)^2$ and $f^2((x_1,x_2)) = 2x_1 \cdot x_2$ both have $\frac{\partial f(\vec{x})}{\partial x_1 \partial x_2} = 2$. For f^1, $f^1((1,1)) = 0$ and $f^1((2,1)) = -1$. For f^2, $f^2((1,1)) = 2$ and $f^2((2,1)) = 4$. So $f^1((1,1)) > f^1((2,1))$ yet $f^2((1,1)) < f^2((2,1))$. Identifying cross-partial derivatives does not tell us whether production is higher at one argument $\vec{x}$ than another argument.

I now focus on the ordinal identification of f. By ordinal identification, I mean that for any two sets of characteristics for a group of matches $\vec{x}_1$ and $\vec{x}_2$, I wish to know whether $f(\vec{x}_1) > f(\vec{x}_2)$ or the reverse. Ordinal identification is helpful in distinguishing whether an individual match characteristic x_k in $\vec{x}$ is actually a "good" that raises output.

Matzkin (1993) proves the ordinal identification of utility functions for the single-agent multinomial choice model under Manski (1975)'s assumptions about the error terms: the single-agent equivalent of Assumption 1. The preliminaries of my analysis follow Matzkin's. However, in the single-agent model, one can vary the characteristics of the choices facing the single agent. In a matching market, an agent must pay the appropriate transfer to match with a partner, and that transfer is both an outcome of the game and assumed to not be in the data. Therefore, I extend the mathematical arguments in Matzkin to show ordinal identification of, here, the production function f, by using only exogenous information on X, the collection of characteristics of all agents and potential matches in a matching market. In other words, I work with the equilibrium structure of the game and variation in the exogenous market-level characteristics of matches to show identification.

6.1 Preliminaries for ordinal identification

Positive monotonic transformations preserve ordinal rankings, so we must rule those transformations out. The following assumption previews the result that production functions can be identified up to a positive monotonic transformation. This is good news, as it might have been the case that the equilibrium concept of pairwise stability, which involves exchanges of one downstream firm per upstream firm, is not enough to identify production functions involving coalitions with many downstream firms. For example, inequalities that involve the exchange of only one downstream firm per upstream firm can identify a production function with ten matches, even if each match being is represented by a vector of five characteristics, for fifty arguments total.[45]

Assumption 3. Let $\mathscr{F}$ be a class of production functions. For any two members of this class $\mathscr{F}$, f^1 and f^2, for no positive, strictly monotonic function m is it the case that $f^1(\vec{x}) = m \circ f^2(\vec{x})$ for all $\vec{x}$.

Matzkin (1993) presents classes of functions that rule out positive monotonic transformations. An example is the class of least-concave functions.

[45] With real data, this would be implemented using a parametric functional form because of the curse of dimensionality in nonparametric estimation.

Manski (1975, 1985, 1988) discusses the semiparametric point identification of single-agent discrete choice models using an assumption that at least one covariate entering the payoff of each choice has a continuous distribution. Matzkin (1991, 1993) extends the semiparametric identification arguments of Manski to single-agent discrete choice models where the deterministic payoff of a choice is a nonparametric function of characteristics. The equivalent characteristic assumption for matching follows. Recall that $\vec{x}$ is one long vector of scalar characteristics. Call the first, scalar element of this vector, x_1. Call all other elements $\vec{x}_{-1}$. The collection of market characteristics is X.

Assumption 4.

- The conditional density of characteristic 1, $g(x_1 \mid \vec{x}_{-1}, X \backslash \vec{x})$, has an everywhere positive density in $\mathbb{R}$.[46]
- Each $f \in \mathscr{F}$ is continuous in its argument x_1 and is either strictly increasing or strictly decreasing in x_1.
- X has support equal to the product of the marginal supports of the scalar elements of X.
- Each scalar element of X has either strictly discrete or strictly continuous support.
- Each $f \in \mathscr{F}$ is continuous in any scalar element of X with continuous support.

The assumption allows all but one of the characteristics to have discrete or qualitative support.[47] This assumption replaces the earlier Assumption 2, which is only for the cardinal identification theorems.[48]

The technical use of Assumption 4 involves the lack of a positive monotonic transformation and its relationship to a strict inequality. I state the argument in a separate lemma because the continuous covariate argument is used in the same way in the proofs of the three ordinal identification theorems.

Lemma 3. Let f^1 and f^2 be production functions in a class $\mathscr{F}$ satisfying Assumption 3. If Assumption 4 holds, then there exists two vectors $\vec{x}_a$ and $\vec{x}_b$ such that either

$$f^1(\vec{x}_a) > f^1(\vec{x}_b) \text{ and } f^2(\vec{x}_a) < f^2(\vec{x}_b)$$

or

$$f^1(\vec{x}_a) < f^1(\vec{x}_b) \text{ and } f^2(\vec{x}_a) > f^2(\vec{x}_b).$$

This lemma will be used where $\vec{x}_a$ and $\vec{x}_b$ are part of X for the same matching market.

6.2 Ordinal identification theorems

Identification proofs in the single-agent maximum score tradition (Matzkin, 1993) typically amount mathematically to Lemmas 2 and 3. Consequently, the identification proof for case each focuses on an issue that is new

[46] The notation $X \backslash \vec{x}$ means all elements of X other than those in the specified vector $\vec{x}$.

[47] The assumption that the support of x_1 is $\mathbb{R}$, rather than some compact subset of $\mathbb{R}$, is made for convenience. Manski (1988) and Horowitz (1998) show how to relax the full support assumption for the identification of single-agent binary choice models. A continuous product quality could be a candidate for the continuous upstream product characteristic x_1.

[48] 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). Identification based on special regressor arguments might be possible if there are match-specific regressors with full support and independence from the error terms. Arguments exist to weaken the full support assumptions (Magnac and Maurin, 2007). Special regressor identification arguments do not lead to tractable estimators for matching games. The Lewbel single-agent, multinomial choice estimator requires multidimensional density estimation and therefore suffers from a data curse of dimensionality in the number of choices.

to matching games: embedding the inequalities from Lemma 3 in a local production maximization inequality, meaning an inequality where each upstream firm switches at most one downstream firm at a time. Thus, the proofs look for market characteristics X where the comparisons in Lemma 3 are decisive in rank ordering the production of two larger, otherwise similar assignments, A_1 and A_2. As with cardinal identification, I consider group, match and firm-specific characteristics separately. Identification is Definition 4 subject to the lack of a positive monotonic transformation in Assumption 3. I list the theorems in increasing difficulty of the proofs.

Group-specific characteristics allow the arguments of production functions to move around more flexibly than in the other cases.

Theorem 4. *Let* $\vec{x}(i,C^u) = \left(x^{\text{group}}_{(i,C^u),1}, \ldots, x^{\text{group}}_{(i,C^u),K}\right)$*. Then the production function is identified in the class* $\mathscr{F}$*.*

Match-specific characteristics make the identification proof more complex than before. The reason is that the equilibrium concept of pairwise stability, Definition 2, involves only one unmatched pair deviating at a time. To show identification, we must start with Lemma 3 and be able to construct local production maximization inequalities where the coalition characteristics differ by the arguments corresponding to only one match between an upstream and a downstream firm. Remember the production function requires a vector of arguments for each match of an upstream firm.

Theorem 5. *Let* $\vec{x}(i,C^u) = \text{cat}\left(\left(x^{u,d}_{\langle a_1,i\rangle,1}, \ldots, x^{u,d}_{\langle a_1,i\rangle,K}\right), \ldots, \left(x^{u,d}_{\langle a_n,i\rangle,1}, \ldots, x^{u,d}_{\langle a_n,i\rangle,K}\right)\right)$*. Also, let there be assignments A that contain as many groups of matches as the maximum quota of an upstream firm. The production function is identified in the class* $\mathscr{F}$*.*

The theorem requires the matching market to be sufficiently large so that the comparisons needed for identification can be formed. The matching market may need to allow several firms on each side of the market because pairwise stability considers firms swapping only one partner at a time, while a production function can have as its arguments the characteristics of the matches involving many downstream firms.

Firm-specific characteristics require an additional normalization. Ordinal identification requires the ability to distinguish production functions with additive, firm-specific components, such as $f\left(\text{cat}\left(\vec{x}^u_i, \vec{x}^d_a\right)\right) = \beta_u'\vec{x}^u_i + \beta_d'\vec{x}^d_a$. The availability of data for firms that are unmatched and an assumption that the production of an unmatched firm is always zero allow for identification. Some location normalization of the production of the matches involving both upstream and downstream firms is necessary, as (6) shows non-interacted terms, of which a constant is a special case, drop out in the local production maximization inequality. If unmatched firms are observable in data and the value of an unmatched firm is set to zero, then non-interacted terms that only "turn on" when a firm is matched are identifiable.[49] Informally, identification considers the probabilities of assignments where certain firms are unmatched.[50] Firms that are more likely to be unmatched in an assignment are likely to have lower contributions to production.

Theorem 6. *Let* $\vec{x}(i,C^u) = \text{cat}\left(\left(x^u_{i,1}, \ldots, x^u_{i,K^u}\right), \left(x^d_{a_1,1}, \ldots, x^d_{a_1,K^d}\right), \ldots, \left(x^d_{a_n,1}, \ldots, x^d_{a_n,K^d}\right)\right)$*. Let the value of any firm remaining unmatched be 0, or* $f\left(\vec{x}^d_a\right) = f\left(\vec{x}^u_i\right) = 0$ *for all* $i \in U$*,* $a \in D$ *and* $f \in \mathscr{F}$*. Further, let there be assignments A that contain as many matched coalitions as three times the maximum quota of an upstream firm. Then the production function is identified in the class* $\mathscr{F}$*.*

[49] What is important for the identification argument is that a specified class of partners for each firm have the same, known value for each firm. Being unmatched is commonly used in the linear programming formulations of one-to-one matching games (Koopmans and Beckmann, 1957).

[50] Choo and Siow (2006) estimate a logit-based one-to-one matching model of marriage that requires data on the fraction of each observable type of man or woman that is single.

7 Estimation as the number of markets gets large

I now discuss how the previous identification arguments can form the basis for a practical estimator. A researcher has some data from M statistically independent matching markets. The econometrician observes the exogenous characteristics X_m and the endogenous market assignment A_m for each market m. The asymptotics will be in M.

There are at least two curses of dimensionality in matching estimation. First, following the identification argument directly leads to a potentially consistent estimator of the production function, but a data curse of dimensionality. One nonparametrically estimates the assignment probabilities $\Pr(A \mid X) = \Pr\left(A \mid X; f^0, S^0\right)$ with data from multiple markets. The estimates $\hat{\Pr}(A \mid X)$ allow some version of the identification arguments to be applied. As $\hat{\Pr}(A \mid X)$ converges in probability to $\Pr(A \mid X)$, the estimator $\hat{f}$ will likely converge to f^0, the true production function in the data generating process. Unfortunately, the first-stage nonparametric estimation of $\Pr(A \mid X)$ is not a tractable strategy for many datasets as the dimensionality of both A and X in $\Pr(A \mid X)$ may be high (thousands or millions of elements in X is not unusual).

Likewise, a traditional maximum likelihood estimator faces a computational curse of dimensionality in having to simulate equation (11), even for the simple case of one-to-one matching. The number of error terms being integrated out in (11) is on the order of $U \cdot D$, the total number of potential matches in a market. The integrand involves computing an equilibrium assignment for a matching market. In one-to-one matching, such a computation is a linear program that itself can be computationally costly. Alternatively, for a given pair (A_m, X_m), the local production maximization inequalities from Lemma 1 might need to be checked to construct at least an upper bound to the likelihood. The text following Lemma 1 discusses why matching combinatorics make checking all inequalities computationally difficult. Multiple equilibrium assignments add additional complications for equilibrium assignment computations.

This section provides a maximum score estimator, which follows the work of Manski (1975) for the single agent model. The maximum score estimator works directly with the production function f instead of $\Pr(A \mid X)$. Also, the maximum score estimator avoids all nested computations. Further, all inequalities do not need to be included with probability 1 to maintain the consistency of the estimator.

7.1 Semiparametric vs nonparametric estimation

I restrict attention to a semiparametric version of the estimator to focus attention on practical implementation.[51] "Semiparametric" refers to not imposing a parametric structure on S in $\Pr\left(A \mid X; f_\beta, S\right)$ while specifying f_β up to a finite vector of parameters β. The parameter space for f_β is a special case of the earlier identification assumptions. Here I do not use any special properties of a given parametric class, such as the linear index $f_\beta(\vec{x}) = \vec{x}'\beta$ in Manski (1975).

Assumption 5.

1. The production function $f_\beta \in \mathscr{F}$ where $\mathscr{F} = \left\{f_\beta\right\}_{\beta \in \mathscr{B}}$ and $\mathscr{B} \subseteq \mathbb{R}^{|\beta|}$, $|\beta| < \infty$. $\mathscr{B}$ is compact.

[51] The matching identification theorems are nonparametric. Matzkin (1993) studies the nonparametric maximum score estimation of single-agent discrete choice models with more than two choices. She examines several technical issues that are needed for estimation using an infinite dimensional space of functions $\mathscr{F}$. These include appropriate metrics to define compactness of $\mathscr{F}$, regularity conditions on $\mathscr{F}$, and the proof of the uniform convergence of the objective function. Matzkin's arguments can be extended to the matching maximum score estimator.

2. Identification: Let $\beta^0 \in \mathscr{B}$ and $S^0 \in \mathscr{S}$ be the true parameters that generate the data. For any $\beta^1 \neq \beta^0$, $\beta_1 \in \mathscr{B}$, and for any $S^1 \in \mathscr{S}$, there exists a set of market characteristics $\mathscr{X}$ with positive probability and two assignments A_1 and A_2 such that $\Pr\left(A_1 \mid X; f_{\beta^0}, S^0\right) > \Pr\left(A_2 \mid X; f_{\beta^0}, S^0\right)$ while $\Pr\left(A_1 \mid X; f_{\beta^1}, S^1\right) < \Pr\left(A_2 \mid X; f_{\beta^1}, S^1\right)$ for any $X \in \mathscr{X}$.

3. Each $\vec{x}$ in X has one or more elements with continuous support.

4. X is independently and identically distributed across markets.

As I have shown nonparametric identification of f^0, or of features of f^0, under various conditions, the parametric restrictions in Assumption 3 are purely for computationally tractable estimation. Part 2 assumes identification to focus on the asymptotic properties of the estimator. Part 2 will be enough to show that β^0 is the unique global maximum to the maximum score objective function. See the example in Section 5.3 for an example of choosing f_β based on the identifiable features from the cardinal nonparametric identification theorems.

7.2 The matching maximum score estimator

First say the researcher examines a one-to-one matching model such as marriage. Assume the researcher uses data on only matched couples. Then the maximum score objective function is

$$Q_M(\beta) = \frac{1}{M} \sum_{m \in M} \sum_{\langle a,i \rangle, \langle b,j \rangle \in A_m} 1\left[f_\beta(\vec{x}(i,\{a\})) + f_\beta(\vec{x}(j,\{b\})) > f_\beta(\vec{x}(i,\{b\})) + f_\beta(\vec{x}(j,\{a\}))\right]. \tag{13}$$

The indicator functions $1[\cdot]$ are equal to 1 when the local production maximization inequality in brackets is true and 0 otherwise. The score of correct predictions increases by 1 when a production maximization inequality holds for a trial guess of β. The matching maximum score estimator $\hat{\beta}_M$ receives the highest score of satisfied inequalities. The fraction of satisfied inequalities is a measure of statistical fit such as R^2 in a regression. As the objective function is a step function, there will always be more than one global maximum; finding one is sufficient for estimation.

More generally, given A_m and X_m, let I_m be the inequalities that the econometrician includes for market m. An inequality in I_m is indexed by the matches $B_1 \subseteq A_m$ that index matches on the left side that will be dissolved and the matches B_2 on the right side formed from those dissolutions. See Definition 3 for the definitions of B_1, B_2 and a local production maximization inequality. The maximum score estimator is any parameter vector $\hat{\beta}_M$ that maximizes

$$Q_M(\beta) = \frac{1}{M} \sum_{m \in M} \sum_{(B_1 B_2) \in I_m} 1\left[\sum_{\langle a,i \rangle \in B_1} f_\beta\left(\vec{x}\left(i, C_i^u(A_m)\right)\right) > \sum_{\langle a,i \rangle \in B_2} f_\beta\left(\vec{x}\left(i, C_i^u((A_m \backslash B_1) \cup B_2)\right)\right)\right]. \tag{14}$$

Evaluating $Q_M(\beta)$ is computationally simple: there is no nested equilibrium computation to a matching game, as say Pakes (1986) and Rust (1987) proposed for dynamic programming problems. Another key idea behind the computational simplicity of maximum score estimation is that there are no error terms $\varepsilon_{\langle a,i \rangle}$ in (14), even though the estimator may perform well if the data are generated from a model with such errors (see Sections 4.4 and 8). Not all inequalities will be satisfied, even at the maximizer $\hat{\beta}_M$ and even at the probability limit of the objective function.[52]

[52]This distinguishes maximum score from a moment inequality approach.

7.3 Consistency and inference

Even given these computational advantages over methods that require nested solutions to matching games or numerical integration, there can still be a very large number of local production maximization inequalities, as Section 3.3 makes clear for the automotive assembler and supplier example. If all such inequalities were included, the estimator would suffer from a computational curse of dimensionality. Luckily, inequalities only need to be included with some positive probability for the estimator to be consistent.[53] This means researchers can sample from the set of theoretically valid inequalities. Let $N(A,X)$ be this set of theoretically valid local production maximization inequalities of the form (B_1,B_2), given assignment A (A_1 in the notation of Lemma 1) and observable characteristics X. Let $I(B_1,B_2 \mid X)$ be the probability, conditional on X, that a researcher includes an inequality when $(B_1,B_2) \in N(A,X)$. Hence, $I(B_2,B_1 \mid X)$ is the probability of sampling (B_2,B_1) when $(B_2,B_1) \in N(A_2,X)$, for some other assignment A_2.

Assumption 6. For all $(B_1,B_2) \in N(A,X)$ and for any feasible pair (A,X),

1. $I(B_1,B_2 \mid X) = I(B_2,B_1 \mid X)$.

2. $I(B_1,B_2 \mid X) > 0$.

Because all inequalities needed for identification are included in the limit as $M \to \infty$, sampling inequalities does not change point identification to set identification.

The following theorem states that the matching maximum score estimator is consistent. The asymptotics are in the number of independent markets.

Theorem 7. *As $M \to \infty$, any $\hat{\beta}_M \in \mathscr{B}$ that maximizes the matching maximum score objective function is a consistent estimator of $\beta^0 \in \mathscr{B}$, the parameter vector in the data generating process.*

The proof is a straightforward application of a general consistency theorem for extremum estimators in Newey and McFadden (1994), which generalizes the early work of Manski (1975, 1985) on maximum score. The insight here is not the consistency proof, but the general idea that maximum score can be interpreted as a necessary conditions approach for inequalities, at least for matching games with transfers. In terms of data requirements and computation, two practical aspects of the estimator are that $\Pr\left(A \mid X; f_\beta, S\right)$ does not have to be manually computed for each guess of β and S and $\Pr(A \mid X)$ does not need to be nonparametrically estimated in a first stage. The maximum score estimator is consistent in part because of a law of large numbers, as by the law of iterated expectations over the random variables A and X,

$$\text{plim}_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} 1\left[A_m = A\right] = E_X \left\{\Pr(A \mid X)\right\},$$

where $1\left[A_m = A\right]$ equals 1 if assignment A occurs in market m.[54]

[53] This estimator will not have a normal distribution, if asymptotics are in the number of markets. Therefore, I will avoid discussing how the choice of inequalities relates to statistical efficiency. For consistency, all valid inequalities from Lemma 1 do not need to be included with positive probability for some parametric functional forms. Consider the standard linear index case $f_\beta(\vec{x}) = \vec{x}'\beta$. Then results in Manski (1975) can be adapted to show only inequalities involving pairs of matches $B_1 = \{\langle a,i\rangle, \langle b,j\rangle\}$ need to be included.

[54] The proof shows that the true production function f^0 maximizes the probability limit of the objective function. 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. The rank order property suggests maximizing the number of correct inequalities, not allowing a violation in one inequality in order to minimize the degree of a violation in another inequality.

Kim and Pollard (1990) show that the binary (two) choice maximum score estimator converges at the rate of $\sqrt[3]{M}$ (instead of the more typical $\sqrt{M}$) and that its limiting distribution is too complex for use in inference. Abrevaya and Huang (2005) show that the bootstrap is inconsistent while Delgado, Rodríguez-Poo and Wolf (2001) show that another resampling procedure, subsampling, is consistent. Subsampling was developed by Politis and Romano (1994). The book Politis, Romano and Wolf (1999) provides a detailed overview of subsampling.

An alternative to subsampling is smoothing the indicator functions in the maximum score objective function. For the single-agent binary-choice maximum score estimator, Horowitz (1992) proves that a smoothed estimator converges at a rate close to $\sqrt{M}$ (the exact rate depends on the smoothing parameter and smoothness assumptions about the model) and is asymptotically normal with a variance-covariance matrix than can be estimated and used for inference. Further, Horowitz (2002) shows the bootstrap is consistent for his smoothed maximum score estimator.

7.4 Computation

Because of the combinatorics inherent in matching games and the multiple equilibrium assignments in some matching games, maximum score has advantages over methods that require repeatedly computing equilibrium assignments or computing high-dimensional integrals. Still, some readers may be concerned about optimization for the maximum score estimator. The maximum score estimator is a step function. 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 local optimization routine to converge to.

Manski and Thompson (1986) and Pinkse (1993) present optimization algorithms for the single-agent maximum score estimator where the parameters enter linearly into the payoff function. In the Monte Carlo study and in the empirical application below, I numerically maximize the maximum score objective function using the global optimization routine known as differential optimization (Storn and Price, 1997).

For a finite sample, the objective function is a step function and so there is a continuum of global maxima, even if the parameter β is point identified asymptotically. Any global maximum is a consistent estimator. Smoothing the indicator functions will remove this numerical indeterminacy.

8 Monte Carlo evidence

This section presents evidence that the maximum score estimator works well in finite samples and with iid match-specific errors. The Monte Carlo study examines games of one-to-one, two-sided matching. Section 4.4 provides background on this class of games and argues that the rank order property holds only approximately under iid match-specific errors. I present an alternative sufficient condition involving a social planner's errors, but in the Monte Carlo study I restrict attention to the iid errors case.

Each agent is distinguished by two characteristics, for upstream firm i, $x^u_{1,i}$ and $x^u_{2,i}$, and for downstream firm a, $x^d_{1,a}$ and $x^d_{2,a}$. The total output from a match of i to a is

$$f_{\beta_1,\beta_2}\left(\text{cat}\left(\vec{x}^u_i,\vec{x}^d_a\right)\right)+\varepsilon_{\langle a,i\rangle}=\beta_1 x^u_{1,i}\cdot x^d_{1,a}+\beta_2 x^u_{2,i}\cdot x^d_{2,a}+\varepsilon_{\langle a,i\rangle}.$$

I impose the scale normalization $\beta_1 = \pm 1$. The sign of β_1 is superconsistently estimable, so I set it to the true value of $+1$ throughout the study. For each side of the market and upstream firms as an example,

$$\begin{bmatrix} x^u_{1,i} \\ x^u_{2,i} \end{bmatrix} \sim N\left(\begin{bmatrix} 10 \\ 10 \end{bmatrix}, \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}\right).$$

The high means of 10 ensure that the characteristic values are usually positive. The nonzero covariance suggests a multivariate estimator might give different estimates than a univariate estimator. I set $\beta_2 = 1.5$, so that the second observable characteristic is more important in sorting.

In the first set of experiments, the match-specific $\varepsilon_{\langle a,i\rangle}$'s are iid with a normal distribution with a standard deviation of either 1 or 5. In the second set of experiments, the match-specific $\varepsilon_{\langle a,i\rangle}$'s are iid with one of two mixed normal distributions, each with two components. The first mixed normal distribution is $0.35 \cdot N\left(-5, 2^2\right) + 0.65 \cdot N\left(2, 2^2\right)$, which has a standard deviation of 1.43. The second distribution is $0.35 \cdot N\left(-5, 2^2\right) + 0.65 \cdot N\left(2, 5^2\right)$, which has a standard deviation of 3.33.

I sample match specific errors and solve for the optimal assignment using the linear programming problem discussed in Section 2.2. The linear programming formulation ensures that all consummated matches provide non-negative surplus. Very few of the agents are unmatched in the fake data as the means of both characteristics are high.

While not shown, I have plotted scatterplots of the characteristics for matched pairs. Consider fake data with 30 upstream and 30 downstream firms and an error standard deviation of 1. Because $\beta_2 = 1.5 > \beta_1 = 1$, typically the matched firms will appear more assortatively matched on characteristic 2 than 1. With an error standard deviation of 5, positive assortative matching on either characteristic will be hard to visually detect in the fake data.[55]

Table 2 reports estimates of the bias and root mean-squared error (RMSE) of the matching maximum score estimates under various specifications. Consider the upper-left panel: normal errors with a standard deviation of 1. The bias and RMSE are high for 3 downstream and 3 upstream firms (6 total) for each market and 100 markets. The bias and RMSE are larger for 10 firms on each side of the market and only 10 markets. However, both the bias and RMSE decrease when more firms are added to each market: the third row reports 30 firms on each side and 10 markets. The bias and RMSE decrease further with 60 firms on each side and 10 markets. The fifth row then shows that increasing the number of markets to 40 further reduces the bias and RMSE.

Another question is how well the estimator works in a finite sample with data on only one fairly large matching market. The seventh row of the upper-left panel uses 100 firms on each side of the market, but only one market. The bias and RMSE then decline in the eighth row as the number of firms on each side increases to 200.

Qualitatively similar changes in the bias and RMSE occur for each of the other three panels: normal errors with a larger standard deviation (no visual sorting pattern in the data) and two forms of the mixed normal distribution. Using a bimodal, mixed normal distribution suggests that the estimator fulfills its semiparametric claims: it is not so sensitive to the distribution of the data generating process. In these experiments, the estimator is not very biased when there are iid errors, despite the discussion in Section 4.4.

[55]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 range of global maxima is small.

Table 2: Monte Carlo results, true value $\beta_2 = 1.5$

Normal errors					Mixed normal errors, two components, asymmetric				
# Upstr. & # Downstr.	# Markets	Errors std. dev.	Bias	RMSE	# Upstr. & # Downstr.	# Markets	Errors std. dev.	Bias	RMSE
3	100	1	0.154	0.896	3	100	1.48	0.216	1.13
10	10	1	0.431	2.61	10	10	1.48	0.254	1.11
30	10	1	0.026	0.406	30	10	1.48	0.099	0.454
60	10	1	0.016	0.264	60	10	1.48	0.028	0.281
60	40	1	0.013	0.154	60	40	1.48	0.010	0.160
100	1	1	0.074	0.551	100	1	1.48	0.084	0.600
200	1	1	0.058	0.355	200	1	1.48	0.022	0.381
3	100	5	0.484	3.32	3	100	3.33	0.426	2.77
10	10	5	0.602	2.80	10	10	3.33	0.336	1.56
30	10	5	0.185	0.852	30	10	3.33	0.072	0.600
60	10	5	0.054	0.442	60	10	3.33	0.058	0.382
60	40	5	0.011	0.056	60	40	3.33	0.030	0.199
100	1	5	0.231	1.05	100	1	3.33	0.129	0.721
200	1	5	0.128	0.398	200	1	3.33	0.056	0.444

The true parameter is $\beta_2 = 1.5$. The population bias is $E\left[\hat{\beta}_2 - 1.5\right]$, and the population RMSE is $\sqrt{E\left[\left(\hat{\beta}_2 - 1.5\right)^2\right]}$, where 1.5 is the value of β_2 used to generate the fake data.

The model is estimated 500 or 1000 times for each simulation of bias and RMSE. A fake dataset consists of the listed number of independent markets. New observable variables X and match-specific errors of the form $\varepsilon_{\langle a,i\rangle}$ are drawn for each market and each replication. Each market is a one-to-one, two-sided matching game. The number of upstream firms (or men) always equals the number of downstream firms (or women). The equilibrium assignment is calculated using a linear programming problem, as discussed in Section 2.2.

The distribution of the fixed agent types is given in the text. On the left table, the errors $\varepsilon_{\langle a,i\rangle}$ have $N\left(0,\sigma^2\right)$ distributions, where σ is the standard deviation listed in the table. In the top half of the right table, the errors have the mixed normal distribution $0.35 \cdot N\left(-5,2^2\right) + 0.65 \cdot N\left(2,2^2\right)$, which has the standard deviation listed in the table. This is a bimodal density. In the bottom half of the right table, the error distribution is $0.35 \cdot N\left(-5,2^2\right) + 0.65 \cdot N\left(2,5^2\right)$.

Each agent has a vector of two types. The coefficient on the product of the first types is normalized to one. The estimate of the sign of the coefficient is superconsistent and so I do not explore its finite sample properties.

9 Empirical application to automotive suppliers

I now present an empirical application about the matching of suppliers to assemblers in the automobile industry. Automobile assemblers are well-known, large manufacturers, such as BMW, Ford or Honda. Automotive suppliers are less well-known to the public, and range from large companies such as Bosch to smaller firms that specialize in one type of car part. A car is one of the most complicated manufacturing goods sold to individual consumers. Making a car be both high quality and inexpensive is a technical challenge. Developing the supply chain is an important part of that challenge. More so than in many other manufacturing industries, suppliers in the automobile industry receive a large amount of coverage in the industry press because of their economic importance.

A matching opportunity in the automotive industry is an individual car part that is needed for a car. Let L_a be the set of parts assembler $a \in D$ needs suppliers for. A particular part $l \in L_a$ in the data is attached to a supplier, $i \in U$. Therefore a match in this industry is a triple $\langle a,i,l \rangle$. The same supplier can supply more than one part to the same assembler: $\langle a,i,l \rangle$ and $\langle a,i,h \rangle$ represent two different matches (car parts) between assembler a and supplier i. This is a two-sided, many-to-many matching game between assemblers and suppliers, with the added wrinkle that a supplier can be matched to the same assembler multiple times.[56]

The data group car parts into component categories, and I treat each component category as a statistically independent matching market. In my data, there are 593 distinct component categories, such as "Pedal Assembly" and "Coolant/Water Hoses." I assume any nonlinearities between multiple matches involving the same supplier occur only within component categories; there are no spillovers across the different matching markets. A triplet $\langle a,i,l \rangle$ in the data then could be the front pads of a Fiat 500 (a car) supplied by Federal-Mogul. Front pads are in the component category (matching market) disk brakes.

The automotive supplier empirical example is a good showcase for the strengths of the matching estimator. The matching markets modeled here contain many more agents than the markets modeled in some other papers on estimating matching games, which are discussed in Section 11. The computational simplicity of maximum score, or some other approach that avoids repeated computations of model outcomes, is needed here. Other than my related use of the estimator in Bajari and Fox (2007), this is the first empirical application to a many-to-many matching market where the payoffs to a set of matches are not additively separable across the individual matches. I focus on specialization in the portfolio of matches for a given supplier. Finally, matched firms exchange money, but the prices of the car parts are not in publicly available data. The matching estimator does not require data on the transfers, even though they are present in the economic model being estimated.

9.1 Where is specialization the most important?

I focus on upstream firms or suppliers. This section examines to what extent suppliers benefit from specialization. My production function specification says suppliers may specialize in four areas: parts (in the same component category) for an individual car, parts for cars from a particular brand (Chevrolet, Audi), parts for cars from a particular parent company or assembler (General Motors, Volkswagen) and parts for cars for brands with headquarters on a particular continent. Given my data, I group brands into three continents: Asia (Japan

[56] Alternatively, this is just a standard two-sided, many-to-many matching game where the car parts are one side of the market and the assembler of each car part is a part-specific characteristic.

and Korea), Europe, and North America.[57] The management literature has suggested that supplier specialization may be a key driver of assembler performance (Dyer, 1996, 1997).[58] Here I focus on how specialization can affect the production from a set of car-part relationships centered around a single supplier.[59]

In a slight generalization of the notation from the earlier part of the paper, let C^u be a collection of car parts (a,l), where a is the assembler and l is the car part, in market (component category) m. The production function for upstream firm i is

$$f_\beta\left(\vec{x}\left(i,C^u\right)\right)=\beta_{\text{Cont.}}x^{\text{Continent}}\left(C^u\right)+\beta_{\text{PG}}x^{\text{ParentGroup}}\left(C^u\right)+\beta_{\text{Brand}}x^{\text{Brand}}\left(C^u\right)+\beta_{\text{Car}}x^{\text{Car}}\left(C^u\right). \tag{15}$$

The parameters $\beta_{\text{Cont.}}$, β_{PG}, β_{Brand} and β_{Car} are estimable parameters. The latter three are real numbers; $\beta_{\text{Cont.}}=\pm 1$, as qualitative data like matches cannot identify the scale of production. The match-specific characteristic $x^{\text{ParentGroup}}\left(C^u\right)$ is the Herfindahl-Hirschman Index (HHI) of specialization at the parent group for that supplier. For example, if the supplier produced car parts for only the three American parent groups, the HHI for parent groups would be

$$x^{\text{ParentGroup}}\left(C^u\right)=\left(\frac{\#\,\text{Chrysler parts in}\,C^u}{\#\,\text{total parts in}\,C^u}\right)^2+\left(\frac{\#\,\text{Ford parts in}\,C^u}{\#\,\text{total parts in}\,C^u}\right)^2+\left(\frac{\#\,\text{GM parts in}\,C^u}{\#\,\text{total parts in}\,C^u}\right)^2.$$

More generally,

$$x^{\text{ParentGroup}}\left(C^u\right)=\sum_{a\in D^{\text{PG}}}\left(\frac{\left|L_a^{\text{PG}}\cap C^u\right|}{\left|C^u\right|}\right)^2,$$

where $|C^u|$ is the number of parts in C^u, D^{PG} is the set of parent groups, L_a^{PG} is the set of car parts for parent group a, and $\left|L_a^{\text{PG}}\cap C^u\right|$ is the number of car parts a sources from i, for this component category. This HHI measure will be computed for both the matches seen in the data and for the counterfactual matches in local production maximization inequalities.

The other three characteristics are similar HHI measures. By construction, two parts for the same car also have the same brand, parent group and continent. Two car parts for cars from the same brand are automatically in the same parent group and the brand only has one headquarters, so the parts are from a brand with a headquarters in the same continent as well. Two cars from the same parent group are not necessarily from the same continent, as the Ford-owned brand Mercury is from North America while the Ford-owned brand Volvo is from Europe. As I model each component category as a separate matching market, I do not consider specialization benefits from producing different component categories of parts for the same assembler or car model. I could allow across-category spillovers at the expense of treating each component category as a statistically independent matching market.

[57] As stated, grouping at the continent of headquarters level occurs by the brand and not the parent company. So Opel is grouped into the European continent even though it has been a subsidiary of General Motors since the 1930s. Some brands have headquarters in one continent but produce cars in other continents as well. The continent specialization measure focuses on the continent where the brand has its headquarters.

[58] Novak and Wernerfelt (2007) study co-production of parts by the same supplier for the same car model. They use data on only eight cars and do not discuss the relative specialization at higher levels of organization, such as brand, parent group and headquarters continent.

[59] A few suppliers are owned by assemblers. I ignore this vertical integration decision in my analysis, in part because I lack data on supplier ownership and in part because vertical integration is just an extreme version of specialization, the focus of my investigation. If a supplier sends car parts to only one assembler, that data is recorded and used as an endogenous matching outcome. Vertical integration in automobile manufacturing has been studied previously (Monteverde and Teece, 1982; Novak and Eppinger, 2001; Novak and Stern, 2007b,a).

It should be clear that the four match-specific characteristics in (15) are highly correlated. Just as univariate linear least squares applied to each covariate separately produces different slope coefficients than multivariate linear least squares when the covariates are correlated, a univariate matching theoretic analysis (such as Becker (1973)) on each characteristic separately will be inadequate here. A univariate analysis of say $\beta_{\text{PG}} x^{\text{ParentGroup}}\left(C_i^u\right)$ would just amount to saying that $\beta_{\text{PG}} > 0$ when each supplier does more business with certain parent groups than others. In principle, even this conclusion about the sign of β_{PG} could be wrong if the correlation with the other three characteristics is not considered in estimation. What is even more interesting in this empirical application is to measure the relative importance of each of the four types of specialization: at which level do the returns to specialization occur? This requires formal statistical analysis to estimate $\beta_{\text{Cont.}}$, β_{PG}, β_{Brand} and β_{Car}. Less parametrically, Theorem 3 says we can identify the relative importances of group-specific characteristics by identifying the ratios of the second derivatives of the production function in each characteristic. Here we are parametric in the choice of a production function, but (15) is a parametric implementation of the spirit behind the nonparametric identification result in Theorem 3.[60]

The data come from SupplierBusiness, an analyst firm. There are 1252 suppliers, 14 parent companies, 52 car brands, 392 car models, and 52,492 car parts divided into 593 distinct matching markets, which again are combinations of component categories and continents of assembly of the car. While the data cover different model years, for simplicity I ignore the time dimension and treat each market as clearing simultaneously.[61] The data also lack complete coverage of all car models. The coverage is best in Europe followed by North America; Asia is the worst. I disregard cars manufactured in Asia during estimation, although Asian brands assembled in Europe and and North America are a major focus below. Again, cars assembled in Europe and North America are treated as separate matching markets, although that could be weakened if a particular economic question required it.[62]

I use the maximum score estimator, (14), to compute point estimates, and subsampling to produce confidence intervals. I use local production maximization inequalities with the left and right side matches being of the form $B_1 = \{\langle a,i,l\rangle, \langle b,j,h\rangle\}$ and $B_2 = \{\langle a,j,h\rangle, \langle b,i,l\rangle\}$, where B_1 and B_2 are as in Definition 3. I include two suppliers per inequality, and they exchange one car part each.[63] These exchanges produce more than enough inequalities for parametric estimation. For matching markets with large numbers of car parts, this scheme's combinatorics will produce a computationally intractable number of inequalities. I randomly sample 2000 inequalities for the large matching markets. All theoretically valid inequalities with two different suppliers are sampled with an equal probability, which satisfies Assumption 6.

Table 3 presents point estimates and subsampled confidence intervals for the four HHI specialization measures.[64] We see that all four estimates are positive, meaning as expected specialization on these dimensions

[60]The estimator is semiparametric as I do not impose a parametric structure for the distribution of error terms.

[61]Car models are refreshed around once every five years. A dynamic matching model would be a different paper.

[62]I do not have any data on the suppliers, other than their portfolio of car parts. Geographic location of a supplier's plant would likely be a good predictor of which assembler and assembler plants the supplier provides parts for. However, geographic location is to a large degree an endogenous matching outcome. Supplier plants are often built to service particular assembly plants. With just-in-time production at many assembly sites, supplier factories are built short distances away so parts can be produced and shipped to the assembly site within hours, in many cases. The production function returns to specialization from a supplier's viewpoint thus encapsulate the cost savings from needing to build only one supplier factory for a particular assembler factory.

[63]The local production maximization inequalities used in estimation keep the number of car parts produced by each supplier the same. With strong returns to specialization, it may be more efficient to have fewer but individually larger suppliers. The optimality of supplier size is not imposed as part of the estimator. Not imposing the optimality of supplier size might be an advantage, as other concerns such as capacity constraints and antitrust rules could limit supplier size.

[64]I estimate $\beta_{\text{Cont.}}$ by optimizing the maximum score objective function over the other parameters, first fixing $\beta_{\text{Cont.}} = +1$ and then fixing $\beta_{\text{Cont.}} = -1$. I then take the set of estimates corresponding to the maximum of the two objective function values as the final set of

Table 3: Different types of supplier specialization: production function parameter estimates

	Production function estimates		Sample statistics for HHI Measures	
HHI Measure	Point Estimate	95% CI	Mean	Standard Deviation
Continent	+1	Superconsistent	0.799	0.192
Parent Group	5.71	(4.06, 8.06)	0.457	0.303
Brand	8.82	(0.611, 12.4)	0.341	0.311
Model	91.2	(73.8, 130)	0.256	0.312
# Inequalities	532,939			
% Satisfied	75.3			

increases match production. Sample statistics for the four measures (taken by weighting each supplier, rather than each car part, once) are also listed in order to help explore the economic magnitudes of the point estimates. The production function parameters show that a given level of specialization at the parent group level is 5.7 times more important in production than the same level of specialization at the continent-of-brand-headquarters level. Most specialization benefits occur within firm boundaries rather than across them. At the same time, the standard deviation of parent group specialization HHI, from each supplier's viewpoint, is 0.303, meaning the variation in this parent group specialization across suppliers is high. A naive researcher might be inclined to interpret this dispersion as evidence parent group specialization is unimportant. This would be wrong: the maximum score estimator accounts for the fact that more available matching opportunities occur across firm boundaries rather than within them. Only an estimate of a structural parameter such as the coefficient on parent group tells us the importance of parent group in the production from a set of supplier relationships.

Table 3 also shows that specialization at the brand and model levels is even more important than specialization at the parent group level, although the brand and parent group confidence intervals substantially overlap. The high point estimate of 91.2 for model specialization is, qualitatively, logical: car models of even the same brand may be built in separate plants and some benefits from specialization may occur from saving on the need to have multiple supplier plants for each model. Also, the technological compatibility of car parts occurs mainly at the model level. Notice how the standard deviation of the HHI specialization measure is about the same (around 0.3) for the parent group, brand and model measures, and how the mean HHI declines from parent group to brand to model. Again, naive researchers might use the means to conclude that specialization at the model level is less important or use the standard deviations to conclude that specialization at all three levels are equally important. The structural estimates of the match production function give statistically consistent estimates of the relative importance of the types of specialization in the production functions for supplier

estimates. The estimate of a parameter that can take only two values is superconsistent, so I do not report a confidence interval. The point estimate was always $\beta_{\text{Cont.}} = +1$ (specialization raises production) in initial specifications with smaller numbers of inequalities. In later specifications with more inequalities, I only fix $\beta_{\text{Cont.}} = +1$ in order to reduce the computational time by half.

I use the numerical optimization routine differential evolution, in Mathematica. For differential evolution, I use a population of 200 points and a scaling factor of 0.5. The numerical optimization is run five times with different initial populations of 200 points. I take the point estimates corresponding to the maximum reported objective function value over the five runs. For inference, I use subsample sizes equal to 1/4 of the matching markets. Unfortunately, the literature on subsampling has not produced data dependent guidelines for choosing the subsample size. I use 100 replications (fake artificial datasets) in subsampling. Following the asymptotic theory, I sample from the 593 distinct matching markets (component categories and continents of final assembly).

To give readers an idea about computational time, constructing the inequalities and producing the estimates in Table 3 took 13.6 hours on a single core of a late 2007 vintage desktop computer. The five estimation runs took 2.1 of those hours and the 100 subsampling replications took 8.2 hours. The remainder of the time was spent in data processing. Computational time is approximately linear in the number of inequalities. Using at most 200 inequalities per market, instead of 2000, reduces the total computational time to 1.0 hours, roughly corresponding to a speed level of $2000/200 = 10$ times compared to the previous level of 13.6 hours.

relationships.[65]

9.2 Do suppliers to Asian assemblers have an edge among non-Asian assemblers?

The magazine *Consumer Reports* and other sources routinely record that brands with headquarters in Asia (Japan, Korea) have higher quality automobiles than brands with headquarters in Europe or North America.[66] Toyota is often rated one of the highest-quality brands. The parts supplied to Toyota must be of high quality in order for Toyota to produce quality cars. Liker and Wu (2000) document that suppliers to Japanese-owned brands in the US produce fewer parts requiring reworking or scrapping, for example. Because of this emphasis on quality, the suppliers to Toyota have to undergo a rigorous screening and training program, the Supplier Development Program, before producing a large volume of car parts for Toyota (Langfield-Smith and Greenwood, 1998). Indeed, there is a hierarchy of suppliers, with more trusted Toyota suppliers being allowed to supply more car parts (Kamath and Liker, 1994; Liker and Wu, 2000).

It is possible that being able to supply a higher-quality assembler such as Toyota coincides with a competitive edge for the supplier, allowing them to win business from non-Asian assemblers as well. There are two plausible reasons that a competitive edge might exist. First, Toyota's Supplier Development Program and similar programs at other manufacturers might upgrade the quality of the participating suppliers. This causal quality upgrade from supplying Toyota would allow the suppliers to better compete for business from other assemblers as well, because all assemblers value quality to some degree. Alternatively, there could be a selection story: only a priori high-quality suppliers are allowed to supply high-quality assemblers. In the data, supplying Toyota is just a proxy for being a high-quality firm. I cannot use cross-sectional matching data to answer whether supplying high-quality assemblers causally upgrades the quality of suppliers or whether the Asian assemblers just select high-quality suppliers. Rather, I seek to learn if there is any competitive edge at all: are suppliers to Asian assemblers more likely to sell parts to non-Asian assemblers?[67]

To my knowledge, no previous empirical paper has directly investigated whether matching with one type of partner increases (even if non-causally) the chance of matching with a different type of partner. To investigate the presence of this competitive edge, I generalize the production function for supplier i in (15) to be

$$f_\beta(\vec{x}(i,C^u)) = \beta_{\text{Cont.}} x^{\text{Continent}}(C^u) + \beta_{\text{PG}} x^{\text{ParentGroup}}(C^u) + \beta_{\text{Brand}} x^{\text{Brand}}(C^u) + \beta_{\text{Car}} x^{\text{Car}}(C^u) + \beta_{\text{AsianCont.}} x^{\text{Continent}}(C^u)\, x^{\text{SupplierToAsian}}(C_i^u(A_m)), \quad (16)$$

for assignment A_m. The new term $x^{\text{Continent}}(C^u)\, x^{\text{SupplierToAsian}}(C_i^u(A_m))$ is an interaction between the specialization HHI at the continent level and a measure of supplying Asian assemblers, which I describe below. The total benefit of specialization at the continent level is $\left(\beta_{\text{Cont.}} + \beta_{\text{AsianCont.}} x^{\text{SupplierToAsian}}(C_i^u(A_m))\right) \cdot x^{\text{Continent}}(C^u)$.

[65] There are 532,939 inequalities in the 593 distinct matching markets. Of those, 400,891 or 75.3% are satisfied at the reported point estimates. The fraction of satisfied inequalities is a measure of statistical fit. In the maximum score objective function, an inequality is satisfied if the left side exceeds the right side by 0.0001. This small perturbation to the sum of productions on the right side ensures that inequalities such as $0 > 0$ will not be counted as being satisfied because of some numerical approximation error for zero, resulting in, say, $2.0 \times 10^{-15} > 1.0 \times 10^{-15}$.

[66] Many brands with headquarters in Asia manufacture cars in Europe and North America.

[67] In the non-causal interpretation, one should not use the production function to explore counterfactuals where $x^{\text{SupplierToAsian}}(C_i^u(A))$ changes because the equilibrium assignment A changes. In this interpretation, $x^{\text{SupplierToAsian}}(C_i^u(A))$ is just a marker for supplier quality that cannot be changed. One parallel for the non-causal interpretation is the best linear predictor interpretation for linear regression. The best linear predictor summarizes factual patterns in the data, just like the non-causal interpretation of production functions summarizes facts about sorting patterns.

If the coefficient $\beta_{\text{AsianCont.}}$ on the interaction term is negative, this means that suppliers selling more car parts to brands with headquarters in Asia tend to benefit less from specialization at the continent level. The potential estimate $\beta_{\text{AsianCont.}} < 0$ is compatible with the suppliers to Asian assemblers having a competitive edge and being able to win business from non-Asian suppliers.

I wish to use measures for $x^{\text{SupplierToAsian}}\left(C_i^u(A_m)\right)$ that do not impose any mechanical relationship between $x^{\text{SupplierToAsian}}\left(C_i^u(A_m)\right)$ and the previous HHI specialization measures. In other words, $x^{\text{SupplierToAsian}}\left(C_i^u(A_m)\right)$ should not be a measure of specialization from the supplier's viewpoint. I use two different Asian supplier measures. The first is just an indicator variable equal to 1 when $C_i^u(A_m)$ contains at least one match with a Asian brand. This represents the supplier being able to meet the quality thresholds of Asian assemblers. The second measure is a measure of the market share of the supplier in the "market" (not a formal matching market) for car parts for Asian assemblers. The second measure is

$$x^{\text{SupplierToAsian},2}\left(C_i^u(A_m)\right) = \frac{\#\,\text{Asian assembler parts in}\, C_i^u\,\text{and market}\, m}{\text{total}\,\#\,\text{Asian assembler parts all suppliers in}\, m},$$

where $C_i^u(A_m)$ is a set of car parts for supplier i in a component category market m with equilibrium assignment A_m. $x^{\text{SupplierToAsian},2}\left(C_i^u(A_m)\right)$ is not a measure of whether a supplier is specialized; it is a measure of the fraction of the available Asian contracts the supplier has. I treat each $x^{\text{SupplierToAsian}}\left(C_i^u(A_m)\right)$ measure as an unchanging characteristic of supplier i in market m in a local production maximization inequality. I do not recompute the measure for the counterfactual exchange of partners on the right side of the inequalities, like I do for the HHI measures. Appendix B explores the alternative specification, where $x^{\text{SupplierToAsian}}\left(C^u\right)$ is recomputed with counterfactual matches C^u, in some detail.

Table 4 produces estimates of a supplier's competitive edge, $\beta_{\text{AsianCont.}}$. There are two sets of estimates corresponding to the two measures of being a supplier to Asian assemblers. Look at the first set of estimates, which uses the indicator variable equal to 1 if a supplier has any Asian contracts. The first four rows represent the point estimates of the HHI specialization measures. Compared to Table 3, the lower point estimates for the non-normalized specialization parameters coincide with the normalized parameter, here $\beta_{\text{Cont.}}$, being relatively more important.[68] For suppliers that do not supply Asian assemblers, the return to specialization at the continental level is relatively more important than in the model without the Asian interaction. The coefficient on the interaction with the Asian dummy (supplying any car part to a brand with an Asian headquarters) is -1.09. For firms supplying at least one car part to an Asian assembler, the effect of specialization at the continental level is $+1-1.09$, or in economic magnitude, approximately 0. This is a large effect: suppliers that can meet the quality standards of Asian assemblers can equally compete for business from assemblers with headquarters in Asia, Europe and North America.

Table 4 lists a separate set of point estimates for the market share measure of being an Asian supplier. The point estimate for $\beta_{\text{AsianCont.}}$ is -5.30. In the data, the mean across suppliers of $x^{\text{SupplierToAsian},2}\left(C_i^u(A_m)\right)$ is 0.111 and its standard deviation is 0.204. This implies that a one-standard deviation increase in the Asian market share lowers the gains from continental specialization by $-5.30 \cdot 0.204 = -1.08$, which compares closely to the coefficient of -1.09 in the specification with the Asian dummy. The interpretation is very similar to the specification with the dummy, except for the fact that the point estimates on the other three HHI specialization

[68] With the interaction term included in (16), the normalized specialization measure is more precisely the HHI for continent specialization for those suppliers with zero parts supplied to assemblers with headquarters in Asia, $x^{\text{SupplierToAsian}}\left(C_i^u(A_m)\right) = 0$. Suppliers with no Asian contracts have a 0 value for the interaction term, $x^{\text{Continent}}\left(C^u\right) x^{\text{SupplierToAsian}}\left(C_i^u(A_m)\right)$. 48% of supplier / matching market combinations do not supply any assembler brand with its headquarters in Asia.

Table 4: Supplier competitive edge from supplying Asian assemblers

HHI Measure	Estimate	95% CI	Estimate	95% CI
Continent	+1	Superconsistent	+1	Superconsistent
Parent Group	2.06	(0.751, 2.76)	5.90	(5.64, 8.50)
Brand	5.08	(3.39, 7.51)	9.41	(6.86, 13.5)
Model	40.9	(10.6, 56.7)	101	(76.0, 147)
Continent * Asian Dummy	-1.09	(-1.14, -0.956)		
Continent * Asian %			-5.30	(-6.67, -4.56)
# Inequalities	532,939		532,939	
% Satisfied	0.760		0.758	

measures have about doubled. This means that the relative importance of specialization at the continental level is lower for all firms than in the specification with the dummy.

Combined, the point estimates in Table 4 are consistent with a story where suppliers to brands with headquarters in Asia have a competitive edge. It may be that matching with an Asian assembler gives a supplier a quality upgrade and thus the power to win more business from other assemblers. Or it may be the case that the Asian assemblers select the suppliers with a priori high quality. Regardless, this example shows the usefulness of the matching estimator in determining the relative importance of the characteristics that affect the production from a match. A lot can be learned about structural parameters just by looking at the sorting patterns of supplier-assembler relationships as an equilibrium outcome to a matching game.

10 Asymptotics in the number of matches

Often a researcher has data on one large matching market. For example, Bajari and Fox (2007) study the matching of bidders to items for sale in a spectrum auction with 85 winning bidders and 480 items. The maximum score matching estimator in (13) can be computed using data from one matching market. There can be a lot of information found in the sorting patterns of a single, large matching market. However, the asymptotic distribution calculated as the number of matching markets grows large probably is a poor approximation for the finite-sample distribution of the maximum score estimator using data on one market. This section introduces an alternative asymptotic argument that lets the number of recorded matches in a matching market grow large. To the extent that any asymptotic approximation to the finite-sample distribution is a good one, asymptotics in the number of recorded agents will likely allow more accurate inference in datasets with a small number of large matching markets.

In terms of computer programming, the estimator in this section is the same as (13). However, I introduce what in a formal sense amounts to a new matching model in order to understand the estimator's properties. I prove consistency and derive the asymptotic distribution under this new data generating process. Consequently, I will not rely on any previous results. For the sake of brevity, I focus on semiparametric estimation and inference. I do not reprove the six nonparametric identification theorems under this new model, although they do go through for the new model under appropriate assumptions.

In this section only, I assume the econometrician observes some finite number of recorded agents from an aggregately deterministic market. I let H represent the number of matches in the dataset and I analyze the asymptotics as $H \to \infty$. Readers may already be thinking of the complications in such an asymptotic argument. As new agents are added to a matching market, the matches of existing agents will likely change as the equilibrium assignment will change. Asymptotics where the existing dependent variable data change with the addition of new data are beyond the scope of the current paper.[69] Instead, I introduce the fiction that the real-life matching market with H agents is a subset of some very large matching market. As H gets larger in the asymptotic approximation, the researcher collects more data on a single matching market. Again, the fiction is not to be taken literally: there were only 480 items for sale in the auction studied by Bajari and Fox (2007). When $M = 1$, the asymptotic distribution produced by this exercise is likely a more accurate approximation to the estimator's distribution than using asymptotics in the number of distinct matching markets.

The same matching maximum score estimator (13) can be thought of as consistent as $M \to \infty$ or $H \to \infty$. For inference, a researcher must choose the asymptotic approximation that is best for the data at hand.

10.1 The new matching model in characteristic space

Let there be one very large matching market. This market is so large that there are an uncountable number of agents in it. Therefore, it is necessary to model each agent as existing in characteristic space, rather than using the indices $a \in D$ and $i \in U$ familiar from the previous large-M model. Because the characteristic-space notation can become a bit complex, in this section I restrict attention to an example: one-to-one, two-sided matching with agent-specific characteristic vectors. The consistency and distribution results can be generalized to many-to-many matching and match and group-specific characteristics, at the cost of making the notation denser.

Because I focus on the example of one-to-one matching, I return to the language of men, women and marriage. Let each man have a characteristic vector $\vec{x}^m$ and each woman have a characteristic vector $\vec{x}^w$. The production from a match is $f_\beta(\vec{x}^m, \vec{x}^w)$, where β is a finite vector of parameters. Again for simplicity, let the characteristic vectors have continuous support and let them have a density. The exogenous features of the matching market include $g_m(\vec{x}^m)$, the density of $\vec{x}^m$, as well as $g_w(\vec{x}^w)$. The equilibrium outcome in this market includes a density $g_{m,w}^{\beta,S}(\langle \vec{x}^m, \vec{x}^w \rangle)$, which gives the frequency of the ordered pair $\langle \vec{x}^m, \vec{x}^w \rangle$, representing a marriage between a man with characteristics $\vec{x}^m$ and a woman with characteristics $\vec{x}^w$. The density $g_{m,w}^{\beta,S}(\langle \vec{x}^m, \vec{x}^w \rangle)$ is the analog of the assignment in the previous model. By the physical feasibility of the matching equilibrium, $g_m(\vec{x}_m) = \int g_{m,w}^{\beta,S}(\vec{x}^m, \vec{x}^w)\, d\vec{x}^w$ and $g_w(\vec{x}^w) = \int g_{m,w}^{\beta,S}(\vec{x}^m, \vec{x}^w)\, d\vec{x}^w$. The special notation $g_{m,w}^{\beta,S}(\vec{x}^m, 0)$ gives the density for single men and $g_{m,w}^{\beta,S}(0, \vec{x}^w)$ gives the density for single women. I assume the density $g_{m,w}^{\beta,S}(\langle \vec{x}^m, \vec{x}^w \rangle)$ is deterministically computed as part of the equilibrium; there are no stochastic elements affecting $g_{m,w}^{\beta,S}(\langle \vec{x}^m, \vec{x}^w \rangle)$, although each individual man $\vec{x}^m$'s match is a random draw from the derived conditional density, $g_{w|m}^{\beta,S}(\vec{x}^w \mid \vec{x}^m)$.[70] I index $g_{m,w}^{\beta,S}(\langle \vec{x}^m, \vec{x}^w \rangle)$ by the parameters β from $f_\beta(\vec{x}^m, \vec{x}^w)$ and the distribution S of any unmeasured stochastic terms, because $g_{m,w}^{\beta,S}(\langle \vec{x}^m, \vec{x}^w \rangle)$ is an equilibrium outcome that depends on model parameters, even if it is not a stochastic outcome. See Section 4 for more on stochastic structures in matching games. As previously, let $\beta \in \mathscr{B}$ and $S \in \mathscr{S}$.

As before, the matching maximum score estimator will use local production maximization inequalities. Let

[69] An asymptotic distribution is used as approximation to the finite-sample distribution, which often cannot be derived analytically. Normally, a researcher does not model the real-world process of collecting more observations to construct an asymptotic approximation for inference. If an observation is a past US president, the researcher does not model presidential elections in the year 3000.

[70] The literature has some examples of aggregately deterministic matching models with a continuum of agents. Sattinger (1979) studies a perfect information, one-to-one, two-sided matching game with a continuum of agents. Each agent is distinguished by a scalar type. Choo and Siow (2006) analyze a model with a deterministic aggregate assignment, where each agent has errors $\varepsilon_{\langle a,i \rangle}$ for each match $\langle a,i \rangle$.

there be values $\vec{x}_1^m$, $\vec{x}_2^w$, $\vec{x}_3^m$ and $\vec{x}_4^w$. Then a local production maximization inequality is

$$f_\beta(\vec{x}_1^m,\vec{x}_2^w)+f_\beta(\vec{x}_3^m,\vec{x}_4^w)>f_\beta(\vec{x}_1^m,\vec{x}_4^w)+f_\beta(\vec{x}_3^m,\vec{x}_2^w),$$

which says that the production from two matches exceeds that from an exchange of partners. This inequality is indexed by the two pairs of matched couples, $\langle\vec{x}_1^m,\vec{x}_2^w\rangle$ and $\langle\vec{x}_3^m,\vec{x}_4^w\rangle$. These two marriages can be seen as independent draws from $g_{m,w}^{\beta,S}(\langle\vec{x}^m,\vec{x}^w\rangle)$. If the two matches in an inequality are chosen with probabilities equal to their probabilities in the economy, an inequality itself has a density equal to $g_{m,w}^{\beta,S}(\langle\vec{x}_1^m,\vec{x}_2^w\rangle)\cdot g_{m,w}^{\beta,S}(\langle\vec{x}_3^m,\vec{x}_4^w\rangle)$.

A rank order property will underlie the consistency of the estimator.

Assumption 7. Let $\langle\vec{x}_1^m,\vec{x}_2^w\rangle$ and $\langle\vec{x}_3^m,\vec{x}_4^w\rangle$ be two hypothetical marriages. Assume, for any $\beta\in\mathscr{B}$ and $S\in\mathscr{S}$,

$$f_\beta(\vec{x}_1^m,\vec{x}_2^w)+f_\beta(\vec{x}_3^m,\vec{x}_4^w)>f_\beta(\vec{x}_1^m,\vec{x}_4^w)+f_\beta(\vec{x}_3^m,\vec{x}_2^w)$$

if and only if

$$g_{m,w}^{\beta,S}(\langle\vec{x}_1^m,\vec{x}_2^w\rangle)\cdot g_{m,w}^{\beta,S}(\langle\vec{x}_3^m,\vec{x}_4^w\rangle)>g_{m,w}^{\beta,S}(\langle\vec{x}_1^m,\vec{x}_4^w\rangle)\cdot g_{m,w}^{\beta,S}(\langle\vec{x}_3^m,\vec{x}_2^w\rangle).$$

This rank order property is a condition on the equilibrium sorting pattern. It says that if an exchange of spouses produces a lower sum of production, then the frequency of observing marriages with the same characteristics as the exchange of partners must be lower than observing matches with characteristics that give higher payoffs. Assumption 7 is a natural extension of the more traditional Assumption 1, which considers different assignments to the same matching market. Like Assumption 1, this rank order property is not formally motivated with a model where each consumer has match-specific shocks $\varepsilon_{\langle a,i\rangle}$. However, it is easy to algebraically show that aggregately deterministic, logit-based matching model in Choo and Siow (2006), which has match-specific shocks, satisfies Assumption 7.[71] Section 11 discusses logit-based matching models.

10.2 Estimation with one market

Let the number of matches in the data be H. Because each dataset is finite, the researcher observes a set of matches of the form $\langle a(i),i\rangle$, where i is a woman and $a(i)$ is a convenience function that gives her husband. Aside from the normalizing constant, the maximum score objective function in (13) can be rewritten as

$$\frac{1}{H(H-1)}\sum_{i=1}^{H-1}\sum_{j=i+1}^{H}1\left[f_\beta\left(\vec{x}_{a(i)}^m,\vec{x}_i^w\right)+f_\beta\left(\vec{x}_{a(j)}^m,\vec{x}_j^w\right)>f_\beta\left(\vec{x}_{a(j)}^m,\vec{x}_i^w\right)+f_\beta\left(\vec{x}_{a(i)}^m,\vec{x}_j^w\right)\right]. \quad (17)$$

The double summation forms all unique pairs of two coalitions. Han (1987) introduced a similar estimator for single-agent ordered choice and showed that it was consistent. He called it a maximum rank correlation estimator.

Note that, other than the denominator $\frac{1}{H(H-1)}$, computationally (17) is identical to (13), from the $M\to\infty$ case. The two estimators are identical in terms of computer programming and point estimates. The new estimator has the same practical implementation advantages as the previous estimator. In matching, the distinction between maximum score and maximum rank correlation will be in the asymptotic distribution used for inference.

[71] In Choo and Siow (2006), Assumption 7 is satisfied with match probabilities replacing the density $g_{m,w}^{\beta,S}(\langle\vec{x}_1^m,\vec{x}_3^w\rangle)$, as Choo and Siow allow only characteristics with discrete support in $\vec{x}_m$ and $\vec{x}_w$.

For space reasons, I do not repeat the investigation of nonparametric identification of the production function, but all the results go through. It is easy to adapt results in Han (1987) to show parametric identification for β if $f_\beta(\vec{x}^m, \vec{x}^w)$ is a linear index. Here I assume identification in order to focus on the asymptotic theory. The following assumption is sufficient for the consistency of the semiparametric estimator in (17).

Assumption 8.

1. The production function $f_\beta \in \mathscr{F}$ where $\mathscr{F} = \{f_\beta\}_{\beta \in \mathscr{B}}$ and $\mathscr{B} \subseteq \mathbb{R}^{|\beta|}$, $|\beta| < \infty$. $\mathscr{B}$ is compact.
2. Identification: Let $\beta^0 \in \mathscr{B}$ and $S^0 \in \mathscr{S}$ be the true parameters that generate the data. For any $\beta^1 \neq \beta^0$, $\beta_1 \in \mathscr{B}$, and for any $S^1 \in \mathscr{S}$, there exists a set of pairs of match characteristics $\mathscr{X}$ with positive probability such that $g_{m,w}^{\beta^0,S^0}(\langle \vec{x}_1^m, \vec{x}_2^w \rangle) \cdot g_{m,w}^{\beta^0,S^0}(\langle \vec{x}_3^m, \vec{x}_4^w \rangle) > g_{m,w}^{\beta^0,S^0}(\langle \vec{x}_1^m, \vec{x}_4^w \rangle) \cdot g_{m,w}^{\beta^0,S^0}(\langle \vec{x}_3^m, \vec{x}_2^w \rangle)$ while $g_{m,w}^{\beta^1,S^1}(\langle \vec{x}_1^m, \vec{x}_2^w \rangle) \cdot g_{m,w}^{\beta^1,S^1}(\langle \vec{x}_3^m, \vec{x}_4^w \rangle) < g_{m,w}^{\beta^1,S^1}(\langle \vec{x}_1^m, \vec{x}_4^w \rangle) \cdot g_{m,w}^{\beta^1,S^1}(\langle \vec{x}_3^m, \vec{x}_2^w \rangle)$ for any $(\vec{x}_1^m, \vec{x}_2^w, \vec{x}_3^m, \vec{x}_4^w) \in \mathscr{X}$.
3. All characteristics are agent-specific and continuous.
4. Each match $\langle \vec{x}^m, \vec{x}^w \rangle$ is an independent draw from $g_{m,w}^{\beta^0,S^0}(\langle \vec{x}^m, \vec{x}^w \rangle)$.

As the maximum score and maximum rank correlation literatures show, for the linear index functional forms of $f_\beta(\vec{x}^m, \vec{x}^w)$, only one characteristic of $\vec{x}^m$ needs to have continuous support. In matching, continuous support for all elements of $\vec{x}^m$ and $\vec{x}^w$ simplifies the statement of the rank order property, Assumption 7.[72]

The estimator is consistent as more data are collected on agents from an underlying very large matching market.

Theorem 8. *As $H \to \infty$, any parameter vector $\hat{\beta}_H \in \mathscr{B}$ that maximizes the maximum rank correlation objective function* (17) *is a consistent estimator for $\beta^0 \in \mathscr{B}$, the parameter vector in the data generating process.*

Sherman (1993) shows that the maximum rank correlation estimator is $\sqrt{H}$-consistent and asymptotically normal. The objective function at a given β is a U-statistic of second order. As H grows, the terms in the double summation grow proportionately to H^2. Intuitively, the inner summation acts like a smoother without requiring an explicit kernel and bandwidth. The derivation relies on a general set of results for the asymptotic distribution of U-processes in Sherman (1994).

Because of the need to impose scale normalizations, we need a specific functional form for the production function. Let $f_\beta(\vec{x}^m, \vec{x}^w) = \sum_{k=1}^{K_m} \sum_{l=1}^{K_w} \beta_{k,l} \vec{x}_k^m \vec{x}_l^w$, where $\vec{x}^m = \left(\vec{x}_1^m, \ldots, \vec{x}_{K_m}^m\right)$ and $\vec{x}^w = \left(\vec{x}_1^w, \ldots, \vec{x}_{K_w}^w\right)$. Normalize $\beta_{1,1} = \pm 1$ and let $\tilde{\beta}$ be a vector of the $K_m K_w - 1$ other elements of β. The kernel of this maximum rank correlation estimator is defined to be

$$\tau\left(\vec{x}_1^m, \vec{x}_2^w, \tilde{\beta}; \beta_{1,1}\right) = E_{\vec{x}_3^m, \vec{x}_4^w}\left[1\left[\sum_{k=1}^{K_m}\sum_{l=1}^{K_w} \beta_{k,l}\vec{x}_{1,k}^m \vec{x}_{2,l}^w + \sum_{k=1}^{K_m}\sum_{l=1}^{K_w} \beta_{k,l}\vec{x}_{3,k}^m \vec{x}_{4,l}^w > \sum_{k=1}^{K_m}\sum_{l=1}^{K_w} \beta_{k,l}\vec{x}_{1,k}^m \vec{x}_{4,l}^w + \sum_{k=1}^{K_m}\sum_{l=1}^{K_w} \beta_{k,l}\vec{x}_{3,k}^m \vec{x}_{2,l}^w\right]\right].$$

If the smoothness conditions on the kernel τ in Assumption A4 in Sherman (1993) hold, then Theorem 4 in Sherman shows

$$\sqrt{H}\left(\tilde{\beta}_H - \tilde{\beta}_0\right) \to N\left(0, V^{-1}\Delta V^{-1}\right),$$

[72] Otherwise the statement would involve a probability measure over inequalities, rather than a density evaluated at a point.

where $2V = E_{\vec{x}_1^m, \vec{x}_2^w} \nabla_2 \tau \left(\vec{x}_1^m, \vec{x}_2^w, \tilde{\beta}^0; \beta_{1,1}^0 \right)$ and $\Delta = \left(E_{\vec{x}_1^m, \vec{x}_2^w} \nabla_1 \tau \left(\vec{x}_1^m, \vec{x}_2^w, \tilde{\beta}^0; \beta_{1,1}^0 \right) \right) \left(E_{\vec{x}_1^m, \vec{x}_2^w} \nabla_1 \tau \left(\vec{x}_1^m, \vec{x}_2^w, \tilde{\beta}^0; \beta_{1,1}^0 \right) \right)'$, where ∇_1 is the gradient operator and ∇_2 creates the Hessian, with all derivatives taken with respect to the argument $\tilde{\beta}$.[73] The elements of $2V$ and Δ can be consistently estimated, so this distribution can be used for asymptotically valid inference.[74]

The estimator as $H \to \infty$ converges at the rate $\sqrt{H}$ while the estimator as $M \to \infty$ converges at the rate $\sqrt[3]{M}$. This should not be taken to mean that the $H \to \infty$ estimator makes more use of the data, because more information is collected when one samples a completely new market than when one samples a new match within a given, large market.

11 Literature comparisons

There are other, more parametric, matching estimators for both matching games where money can be exchanged (like in this paper) and matching games where money is not used. I review these two literatures separately.

11.1 Other estimators for matching games with transfers

Dagsvik (2000), Choo and Siow (2006) and Weiss (2007) introduce logit matching models one-to-one (marriage), two-sided matching games with transferable utility.[75] These estimators are computationally simple because they exploit the mathematics behind the aggregate data multinomial logit model (McFadden, 1973; Berry, 1994). However, to my knowledge, these estimators have not been expanded to the case of many-to-many matching. The discussion about the combinatorics in the number of necessary conditions at the end of Section 3.3 suggests that any extension of the logit estimators to many-to-many matching will be difficult.

There is a subtler distinction between my matching model and the logit-based matching models that focuses on the timing of when equilibrium transfers $t_{\langle a,i \rangle}$ are computed, and what the timing implies about the models' abilities to give positive probability to any feasible assignment: $\Pr(A \mid X) > 0$ for any feasible A. Focus on Choo and Siow (2006). In their paper, the data generating process is not (11): there is no linear program. Rather, men and women are divided into a finite number of classes. Each man has error terms for women of a certain class, but not each woman individually. Likewise, each woman has error terms for each male type. Then prices are set to equate the supply and demand of men and women for each type of marriage. Therefore, this model is deterministic at the aggregate level: the iid logit shocks average out because an infinite number of each type of man and each type of woman are assumed to exist. In effect, each agent plays the equivalent of a mixed strategy from a Nash game, where the randomness across matching partners is governed by the parametric logit distribution.

This type of model may be appropriate to apply to the US marriage market, where there are a large number of agents and a coarse set of demographics to distinguish them. However, the model will not be compatible with typical assignment data if applied to a dataset with a smaller number of men and women. Say there are

[73] A previous draft computed these derivatives for a case with match-specific characteristics, which more closely follows the derivative calculations in Sherman (1993).

[74] A population mean such as $E_{\vec{x}_1^m, \vec{x}_2^w} \nabla_2 \tau \left(\vec{x}_1^m, \vec{x}_2^w, \tilde{\beta}^0; \beta_{1,1}^0 \right)$ can be estimated by a sample average evaluated at the point estimates $\hat{\beta}_H$ instead of the true parameter values β^0. First and second derivatives with respect to $\tilde{\beta}$ can be computed using finite differences (perturb elements of the estimated $\hat{\tilde{\beta}}_H$ by $e > 0$) or, given a functional form for $f_\beta(\vec{x}^m, \vec{x}^w)$, by symbolic calculation.

[75] Dagsvik (2000) actually analyzes a more general model of matching in contract space; transferable utility is a special case.

only two men, a and b, and two women, i and j, in the market. Prices are set before the logit shocks are realized and after the two men make unilateral decisions to marry the two women. If $\Pr(\langle a,i\rangle)$ is the probability a marries i at the equilibrium prices, $\Pr(\langle a,i\rangle)\cdot\Pr(\langle b,i\rangle)$ is the probability that both men marry woman i. A woman cannot marry two men in most countries, so this prediction of the model will be counterfactual and the model will be rejected by the data. By contrast, the data generation process in (11) has the error terms enter a linear programming problem that ensures, for every realization of the errors for all agents, that the resulting assignment is physically feasible.[76]

11.2 Estimators for games without transfers

Recently, Boyd, Lankford, Loeb and Wyckoff (2003), Sørensen (2007), and Gordon and Knight (2006) estimate Gale and Shapley (1962) matching games, which do not use transfers as part of the equilibrium concept.[77] Whether a researcher should estimate a game with or without endogenous transfers depends on the market in question. Games with endogenous transfers often give different equilibrium predictions than games without transfers. Nonparametric identification has not previously been studied for any matching games. Their empirical applications study many-to-many matching, but all the papers rule out preferences over sets of partners; rather utilities are defined over only singleton matches.[78]

The main drawback of these approaches is computational. For a given value of parameters, these approaches use simulation to evaluate a likelihood or moments-based objective function. In Boyd et al. and Gordon and Knight, a nested equilibrium computation produces the model's prediction for the data for each draw of the error terms from some parametric distribution. Sørensen treats the unobservables as nuisance parameters and samples them from a parametric likelihood that enforces sufficient conditions for the data to be the equilibrium to a matching game.

Several simplifications must be imposed that the current paper weakens in the class of games with transfers. First, a researcher must take a stand and on all model components needed to compute an equilibrium. For example, quotas, the number of matches a firm can make, are typically unobserved. Boyd et al. and Sørensen assume that a firm can make only as many matches as are observed in the data. By contrast, a necessary conditions approach does not force one to consider inequalities that raise the number of matches, which preserves consistency without violating unobserved quotas.

The definition of a matching market may be unclear to the econometrician. Boyd et al. and Sørensen limit the size of markets for computational reasons because an equilibrium to a matching game must be calculated or enforced for every trial parameter vector and realization of the error terms. Consistency is broken if the market is defined too narrowly. By contrast, the current paper uses necessary conditions. A market can be defined conservatively for robustness without damaging the validity of the necessary conditions. A researcher can use

[76] A related distinction between the two models lies in how prices are formed. In Choo and Siow (2006), prices are only functions of the discrete type of one's marriage partner. Prices are formed before the logit shocks are realized. By contrast, in (11) a full matching game is solved for each realization of the error terms. In this model, the distinction between error terms $\varepsilon_{\langle a,i\rangle}$ and the characteristics in X is only whether the exogenous variable in question is recorded in the data or not. Equilibrium prices given by the model, even if not recorded in the data, will be a function of the error terms $\varepsilon_{\langle a,i\rangle}$ for all potential matches $\langle a,i\rangle$.

[77] Hitsch, Hortaçsu and Ariely (2008) use data on both desired and rejected matches to estimate preferences without using an equilibrium model. They then find that a calibrated model's prediction fits observed matching behavior. Echenique (2006) examines testable restrictions on the lattice of equilibrium assignments of the Gale and Shapley (1962) model.

[78] A similar assumption for matching games with transfers would be that production functions are additively separable across multiple matches: $f(X(i,\{a,b\})) = f(\vec{x}(i,\{a\})) + f(\vec{x}(i,\{b\}))$. This would rule out the study of spectrum auctions in Bajari and Fox (2007) and the automotive supplier specialization empirical example in this paper.

a constant number of inequalities from each market, so there is no need to limit the size of a matching market for computational reasons.

Matching games without transfers have a lattice of multiple equilibrium assignments. Nested solutions methods require auxiliary assumptions to resolve the multiplicity problem. Sørensen and Gordon and Knight restrict preferences to generate a unique equilibrium. Boyd et al. impose an auxiliary equilibrium selection rule. By contrast, Section 4.5 argues that the maximum score necessary conditions approach can be valid in the presence of multiple equilibrium assignments.[79]

12 Conclusions

This paper discusses identification and estimation of production functions in matching games first studied by Koopmans and Beckmann (1957), Shapley and Shubik (1972) and Becker (1973). These matching games allow endogenous transfers that are additively separable in payoffs. Under a pairwise stable equilibrium, production functions must satisfy inequalities that I call local production maximization: if an exchange of one downstream firm per upstream firm produces a higher production level, than it cannot be individually rational for some agent. For some simple matching games this condition is related to social efficiency, but for general many-to-many matching games it is not.

It is not obvious what types of economic parameters are identified from data on only who matches with whom. The identification theorems cover both cardinal and ordinal properties of production functions. The cardinal results generalize the work of Becker (1973) to the case of each agent having a vector of types, many-to-many matching as well as production functions where pairs of inputs are not complements over their entire supports. The ordinal results extend the single-agent work of Matzkin (1993) to matching games, where agents cannot unilaterally choose partners and so identification requires working with the equilibrium structure of the game.

I introduce a semiparametric estimator for matching games. The matching maximum score estimator has computational advantages that eliminate three aspects of a computational curse of dimensionality in the size of the market. First, the estimator avoids the need to nest an equilibrium computation in the statistical objective function. Second, the maximum score estimator does not require numerical integrals over match-specific error terms. Third, inequalities need to be included only with some positive probability, which is extremely important given the combinatorics of necessary conditions for pairwise stability in many-to-many matching. Also, evaluating the objective function involves only calculating match production levels and checking inequalities. Numerical optimization can use global optimization routines.

There are also data advantages. The estimator uses data on only observed matches and agent characteristics. It does not require the often unavailable data on endogenous transfers, quotas and production levels. For example, the empirical application to automotive suppliers and assemblers is typical in that the parties exchange transfers but those transfers are not shared with researchers. Also, the estimator does not require any first-stage nonparametric estimates of assignment probabilities as a function of all exogenous characteristic data.

[79]Further, game with transfers with an outcome in the core (say a marriage game) have unique equilibrium assignments with probability 1, without resorting to preference restrictions or equilibrium selection rules.

A Proofs

A.1 Lemma 1: Pairwise stability implies local production maximization

Substitute $\tilde{t}_{\langle b,i\rangle}$ into (3) and cancel the transfers $\sum_{c\in C_i^u(A)\setminus\{a\}} t_{\langle c,i\rangle}$ to give

$$r^u\left(\vec{x}\left(i,C_i^u(A)\right)\right)+t_{\langle a,i\rangle}\geq r^u\left(\vec{x}\left(i,\left(C_i^u(A)\setminus\{a\}\right)\cup\{b\}\right)\right)+r^d\left(\vec{x}\left(i,\left(C_i^u(A)\setminus\{a\}\right)\cup\{b\}\right)\right)-\left(r^d\left(\vec{x}\left(j,C_j^u(A)\right)\right)-t_{\langle b,j\rangle}\right).$$

Call this no-deviation inequality $\text{nd}\left(\langle a,i\rangle,\langle b,j\rangle,A,t_{\langle a,i\rangle},t_{\langle b,j\rangle}\right)$. If $\pi\langle a,i\rangle$ is the new downstream partner of i in the permutation, let $j(\pi\langle a,i\rangle)$ be a function that gives the original partner of $\pi\langle a,i\rangle$, in B_1. So $\langle\pi\langle a,i\rangle,j(\pi\langle a,i\rangle)\rangle\in B_1$. Now form $\sum_{\langle a,i\rangle\in B_1}\text{nd}\left(\langle a,i\rangle,\langle\pi\langle a,i\rangle,j(\pi\langle a,i\rangle)\rangle,A,t_{\langle a,i\rangle},t_{\langle\pi\langle a,i\rangle,j(\pi\langle a,i\rangle)\rangle}\right)$. This gives

$$\begin{aligned}\sum_{\langle a,i\rangle\in B_1} r^u\left(\vec{x}\left(i,C_i^u(A)\right)\right)+\sum_{\langle a,i\rangle\in B_1} t_{\langle a,i\rangle}\geq\\ \sum_{\langle a,i\rangle\in B_1} r^u\left(\vec{x}\left(i,\left(C_i^u(A)\setminus\{a\}\right)\cup\{\pi\langle a,i\rangle\}\right)\right)+\\ \sum_{\langle a,i\rangle\in B_1}\left\{r^d\left(\vec{x}\left(i,\left(C_i^u(A)\setminus\{a\}\right)\cup\{\pi\langle a,i\rangle\}\right)\right)-\left(r^d\left(\vec{x}\left(j(\pi\langle a,i\rangle),C^u_{j(\pi\langle a,i\rangle)}(A)\right)\right)-t_{\langle\pi\langle a,i\rangle,j(\pi\langle a,i\rangle)\rangle}\right)\right\}.\end{aligned}$$

By the definition of a permutation, each $t_{\langle a,i\rangle}$ for $\langle a,i\rangle\in B_1$ appears on the both left and right sides. The transfers cancel. Similarly, each equilibrium $r^d\left(\vec{x}\left(j(\pi\langle a,i\rangle),C^u_{j(\pi\langle a,i\rangle)}(A)\right)\right)$ appears on the right side with a negative sign and each deviation $r^d\left(\vec{x}\left(i,\left(C_i^u(A)\setminus\{a\}\right)\cup\{\pi\langle a,i\rangle\}\right)\right)$ appears on the right side with a positive sign. Moving $\sum_{\langle a,i\rangle\in B_1} r^d\left(\vec{x}\left(j(\pi\langle a,i\rangle),C^u_{j(\pi\langle a,i\rangle)}(A)\right)\right)$ to the left side and substituting the definition of a production function, Definition 1, gives the local production maximization inequality in the lemma.

A.2 Lemma 2: Sufficient condition for the rank order property

Inspect the formula

$$\Pr(A_1\mid Q,X;f,S)=\int_{-\infty}^{\infty}\int_{-\infty}^{\sum_{\langle a,i\rangle\in A_1} f\left(\vec{x}\left(i,C_i^u(A_1)\right)\right)+\psi_{A_1}-\sum_{\langle a,i\rangle\in A_2} f\left(\vec{x}\left(i,C_i^u(A_2)\right)\right)}\cdots\int_{-\infty}^{\sum_{\langle a,i\rangle\in A_1} f\left(\vec{x}\left(i,C_i^u(A_1)\right)\right)+\psi_{A_1}-\sum_{\langle a,i\rangle\in A_J} f\left(\vec{x}\left(i,C_i^u(A_J)\right)\right)} S(\psi)\,d\psi,$$

If there are J physically feasible assignments $A_1,\ldots A_J$, $\Pr(A_2\mid Q,X;f,S)$ is just the same function, with $\sum_{\langle a,i\rangle\in A_2} f\left(\vec{x}\left(i,C_i^u(A_2)\right)\right)+\psi_{A_2}$ replacing $\sum_{\langle a,i\rangle\in A_1} f\left(\vec{x}\left(i,C_i^u(A_1)\right)\right)+\psi_{A_1}$ in the upper limits of the integrals and $\sum_{\langle a,i\rangle\in A_1} f\left(\vec{x}\left(i,C_i^u(A_1)\right)\right)$ replacing $\sum_{\langle a,i\rangle\in A_2} f\left(\vec{x}\left(i,C_i^u(A_2)\right)\right)$ in the first upper limit. $\Pr(A_1\mid Q,X;f,S)$ is the same function as $\Pr(A_2\mid Q,X;f,S)$ by the exchangeability of $S(\psi)$, except for the A_1 and A_2 arguments. Because $\Pr(A_1\mid Q,X;f,S)$ is increasing in $\sum_{\langle a,i\rangle\in A_1} f\left(\vec{x}\left(i,C_i^u(A_1)\right)\right)$ and decreasing in $\sum_{\langle a,i\rangle\in A_2} f\left(\vec{x}\left(i,C_i^u(A_2)\right)\right)$, $\Pr(A_1\mid Q,X;f,S)>\Pr(A_2\mid Q,X;f,S)$. Reversing the order of these arguments proves the other direction.

A.3 Theorem 1: Cardinal identification with firm-specific covariates

A.3.1 Part 1

The vector $\vec{x}$ is given in the statement of the theorem. To avoid confusion of the point $\vec{x}$ with the function $\vec{x}\left(i,C_i^u\right)$, I relabel the vector $\vec{x}$ as $\vec{y}$ inside this proof. I will focus on the case where $\frac{\partial f^0(\vec{y})}{\partial x_1\partial x_2}>0$. The proof for the case where $\frac{\partial f(\vec{y})}{\partial x_1\partial x_2}<0$ is very similar.

For an arbitrary $f^1 \in \mathscr{F}, f^1 \neq f^0$ where $\frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} < 0$, by Definition 4 I must show that there does not exist S^1 corresponding to f^1 where $\Pr\left(A \mid X; f^0, S^0\right) = \Pr\left(A \mid X; f^1, S^1\right)$ for all (A, X) except perhaps a set of X of probability 0. By the key Assumption 1, a sufficient condition involves showing that there exists a continuum of market characteristics X with positive probability and a corresponding matching situation where f^0 and f^1 give different implications for a local production maximization inequality, of the form in Definition 3. At each of these markets X, there will be a particular assignment A_1 and another assignment A_2 where, by Assumption 1, $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$. Therefore, the conditions of Definition 4 will be satisfied.

Let me explain the steps of the proof. First, I derive an appropriate local production maximization inequality and show that the inequality will be reversed if the production function is f^1 instead of f^0. Second, I show how I can embed the characteristics in the local production maximization inequality into a matching market with characteristics X. Third, I show that I can locally vary all the characteristics in X to find a continuum of markets X with the property of $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$.

First, I explore deriving a local production maximization inequality. Let $e_k = (0, \ldots, 0, 1, 0, \ldots, 0)$, where the 1 is in the kth position. Without loss of generality, let x_1 be the first position of $\vec{x}$ and let x_2 be the second position. The definition of a cross-partial derivative is the limit of the middle difference quotient:

$$\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2} = \lim_{h \to 0} \frac{f(\vec{y} + he_1 + he_2) - f(\vec{y} + he_1) - f(\vec{y} + he_2) + f(\vec{y})}{h^2}. \tag{18}$$

Let $\nu > 0$ be given. By the definition of a limit, we can find $h > 0$ such that

$$\left| \frac{\partial f(\vec{x})}{\partial x_1 \partial x_2} - \frac{f(\vec{y} + he_1 + he_2) - f(\vec{y} + he_1) - f(\vec{y} + he_2) + f(\vec{y})}{h^2} \right| < \nu.$$

As $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} > 0$, there will be a $h^0 > 0$ such that the numerator of the middle difference quotient at $f = f^0$ is positive, or

$$f^0\left(\vec{y} + h^0 e_1 + h^0 e_2\right) - f^0\left(\vec{y} + h^0 e_1\right) - f^0\left(\vec{y} + h^0 e_2\right) + f^0(\vec{y}) > 0,$$

or

$$f^0\left(\vec{y} + h^0 e_1 + h^0 e_2\right) + f^0(\vec{y}) > f^0\left(\vec{y} + h^0 e_1\right) + f^0\left(\vec{y} + h^0 e_2\right). \tag{19}$$

As $\frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} < 0$, there exists $h^1 > 0$ where

$$f^1(\vec{y} + he_1 + he_2) + f^1(\vec{y}) < f^1(\vec{y} + he_1) + f^1(\vec{y} + he_2). \tag{20}$$

The argument for f^1 is symmetric to the argument for f^0 and is omitted. Set $h = \min\left\{h^0, h^1\right\}$.

Now let me argue that (19), and by a similar argument (20), is a local production maximization inequality: it satisfies Definition 3. To do this I need to form B_1 and B_2, as in the definition, and show how a hypothetical swap of downstream firm partners could produce (19). Let $B_1 = \{\langle a, i\rangle, \langle b, j\rangle\}$ and $B_2 = \{\langle a, j\rangle, \langle b, i\rangle\}$, where these indices refer to arbitrary firms I am creating to show (19) satisfies Definition 3. Also, let there be some C_i^u and C_j^u of sufficient size to reproduce the number of non-empty ($\emptyset$, representing unfilled matches) elements of $\vec{y}$. We require $a \in C_i^u$, $a \notin C_j^u$, $b \in C_j^u$, and $b \notin C_i^u$. Then define $\vec{x}\left(i, C_i^u\right) = \vec{y} + he_1 + he_2$, $\vec{x}\left(j, C_j^u\right) = \vec{y}$, $\vec{x}\left(i, \left(C_i^u \setminus \{a\}\right) \cup \{b\}\right) =$

$\vec{y}+he_1$ and $\vec{x}\left(j,\left(C_j^u\backslash\{b\}\right)\cup\{a\}\right)=\vec{y}+he_2$. With $\pi\langle a,i\rangle=b$ and $\pi\langle b,j\rangle=a$, inspection shows (19) satisfies Definition 3.

Further, it is important to show that this exchange of partners can be accomplished with firm-specific characteristics, as that is a maintained hypothesis in the theorem being proved. The theorem requires that x_1 and x_2 be from different firms. As only one upstream firm's characteristics enter each production function, it is without loss of generality to say that x_2 is a characteristic of a downstream firm. To complete the argument that (19) satisfies Definition 3 for the case of firm-specific characteristics, let downstream firms a and b have the same baseline characteristics, except that firm a has he_2 more of characteristic x_2 than firm b. Using the notation $\vec{x}(i,C^u)=\text{cat}\left(\left(x_{i,1}^u,\ldots,x_{i,K^u}^u\right),\left(x_{d_1,1}^d,\ldots,x_{d_1,K^d}^d\right),\ldots,\left(x_{d_n,1}^d,\ldots,x_{d_n,K^d}^d\right)\right)$ for $C^u=\{d_1,\ldots,d_n\}$, then $x_{a,2}^d-he_2=x_{b,2}^d$, where now a is one of the firms in C_i^u. On the left of (19), the match $\langle a,i\rangle$ puts the downstream firm a with he_2 more x_2 in either a direct partnership with an upstream firm i with h_1e_1 more x_1 than upstream firm j or an indirect partnership with another downstream firm, say $c_i\in C_i^u$, with h_1e_1 more x_1 than the corresponding downstream firm c_j in C_j^u. In notation, either $x_{i,1}^u-he_1=x_{j,1}^u$ or $x_{c_i,1}^d-he_1=x_{c_j,1}^d$.[80]

The matches $\langle a,j\rangle$ and $\langle b,i\rangle$ form on the right side of (19). The firm a with h_2e_2 more of x_2 is transferred from the set of matches involving i with h_1e_1 more x_1 to the set of matches involving j without any more x_1. Likewise, the downstream firm b without any more x_2 matches to i and its downstream firm partners, which together have h_1e_1 more x_1 than the matches involving j. The important requirement that is satisfied is that each move switches the characteristics of only the firm that is actually switching. Therefore, (19) satisfies Definition 3 for the case of firm-specific characteristics.

The second step of the proof is that I will argue that I can embed B_1 and B_2 in an entire matching market. Let B_3 be a larger set of matches that includes the matches corresponding to the downstream firms in $C_i^u\backslash\{a\}$ and $C_j^u\backslash\{b\}$. The exact choice of B_3 plays no role in the proof, other than to ensure $C_i^u\backslash\{a\}$ and $C_j^u\backslash\{b\}$ are large enough given the number of non-empty elements (representing filled quota slots) in $\vec{y}$. Then set $A_1=B_1\cup B_3$ and $A_2=B_2\cup B_3$. Likewise, let there be some collection X of characteristics for all potential matches, including but not limited to the matches in $A_1\cup A_2$. The choice of X plays no role in the proof, except that X must contain $\vec{x}\left(i,C_i^u\right)=\vec{y}+he_1+he_2$, $\vec{x}\left(j,C_j^u\right)=\vec{y}$, $\vec{x}\left(i,\left(C_i^u\backslash\{a\}\right)\cup\{b\}\right)=\vec{y}+he_1$ and $\vec{x}\left(j,\left(C_j^u\backslash\{b\}\right)\cup\{a\}\right)=\vec{y}+he_2$.

Assumption 1 and (19) imply $\Pr\left(A_1\mid X;f^0,S^0\right)>\Pr\left(A_2\mid X;f^0,S^0\right)$ while Assumption 1 and (20) imply $\Pr\left(A_1\mid X;f^1,S^1\right)<\Pr\left(A_2\mid X;f^1,S^1\right)$ for any $S_1\in\mathscr{S}$.

I now move to the third stage of the proof. Definition 4 requires a continuum of markets with characteristics X with this property. The definition of a continuous function $g(z)$ states that $g^{-1}(V)$ is an open neighborhood of z whenever V is an open neighborhood of $g(x)$. $\mathscr{F}$ contains only continuous functions, so

$$g\left(\vec{z}_1,\vec{z}_2,\vec{z}_3,\vec{z}_4\mid f\right)\equiv f\left(\vec{z}_1\right)+f\left(\vec{z}_2\right)-f\left(\vec{z}_3\right)-f\left(\vec{z}_4\right)$$

is also a continuous function in the tuple $\left(\vec{z}_1,\vec{z}_2,\vec{z}_3,\vec{z}_4\right)$, for any $f\in\mathscr{F}$. Because the inequalities are strict, there is an open neighborhood V^0 around

$$g\left(\vec{x}\left(i,C_i^u\right),\vec{x}\left(j,C_j^u\right),\vec{x}\left(i,\left(C_i^u\backslash\{a\}\right)\cup\{b\}\right),\vec{x}\left(j,\left(C_j^u\backslash\{b\}\right)\cup\{a\}\right);f^0\right)$$

[80] Note some abuses of notation: if x_2 is the second characteristic of $\vec{x}\left(i,C_i^u\right)$, I then say it is also the second characteristic of $\vec{x}_a^d$, $x_{a,2}^d$. This is done for clarity: to keep "2" referring to the same variable whether I am referring to it as an element of the entire vector of production function arguments $\vec{x}\left(i,C_i^u\right)$ or as an element of the vector of characteristics of firm a, $\vec{x}_a^d$.

where $p > 0$ for $p \in V^0$ and there is an open neighborhood V^1 around

$$g\left(\vec{x}\left(i, C_i^u\right), \vec{x}\left(j, C_j^u\right), \vec{x}\left(i, \left(C_i^u \setminus \{a\}\right) \cup \{b\}\right), \vec{x}\left(j, \left(C_j^u \setminus \{b\}\right) \cup \{a\}\right); f^1\right) < 0$$

where $p < 0$ for $p \in V^1$. Because the tuple $v = \left(\vec{x}\left(i, C_i^u\right), \vec{x}\left(j, C_j^u\right), \vec{x}\left(i, \left(C_i^u \setminus \{a\}\right) \cup \{b\}\right), \vec{x}\left(j, \left(C_j^u \setminus \{b\}\right) \cup \{a\}\right)\right)$ is the same in both the f^0 and f^1 constructions, $W = g^{-1}\left(V^0; f^0\right) \cap g^{-1}\left(V^1; f^1\right) \neq \emptyset$ and is itself an open neighborhood of v, as both $g^{-1}\left(V^0; f^0\right)$ and $g^{-1}\left(V^1; f^1\right)$ are open sets by the definition of continuity and topologies are closed under finite intersections.

Divide any market characteristics X into X_1 and X_2, where X_1 is the set comprised of $\vec{x}\left(i, C_i^u\right)$, $\vec{x}\left(j, C_j^u\right)$, $\vec{x}\left(i, \left(C_i^u \setminus \{a\}\right) \cup \{b\}\right)$ and $\vec{x}\left(j, \left(C_j^u \setminus \{b\}\right) \cup \{a\}\right)$, where here these terms represent the characteristics of the corresponding four matches given an arbitrary X (the function arguments $(\vec{z}_1, \vec{z}_2, \vec{z}_3, \vec{z}_4)$ above) rather than terms involving the $\vec{y}$ from the statement of the theorem. Also, $X_2 = X \setminus X_1$, so X is a one-to-one change of variables (or just notation) of (X_1, X_2). By Assumption 2, X has support equal to the product of the marginal supports of its elements. Therefore, the probability that (X_1, X_2) lies in $W \times \mathbb{R}^s$ is strictly positive, where s is the number of elements of X_2. By construction, all market characteristics (X_1, X_2) in $W \times \mathbb{R}^s$ satisfy $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$.

A.3.2 Part 2

We are given a point $\vec{y}$ (relabeled from $\vec{x}$ in the statement of the theorem) and there is an arbitrary $f^1 \in \mathscr{F}$ where $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_3 \partial x_4} \neq \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^1(\vec{y})}{\partial x_3 \partial x_4}$. The goal in broad generality is the same as Part 1: show there exists a continuum of X and two assignments A_1 and A_2 where $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$. The proof is more challenging than the proof of Part 1 because now we are trying to identify the value of some feature of f^0, here $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_3 \partial x_4}$, rather than just the sign of some feature, as before.

I will show that the term $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3}$ is identified, where x_1 is the same characteristic in the numerator and the denominator. Then, by Young's / Clairaut's / Schwarz's theorem, arbitrary ratios $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_3 \partial x_4}$ can be identified by comparing, say, $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3}$ to $\frac{\partial f^0(\vec{y})}{\partial x_3 \partial x_4} / \frac{\partial f^0(\vec{y})}{\partial x_3 \partial x_1}$. Cross-partial derivatives are symmetric if the second partial derivatives are continuous, which they are because Assumption 3 states f is three times differentiable.

By Part 1 of theorem, we know the signs of $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2}$ and $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3}$ if they are nonzero, as Part 2 requires. If f^1 implies different signs for $\frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2}$ and $\frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_3}$, then by Part 1 we can distinguish f^0 and f^1. So we can restrict attention to the case where the signs of $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2}$ and $\frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2}$ as well as $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3}$ and $\frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_3}$ are the same. I will first consider the case where $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2} > 0$ and $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_3} > 0$ for $f \in \left\{f^0, f^1\right\}$. The other cases are discussed at the end of the proof.

The outline of the proof follows. The most novel step comes first: I find a key inequality that arises from the numerator of the middle difference quotient, (18), and that has a different direction for f^0 and f^1. For example, this can be seen as a situation where f^0 would predict sorting on characteristics x_1 and x_2 while f^1 would predict sorting on characteristics x_1 and x_3 when sorting on both pairs simultaneously is physically impossible. The second step is that I show that this inequality is a local production maximization inequality. The third step is that I embed the firms in the local production maximization inequality into a larger matching market with characteristics X. The fourth and final step is that I show there is a continuum of matching market characteristics X where f^0 and f^1 give different predictions the local production maximization inequality and so where $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$, for appropriate choices of A_1 and A_2.

Let $h_{1,2}$ be the limit argument from the middle difference quotient, (18), for $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2}$. Likewise, let $h_{1,3}$ be the limit argument for $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_3}$. Consider the case $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3} > \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_3}$ and let $\{h_{1,2,n}\}_{n \in \mathbb{N}}$ be a sequence that converges to 0. Let $\{h_{1,3,n}\}_{n \in \mathbb{N}}$ be a sequence

$$h_{1,3,n} = h_{1,2,n} \sqrt{\frac{1}{2} \left(\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} \Big/ \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3} + \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} \Big/ \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_3} \right)}. \tag{21}$$

$\{h_{1,3,n}\}_{n \in \mathbb{N}}$ converges to 0 and

$$\frac{h_{1,3,n}^2}{h_{1,2,n}^2} = \frac{1}{2} \left(\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} \Big/ \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3} + \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} \Big/ \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_3} \right)$$

is the mean of the two ratios of cross-partial derivatives for all $n \in \mathbb{N}$. This choice of $h_{1,3,n}$ ensures

$$\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} \Big/ \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3} > \frac{h_{1,3,n}^2}{h_{1,2,n}^2} > \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} \Big/ \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_3} \tag{22}$$

for all $n \in \mathbb{N}$.

Let $\tau = \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3} - \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_3}$ and

$$\Upsilon\left(h_{1,2}, h_{1,3}; f\right) = \frac{f\left(\vec{y} + h_{1,2}e_1 + h_{1,2}e_2\right) - f\left(\vec{y} + h_{1,2}e_1\right) - f\left(\vec{y} + h_{1,2}e_2\right) + f(\vec{y})}{h_{1,2}^2} \cdot \left(\frac{f\left(\vec{y} + h_{1,3}e_1 + h_{1,3}e_3\right) - f\left(\vec{y} + h_{1,3}e_1\right) - f\left(\vec{y} + h_{1,3}e_3\right) + f(\vec{y})}{h_{1,3}^2} \right)^{-1} \tag{23}$$

for $f \in \mathscr{F}$. By the definition of a cross-partial derivative, (18), the ratio $\Upsilon\left(h_{1,2,n}, h_{1,3,n}; f\right)$ converges to $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f(\vec{y})}{\partial x_1 \partial x_3}$ for $f \in \{f^0, f^1\}$ as $n \to 0$ and $\left(h_{1,2,n}, h_{1,3,n}\right) \to (0,0)$. Then there exists some $n_1 \in \mathbb{N}$ where, for all $n \geq n_1$, $n \in \mathbb{N}$, $\left|\Upsilon\left(h_{1,2,n}, h_{1,3,n}; f^0\right) - \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3}\right| < \frac{\tau}{3}$ and $\left|\Upsilon\left(h_{1,2,n}, h_{1,3,n}; f^1\right) - \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_3}\right| < \frac{\tau}{3}$. The choice of distance $\frac{\tau}{3}$ ensures that

$$\Upsilon\left(h_{1,2,n}, h_{1,3,n}; f^0\right) > \frac{h_{1,3,n}^2}{h_{1,2,n}^2} > \Upsilon\left(h_{1,2,n}, h_{1,3,n}; f^1\right) \tag{24}$$

for all $n \geq n_1$, $n \in \mathbb{N}$. Define

$$\Delta\left(h_{1,2}, h_{1,3}; f\right) = f\left(\vec{y} + h_{1,2}e_1 + h_{1,2}e_2\right) - f\left(\vec{y} + h_{1,2}e_1\right) - f\left(\vec{y} + h_{1,2}e_2\right) + f(\vec{y}) - \left(f\left(\vec{y} + h_{1,3}e_1 + h_{1,3}e_3\right) - f\left(\vec{y} + h_{1,3}e_1\right) - f\left(\vec{y} + h_{1,3}e_3\right) + f(\vec{y})\right)$$

for $f \in \mathscr{F}$. Choose $\left(h_{1,2}, h_{1,3}\right) = \left(h_{1,2,n}, h_{1,3,n}\right)$. Substituting the definition of $\Upsilon\left(h_{1,2,n}, h_{1,3,n}; f\right)$ into (24) and resulting algebra shows that, at $(h_{1,2},h_{1,3})$, the ratios $h_{1,3,n}^2/h_{1,2,n}^2$ cancel in all terms and

$$\frac{f^0(\vec{y}+h_{1,2}e_1+h_{1,2}e_2)-f^0(\vec{y}+h_{1,2}e_1)-f^0(\vec{y}+h_{1,2}e_2)+f^0(\vec{y})}{f^0(\vec{y}+h_{1,3}e_1+h_{1,3}e_3)-f^0(\vec{y}+h_{1,3}e_1)-f^0(\vec{y}+h_{1,3}e_3)+f^0(\vec{y})}>1>\frac{f^1(\vec{y}+h_{1,2}e_1+h_{1,2}e_2)-f^1(\vec{y}+h_{1,2}e_1)-f^1(\vec{y}+h_{1,2}e_2)+f^1(\vec{y})}{f^1(\vec{y}+h_{1,3}e_1+h_{1,3}e_3)-f^1(\vec{y}+h_{1,3}e_1)-f^1(\vec{y}+h_{1,3}e_3)+f^1(\vec{y})} \quad (25)$$

and so

$$\Delta\left(h_{1,2},h_{1,3};f^0\right)>0>\Delta\left(h_{1,2},h_{1,3};f^1\right).$$

At this value $(h_{1,2},h_{1,3})$, f^0 and f^1 have different signs for a key term $\Delta(h_{1,2},h_{1,3};f)$. The same style of arguments and the same choice of $h_{1,3,n}$, (21), will apply to the case $\frac{\partial f^0(\vec{y})}{\partial x_1\partial x_2}/\frac{\partial f^0(\vec{y})}{\partial x_1\partial x_3}<\frac{\partial f^1(\vec{y})}{\partial x_1\partial x_2}/\frac{\partial f^1(\vec{y})}{\partial x_1\partial x_3}$. Only a few inequalities are reversed.

Now I will argue that $\Delta(h_{1,2},h_{1,3};f)$ can be used to form a local production maximization inequality. Rearrange the inequality $\Delta(h_{1,2},h_{1,3};f)>0$ so that all signs are positive:

$$f(\vec{y}+h_{1,2}e_1+h_{1,2}e_2)+f(\vec{y}+h_{1,3}e_1)+f(\vec{y}+h_{1,3}e_3)+f(\vec{y})> \\ f(\vec{y}+h_{1,3}e_1+h_{1,3}e_3)+f(\vec{y}+h_{1,2}e_1)+f(\vec{y}+h_{1,2}e_2)+f(\vec{y}). \quad (26)$$

Clearly this inequality is satisfied when $\Delta(h_{1,2},h_{1,3};f)>0$. The inequality (26) satisfies Definition 3 for some choice of B_1 and B_2. Let $B_1=\{\langle a,i\rangle,\langle b,j\rangle,\langle c,k\rangle,\langle g,l\rangle\}$ and $B_2=\{\langle g,i\rangle,\langle a,j\rangle,\langle b,k\rangle,\langle c,l\rangle\}$, where the permutation π is implied by the definitions of B_1 and B_2. Also, let $a\in C_i^u$, $b\in C_j^u$, $c\in C_k^u$ and $g\in C_l^u$. Let $\vec{x}(i,C_i^u)=\vec{y}+h_{1,2}e_1+h_{1,2}e_2$, $\vec{x}\left(j,C_j^u\right)=\vec{y}+h_{1,3}e_1$, $\vec{x}(k,C_k^u)=\vec{y}+h_{1,3}e_3$, $\vec{x}(l,C_l^u)=\vec{y}$, $\vec{x}(i,(C_i^u\setminus\{a\})\cup\{g\})=\vec{y}+h_{1,2}e_1$, $\vec{x}\left(j,\left(C_j^u\setminus\{b\}\right)\cup\{a\}\right)=\vec{y}+h_{1,3}e_1+h_{1,3}e_3$, $\vec{x}(k,\{C_k^u\setminus\{c\}\}\cup\{b\})=\vec{y}$, and $\vec{x}(l,(C_l^u\setminus\{g\})\cup\{c\})=\vec{y}+h_{1,2}e_2$. Using the notation $\vec{x}(i,C_i^u)=\text{cat}\left(\left(x_{i,1}^u,\ldots x_{i,K^u}^u\right),\left(x_{d_1,1}^d,\ldots x_{d_1,K^d}^d\right),\ldots,\left(x_{d_n,1}^d,\ldots x_{d_n,K^d}^d\right)\right)$: $x_{a,2}^d-h_{1,2}e_2=x_{g,2}^d$; either $x_{i,1}^u-h_{1,2}e_1=x_{l,1}^u$ or $x_{m_i,1}^d-h_{1,2}e_1=x_{m_l,1}^d$ for two firms $m_i\in C_i^u$, $m_i\neq a$ and $m_l\in C_l^u$, $m_l\neq g$; $x_{c,2}^d-h_{1,3}e_3=x_{b,2}^d$; and either $x_{j,1}^u-h_{1,3}e_1=x_{k,1}^u$ or $x_{m_j,1}^d-h_{1,3}e_1=x_{m_k,1}^d$ for two firms $m_j\in C_j^u$, $m_j\neq b$ and $m_k\in C_k^u$, $m_k\neq c$. By inspection, it can be seen that each match in B_1 exchanges a downstream firm partner for a match in B_2. Meanwhile, each set of arguments $\vec{x}(i,C_i^u)$ on the right can be formed by an exchange of single downstream firm's characteristics from a set of arguments on the left. Therefore, this construction satisfies Definition 3 for the case of firm-specific characteristics.

As in the proof of Part 1, I can embed B_1 and B_2 into a larger matching market. Let B_3 be a larger set of matches that includes the matches corresponding to the downstream firms in $C_i^u\setminus\{a\}$, $C_j^u\setminus\{b\}$, $C_k^u\setminus\{c\}$ and $C_l^u\setminus\{g\}$. The exact choice of B_3 plays no role in the proof, other than to ensure the sets of the form $C_i^u\setminus\{a\}$ are large enough given the number of non-empty elements in $\vec{y}$. Then set $A_1=B_1\cup B_3$ and $A_2=B_2\cup B_3$. Likewise, I will invent some collection X of characteristics for all potential matches, including but not limited to the matches in $A_1\cup A_2$. The choice of X plays no role in the proof, except that X must contain the eight terms of the form $\vec{x}(i,C_i^u)$ listed in the previous paragraph.

Assumption 1 states that if (26) holds for $f\in\mathscr{F}$, then $\Pr(A_1\mid X;f,S)>\Pr(A_2\mid X;f,S)$ for any $S\in\mathscr{S}$. Above, we found $(h_{1,2},h_{1,3})$ where $\Delta(h_{1,2},h_{1,3};f^0)>0$ and hence where (26) holds for the true f^0. Likewise, as this point $\Delta(h_{1,2},h_{1,3};f^1)<0$ and hence (26) does not hold for the alternative f^1.

Now we need to show that there exists a continuum of markets X where $\Pr(A_1\mid X;f^0,S^0)>\Pr(A_2\mid X;f^0,S^0)$

while $\Pr(A_1 \mid X; f^1, S^1) < \Pr(A_2 \mid X; f^1, S^1)$ for any $S^1 \in \mathscr{S}$. Because each $f \in \mathscr{F}$ is continuous, the function

$$g(z; f) = g(\vec{z}_1, \ldots, \vec{z}_8; f) = f(\vec{z}_1) + f(\vec{z}_2) + f(\vec{z}_3) + f(\vec{z}_4) - (f(\vec{z}_5) - f(\vec{z}_6) - f(\vec{z}_7) - f(\vec{z}_8))$$

is itself continuous. Let $v = \left(\vec{x}\left(i, C_i^u\right), \vec{x}\left(j, C_j^u\right), \ldots, \vec{x}\left(l, \left(C_l^u \setminus \{g\}\right) \cup \{c\}\right)\right)$. Let V^0 be an open set around $g(v; f^0)$ where $p > 0$ for $p \in V^0$ and let V^1 be an open set around $g(v; f^1)$ where $p < 0$ for $p \in V^1$. Such open neighborhoods exist as g is continuous. Because the tuple v is the same in both the f^0 and f^1 constructions, $W = g^{-1}(V^0; f^0) \cap g^{-1}(V^1; f^1) \neq \emptyset$ and is itself an open neighborhood of v, as both $g^{-1}(V^0; f^0)$ and $g^{-1}(V^1; f^1)$ are open sets by the definition of continuity and topologies are closed under finite intersections. By construction, $g(z; f^0) > 0$ is equivalent to saying a local production maximization inequality is satisfied; the opposite local production maximization inequality is satisfied for f^1 when $g(z; f^1) < 0$.

Divide any market characteristics X into X_1 and X_2, where X_1 is the set comprised of the eight elements of the tuple v or the eight arguments of the g function. Also, $X_2 = X \setminus X_1$, so X is a one-to-one change of variables (or just notation) of (X_1, X_2). By Assumption 2, X has support equal to the product of the marginal supports of its elements. Therefore, the probability that (X_1, X_2) lies in $W \times \mathbb{R}^s$ is strictly positive, where s is the number of elements of X_2. All market characteristics (X_1, X_2) in $W \times \mathbb{R}^s$ satisfy $\Pr(A_1 \mid X; f^0, S^0) > \Pr(A_2 \mid X; f^0, S^0)$ while $\Pr(A_1 \mid X; f^1, S^1) < \Pr(A_2 \mid X; f^1, S^1)$ for any $S_1 \in \mathscr{S}$, by the key Assumption 1. By the arguments at the beginning of the proof, identification has been shown for the case under consideration.

Recall that by Part 1 of the theorem we can focus on cases where the signs of the cross partials are the same for f^0 and f^1. Before we restricted attention to the case $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2} > 0$, $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_3} > 0$ and $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2} > \frac{\partial f(\vec{y})}{\partial x_1 \partial x_3}$ for $f \in \{f^0, f^1\}$. $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2} > \frac{\partial f(\vec{y})}{\partial x_1 \partial x_3}$ is without loss of generality,[81] but $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2} > 0$ and $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_3} > 0$ for $f \in \{f^0, f^1\}$ are conditions with some loss of generality. Now we need to argue that the above arguments go through for the other three cases: $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2} < 0$ and $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_3} > 0$; $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2} > 0$ and $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_3} < 0$; as well as $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2} < 0$ and $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_3} < 0$. This is simple: in some of these new cases key inequalities may reverse direction, but as $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3} \neq \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_3}$ for all cases, the same arguments as above will show $(h_{1,2}, h_{1,3})$ can be chosen to lie in between $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3}$ and $\frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_3}$. One this is done, the arguments about local production maximization inequalities, embedding these matches in a larger matching market, and then finding a continuum of matching markets where f^0 and f^1 make different predictions can all be repeated without much change.[82]

A.4 Theorem 2: Cardinal identification, match-specific covariates

To a large degree, the arguments for the proof of Theorem 2 mirror those of Theorem 1. However, as Theorem 2 involves match-specific characteristics and second derivatives rather than firm-specific characteristics and cross-partial derivatives, it is not enough to simply assert that the proofs are similar enough that the proof of Theorem 2 can be disregarded.

[81] In part, there is no loss in generality because identifying $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3}$ is equivalent to identifying its inverse, $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3} / \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2}$.

[82] If $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f(\vec{y})}{\partial x_1 \partial x_3} < 0$ for $f \in \{f^0, f^1\}$, then (21) will involve the square root of a negative number. To fix this, let (21) involve the absolute values of $\frac{\partial f(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f(\vec{y})}{\partial x_1 \partial x_3}$ for $f \in \{f^0, f^1\}$. For the case $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3} > \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_3}$, (31) will become $\frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^0(\vec{y})}{\partial x_1 \partial x_3} > -\frac{h_{1,3,n}^2}{h_{1,2,n}^2} > \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial f^1(\vec{y})}{\partial x_1 \partial x_3}$. Following the steps of the algebra in the earlier argument, the 1 in (25) will be a -1 and the pair of inequalities in (25) will reverse directions once both sides are multiplied by the -1. A different local production maximization inequality will arise, but otherwise the argument is similar to the earlier argument.

A.4.1 Part 1

The vector $\vec{x}$ is given in the statement of the theorem. To avoid confusion of the point $\vec{x}$ with the function $\vec{x}\left(i, C_i^u\right)$, I relabel the vector $\vec{x}$ as $\vec{y}$ inside this proof. I will prove Part 1 for the case where f^0 is concave in x_1 at x_1, $\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_1} > 0$. The case when f^0 is convex is similar mathematically and so is omitted.

For an arbitrary $f^1 \in \mathscr{F}, f^1 \neq f^0$ where $\frac{\partial^2 f^1(\vec{y})}{\partial^2 x_1} < 0$, by Definition 4 I must show that there does not exist S^1 corresponding to f^1 where $\Pr\left(A \mid X; f^0, S^0\right) = \Pr\left(A \mid X; f^1, S^1\right)$ for all (A, X) except perhaps a set of X of probability 0. By the key Assumption 1, a sufficient condition involves showing that there exists a continuum of market characteristics X with positive probability and a corresponding matching situation where f^0 and f^1 give different implications for a local production maximization inequality, of the form in Definition 3. At each of these markets X, there will be a particular assignment A_1 and another assignment A_2 where, by Assumption 1, $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$. Therefore, the conditions of Definition 4 will be satisfied.

One definition of a second derivative is

$$\frac{\partial^2 f(\vec{y})}{\partial^2 x_1} = \lim_{h \to \infty} \frac{f(\vec{y} + 2he_1) - 2f(\vec{y} + he_1) + f(\vec{y})}{h^2}, \tag{27}$$

where as in the proof of Theorem 1, $e_1 = (1, 0, 0, \ldots, 0)$ is a vector of 0's except in the first element. Because both $\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_1}$ and $\frac{\partial^2 f^1(\vec{y})}{\partial^2 x_1}$ are limits, there will be some $h = \min\left\{h^0, h^1\right\}$, where h^0 and h^1 are the respective limit arguments for f^0 and f^1, where both

$$f^0(\vec{y} + 2he_1) + f^0(\vec{y}) > 2f^0(\vec{y} + he_1) \tag{28}$$

and

$$f^1(\vec{y} + 2he_1) + f^1(\vec{y}) < 2f^1(\vec{y} + he_1) \tag{29}$$

hold.

Now I will argue that (28) and by the same argument (29) are local production maximization equations, Definition 3. Let $B_1 = \{\langle a, i\rangle, \langle b, j\rangle\}$ and $B_2 = \{\langle a, j\rangle, \langle b, i\rangle\}$. Also let there be sufficiently large (to handle $\vec{y}$) sets C_i^u and C_j^u where $a \in C_i^u$, $a \notin C_j^u$, $b \in C_j^u$, and $b \notin C_i^u$. Also define $\vec{x}\left(i, C_i^u\right) = \vec{y} + 2he_1$, $\vec{x}\left(j, C_j^u\right) = \vec{y}$, $\vec{x}\left(i, \left(C_i^u \backslash \{a\}\right) \cup \{b\}\right) = \vec{y} + he_1$ and $\vec{x}\left(j, \left(C_j^u \backslash \{b\}\right) \cup \{a\}\right) = \vec{y} + he_1$. Using the match-specific characteristics notation $\vec{x}(i, C^u) = \text{cat}\left(\left(x^{u,d}_{\langle d_1, i\rangle, 1}, \ldots x^{u,d}_{\langle d_1, i\rangle, K}\right), \ldots, \left(x^{u,d}_{\langle d_n, i\rangle, 1}, \ldots x^{u,d}_{\langle d_n, i\rangle, K}\right)\right)$, let $x^{u,d}_{\langle a, i\rangle, 1} - he_1 = x^{u,d}_{\langle a, j\rangle, 1} = x^{u,d}_{\langle b, i\rangle, 1}$ and $x^{u,d}_{\langle a, i\rangle, 1} - 2he_1 = x^{u,d}_{\langle b, j\rangle, 1}$. With $\pi\langle a, i\rangle = b$ and $\pi\langle b, j\rangle = a$, inspection shows (19) satisfies Definition 3 for the case of match-specific characteristics.

I can embed B_1 and B_2 in a matching market. Let B_3 be a larger set of matches that includes the matches corresponding to the downstream firms in $C_i^u \backslash \{a\}$ and $C_j^u \backslash \{b\}$. The exact choice of B_3 plays no role in the proof, other than to ensure $C_i^u \backslash \{a\}$ and $C_j^u \backslash \{b\}$ are large enough given the number of non-empty elements in $\vec{y}$. Then set $A_1 = B_1 \cup B_3$ and $A_2 = B_2 \cup B_3$. Likewise, let there be a collection X of characteristics for all potential matches, including but not limited to the matches in $A_1 \cup A_2$. The choice of X plays no role in the proof, except that X must contain the four terms of the form $\vec{x}(i, C^u)$ described in the previous paragraph.

Assumption 1 and (19) imply $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while Assumption 1 and (20) imply $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$. Now we need to find a continuum of markets X with this property. Define

$$g(z;f) = g(\vec{z}_1,\vec{z}_2,\vec{z}_3,\vec{z}_4;f) = f(\vec{z}_1) + f(\vec{z}_2) - f(\vec{z}_3) - f(\vec{z}_4).$$

As $f \in \mathscr{F}$ is continuous, g is continuous in the tuple $(\vec{z}_1,\vec{z}_2,\vec{z}_3,\vec{z}_4)$. Let

$$v = \left(\vec{x}(i,C_i^u), \vec{x}\left(j,C_j^u\right), \vec{x}(i,(C_i^u\backslash\{a\})\cup\{b\}), \vec{x}\left(j,\left(C_j^u\backslash\{b\}\right)\cup\{a\}\right)\right).$$

Let V^0 be an open set around $g\left(v;f^0\right)$ where $p>0$ for $p \in V^0$ and let V^1 be an open set around $g\left(v;f^1\right)$ where $p<0$ for $p \in V^1$. Such sets exist by the continuity of g. By the continuity of g and the fact that topologies are closed under finite intersections, $W = g^{-1}\left(V^0;f^0\right) \cap g^{-1}\left(V^1;f^1\right)$ is a nonempty, open neighborhood of v where $g\left(z;f^0\right)>0$ and $g\left(z;f^1\right)<0$ for $z \in W$.

Divide any market characteristics X into X_1 and X_2, where X_1 is the set comprised of the eight elements of the tuple v. Also, $X_2 = X\backslash X_1$, so X is a one-to-one change of variables (or just notation) of (X_1,X_2). By Assumption 2, X has support equal to the product of the marginal supports of its elements. Therefore, the probability that (X_1,X_2) lies in $W \times \mathbb{R}^s$ is strictly positive, where s is the number of elements of X_2. All market characteristics (X_1,X_2) in $W \times \mathbb{R}^s$ satisfy $\Pr\left(A_1 \mid X;f^0,S^0\right) > \Pr\left(A_2 \mid X;f^0,S^0\right)$ while $\Pr\left(A_1 \mid X;f^1,S^1\right) < \Pr\left(A_2 \mid X;f^1,S^1\right)$ for any $S_1 \in \mathscr{S}$, by the key Assumption 1. By the arguments at the beginning of the proof, identification has been shown for the case under consideration.

A.4.2 Part 2

We are given a point $\vec{y}$ (relabeled from $\vec{x}$ in the statement of the theorem) and there is an arbitrary $f^1 \in \mathscr{F}$ where $\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_1}/\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_2} \neq \frac{\partial^2 f^1(\vec{y})}{\partial^2 x_1}/\frac{\partial^2 f^1(\vec{y})}{\partial^2 x_2}$. The goal is to show there exists a continuum of X and two assignments A_1 and A_2 where $\Pr\left(A_1 \mid X;f^0,S^0\right) > \Pr\left(A_2 \mid X;f^0,S^0\right)$ while $\Pr\left(A_1 \mid X;f^1,S^1\right) < \Pr\left(A_2 \mid X;f^1,S^1\right)$ for any $S_1 \in \mathscr{S}$. By Part 1, we can restrict attention to the case where the pair $\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_1}$ and $\frac{\partial^2 f^1(\vec{y})}{\partial^2 x_1}$ as well as the pair $\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_2}$ and $\frac{\partial^2 f^1(\vec{y})}{\partial^2 x_2}$ have the same signs. I first consider the case where $\frac{\partial^2 f(\vec{y})}{\partial^2 x_1} > 0$ and $\frac{\partial^2 f(\vec{y})}{\partial^2 x_2} > 0$ for $f \in \left\{f^0,f^1\right\}$. The other cases are discussed at the end of the proof.

Let h_1 be the index for the approximation term on the right side of (27) for $\frac{\partial^2 f(\vec{y})}{\partial^2 x_1}$ and let h_2 be the index for $\frac{\partial^2 f(\vec{y})}{\partial^2 x_2}$. Consider the case $\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_1}/\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_2} > \frac{\partial^2 f^1(\vec{y})}{\partial^2 x_1}/\frac{\partial^2 f^1(\vec{y})}{\partial^2 x_2}$ and let $\left\{h_{1,n}\right\}_{n\in\mathbb{N}}$ be a sequence that converges to 0. Let $\left\{h_{2,n}\right\}_{n\in\mathbb{N}}$ be a sequence

$$h_{2,n} = h_{1,n}\sqrt{\frac{1}{2}\left(\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_1}/\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_2} + \frac{\partial^2 f^1(\vec{y})}{\partial^2 x_1}/\frac{\partial^2 f^1(\vec{y})}{\partial^2 x_2}\right)}. \tag{30}$$

$\left\{h_{2,n}\right\}_{n\in\mathbb{N}}$ converges to 0 and

$$\frac{h_{2,n}^2}{h_{1,n}^2} = \frac{1}{2}\left(\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_1}/\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_2} + \frac{\partial^2 f^1(\vec{y})}{\partial^2 x_1}/\frac{\partial^2 f^1(\vec{y})}{\partial^2 x_2}\right)$$

is the mean of the two ratios of second partial derivatives for all $n \in \mathbb{N}$. This choice of $h_{2,n}$ ensures

$$\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_1}/\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_2} > \frac{h_{2,n}^2}{h_{1,n}^2} > \frac{\partial^2 f^1(\vec{y})}{\partial^2 x_1}/\frac{\partial^2 f^1(\vec{y})}{\partial^2 x_2} \tag{31}$$

for all $n \in \mathbb{N}$.

Let $\tau = \frac{\partial^2 f^0(\vec{y})}{\partial^2 x_1} / \frac{\partial^2 f^0(\vec{y})}{\partial^2 x_2} - \frac{\partial^2 f^1(\vec{y})}{\partial^2 x_1} / \frac{\partial^2 f^1(\vec{y})}{\partial^2 x_2}$ and

$$\Upsilon(h_1, h_2; f) = \frac{f(\vec{y}+2h_1e_1) - 2f(\vec{y}+h_1e_1) + f(\vec{y})}{h_1^2} \cdot \left(\frac{f(\vec{y}+2h_2e_2) - 2f(\vec{y}+h_2e_2) + f(\vec{y})}{h_2^2} \right)^{-1} \tag{32}$$

for $f \in \mathscr{F}$. By the definition of a second partial derivative, (27), the ratio $\Upsilon(h_1, h_2; f)$ converges to $\frac{\partial^2 f(\vec{y})}{\partial^2 x_1} / \frac{\partial^2 f(\vec{y})}{\partial^2 x_2}$ for $f \in \{f^0, f^1\}$ as $n \to 0$ and $(h_{1,n}, h_{2,n}) \to (0,0)$. Then there exists some $n_1 \in \mathbb{N}$ where, for all $n \geq n_1$, $n \in \mathbb{N}$, $\left| \Upsilon(h_1, h_2; f^0) - \frac{\partial^2 f^0(\vec{y})}{\partial^2 x_1} / \frac{\partial^2 f^0(\vec{y})}{\partial^2 x_2} \right| < \frac{\tau}{3}$ and $\left| \Upsilon(h_1, h_2; f^1) - \frac{\partial^2 f^1(\vec{y})}{\partial^2 x_1} / \frac{\partial^2 f^1(\vec{y})}{\partial^2 x_2} \right| < \frac{\tau}{3}$. The choice of distance $\frac{\tau}{3}$ ensures that

$$\Upsilon\left(h_1, h_2; f^0\right) > \frac{h_{2,n}^2}{h_{1,n}^2} > \Upsilon\left(h_1, h_2; f^1\right) \tag{33}$$

for all $n \geq n_1$, $n \in \mathbb{N}$. Define

$$\Delta(h_1, h_2; f) = f(\vec{y}+2h_1e_1) - 2f(\vec{y}+h_1e_1) + f(\vec{y}) - (f(\vec{y}+2h_2e_2) - 2f(\vec{y}+h_2e_2) + f(\vec{y}))$$

for $f \in \mathscr{F}$. Choose $(h_1, h_2) = (h_{1,n}, h_{2,n})$ for $n \geq n_1$. Substituting the definition of $\Upsilon(h_1, h_2; f)$ into (33) and resulting algebra shows that at (h_1, h_2), the ratios $h_{2,n}^2 / h_{1,n}^2$ cancel in all terms and

$$\frac{f^0(\vec{y}+2h_1e_1) - 2f^0(\vec{y}+h_1e_1) + f^0(\vec{y})}{f^0(\vec{y}+2h_2e_2) - 2f^0(\vec{y}+h_2e_2) + f^0(\vec{y})} > 1 > \frac{f^1(\vec{y}+2h_1e_1) - 2f^1(\vec{y}+h_1e_1) + f^1(\vec{y})}{f^1(\vec{y}+2h_2e_2) - 2f^1(\vec{y}+h_2e_2) + f^1(\vec{y})}$$

and so

$$\Delta\left(h_1, h_2; f^0\right) > 0 > \Delta\left(h_1, h_2; f^1\right).$$

At this value $(h_{1,2}, h_{1,3})$, f^0 and f^1 have different signs for a key term $\Delta(h_1, h_2; f^0)$. The same style of arguments will apply to the case $\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_1} / \frac{\partial^2 f^0(\vec{y})}{\partial^2 x_2} < \frac{\partial^2 f^1(\vec{y})}{\partial^2 x_1} / \frac{\partial^2 f^1(\vec{y})}{\partial^2 x_2}$. Only a few inequalities are reversed.

Now we can rearrange the inequality $\Delta(h_1, h_2; f^0) > 0$, giving

$$f(\vec{y}+2h_1e_1) + 2f(\vec{y}+h_2e_2) + f(\vec{y}) > f(\vec{y}+2h_2e_2) + 2f(\vec{y}+h_1e_1) + f(\vec{y}). \tag{34}$$

We can show that this is a local production maximization inequality, Definition 3, for some choice of B_1 and B_2. Let $B_1 = \{\langle a,i \rangle, \langle b,j \rangle, \langle c,k \rangle, \langle g,l \rangle\}$ and $B_2 = \{\langle g,i \rangle, \langle a,j \rangle, \langle b,k \rangle, \langle c,l \rangle\}$, where the permutation π is implied by the definitions of B_1 and B_2. Also, let $a \in C_i^u$, $b \in C_j^u$, $c \in C_k^u$ and $g \in C_l^u$. Let $\vec{x}(i, C_i^u) = \vec{y} + 2h_1e_1$, $\vec{x}\left(j, C_j^u\right) = \vec{y} + h_2e_2$, $\vec{x}(k, C_k^u) = \vec{y} + h_2e_2$, $\vec{x}(l, C_l^u) = \vec{y}$, $\vec{x}(i, (C_i^u \setminus \{a\}) \cup \{g\}) = \vec{y}$, $\vec{x}\left(j, \left(C_j^u \setminus \{b\}\right) \cup \{a\}\right) = \vec{y} + h_1e_1$, $\vec{x}(k, \{C_k^u \setminus \{c\}\} \cup \{b\}) = \vec{y} + h_1e_1$, and $\vec{x}(l, (C_l^u \setminus \{g\}) \cup \{c\}) = \vec{y} + 2h_2e_2$. Using the match-specific characteristics notation $\vec{x}(i, C^u) = \text{cat}\left(\left(x^{u,d}_{\langle d_1,i \rangle,1}, \ldots x^{u,d}_{\langle d_1,i \rangle,K}\right), \ldots, \left(x^{u,d}_{\langle d_n,i \rangle,1}, \ldots x^{u,d}_{\langle d_n,i \rangle,K}\right)\right)$, let $x^{u,d}_{\langle a,i \rangle,1} - 2h_1e_1 = x^{u,d}_{\langle g,l \rangle,1} = x^{u,d}_{\langle g,i \rangle,1} = x^{u,d}_{\langle b,j \rangle,1} = x^{u,d}_{\langle c,k \rangle,1} = x^{u,d}_{\langle c,l \rangle,1}$; $x^{u,d}_{\langle a,i \rangle,1} - h_1e_1 = x^{u,d}_{\langle a,j \rangle,1} = x^{u,d}_{\langle b,k \rangle,1}$; $x^{u,d}_{\langle c,l \rangle,2} - 2h_2e_2 = x^{u,d}_{\langle a,i \rangle,2} = x^{u,d}_{\langle c,j \rangle,2} = x^{u,d}_{\langle b,k \rangle,2} = x^{u,d}_{\langle g,l \rangle,2} = x^{u,d}_{\langle g,i \rangle,2}$; and $x^{u,d}_{\langle c,l \rangle,2} - h_2e_2 = x^{u,d}_{\langle b,j \rangle,2} = x^{u,d}_{\langle c,k \rangle,2}$. By inspection, it can be seen that each match in B_1 exchanges a downstream firm partner for a match in B_2. Meanwhile, each set of arguments $\vec{x}(i, C_i^u)$ on the right can be formed by replacing the characteristics associated with a single match in a set of arguments $\vec{x}(i, C_i^u)$ on the left side. Therefore, Definition 3 is satisfied.

Now we embed B_1 and B_2 into a larger matching market. Let B_3 be a larger set of matches that includes the matches corresponding to the downstream firms in $C_i^u \setminus \{a\}$, $C_j^u \setminus \{b\}$, $C_k^u \setminus \{c\}$ and $C_l^u \setminus \{g\}$. The exact choice of B_3

plays no role in the proof, other than to ensure the sets of the form $C_i^u \backslash \{a\}$ are large enough given the number of non-empty elements in $\vec{y}$. Then set $A_1 = B_1 \cup B_3$ and $A_2 = B_2 \cup B_3$. Likewise, let there be some collection X of characteristics for all potential matches, including but not limited to the matches in $A_1 \cup A_2$. The choice of X plays no role in the proof, except that X must contain the eight terms of the form $\vec{x}\left(i, C_i^u\right)$ listed in the previous paragraph.

Assumption 1 states that if (34) holds for $f \in \mathscr{F}$, then $\Pr(A_1 \mid X; f, S) > \Pr(A_2 \mid X; f, S)$ for any $S \in \mathscr{S}$. Above, we found a h_1 and an h_2 where $\Delta\left(h_1, h_2; f^0\right) > 0$ and hence where (26) holds for the true f^0. Likewise, as this point $\Delta\left(h_1, h_2; f^1\right) < 0$ and hence (26) does not hold for the alternative f^1.

Now we need to show that there exists a continuum of markets X where $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S^1 \in \mathscr{S}$. Because each $f \in \mathscr{F}$ is continuous, the function

$$g(z; f) = g(\vec{z}_1, \ldots, \vec{z}_8; f) = f(\vec{z}_1) + f(\vec{z}_2) + f(\vec{z}_3) + f(\vec{z}_4) - (f(\vec{z}_5) - f(\vec{z}_6) - f(\vec{z}_7) - f(\vec{z}_8))$$

is itself continuous. Let $v = \left(\vec{x}\left(i, C_i^u\right), \vec{x}\left(j, C_j^u\right), \ldots, \vec{x}\left(l, \left(C_l^u \backslash \{g\}\right) \cup \{c\}\right)\right)$. Let V^0 be an open set around $g\left(v; f^0\right)$ where $p > 0$ for $p \in V^0$ and let V^1 be an open set around $g\left(v; f^1\right)$ where $p < 0$ for $p \in V^1$. Such open neighborhoods exist as g is continuous. Because the tuple v is the same in both the f^0 and f^1 constructions, $W = g^{-1}\left(V^0; f^0\right) \cap g^{-1}\left(V^1; f^1\right) \neq \emptyset$ and is itself an open neighborhood of v, as both $g^{-1}\left(V^0; f^0\right)$ and $g^{-1}\left(V^1; f^1\right)$ are open sets by the definition of continuity and topologies are closed under finite intersections. By construction, $g\left(z; f^0\right) > 0$ is equivalent to saying a local production maximization inequality is satisfied; the opposite local production maximization inequality is satisfied for f^1 when $g\left(z; f^1\right) < 0$.

Divide any market characteristics X into X_1 and X_2, where X_1 is the set comprised of the eight elements of the tuple v or the eight arguments of the g function. Also, $X_2 = X \backslash X_1$, so X is a one-to-one change of variables (or just notation) of (X_1, X_2). By Assumption 2, X has support equal to the product of the marginal supports of its elements. Therefore, the probability that (X_1, X_2) lies in $W \times \mathbb{R}^s$ is strictly positive, where s is the number of elements of X_2. All market characteristics (X_1, X_2) in $W \times \mathbb{R}^s$ satisfy $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$, by the key Assumption 1. By the arguments at the beginning of the proof, identification has been shown for the case under consideration.

Before we restricted attention to the case $\frac{\partial^2 f(\vec{y})}{\partial^2 x_1} > 0$ and $\frac{\partial^2 f(\vec{y})}{\partial^2 x_2} > 0$ for $f \in \left\{f^0, f^1\right\}$. However, it should be clear that for other cases, we can always find two sequences so that $h_{2,n}^2 / h_{1,n}^2$ lies in between $\frac{\partial^2 f^0(\vec{y})}{\partial^2 x_1} / \frac{\partial^2 f^0(\vec{y})}{\partial^2 x_2}$ and $\frac{\partial^2 f^1(\vec{y})}{\partial^2 x_1} / \frac{\partial^2 f^1(\vec{y})}{\partial^2 x_2}$. That is the key insight of the proof: find a market where f^0 and f^1 give different implications for the local production maximization inequality.

A.5 Lemma 3: Continuous characteristics for ordinal identification

Without loss of generality, the goal of the proof is to show that the set

$$W^1 = \left\{(\vec{x}_a, \vec{x}_b) \mid f^1(\vec{x}_a) > f^1(\vec{x}_b) \text{ and } f^2(\vec{x}_a) < f^2(\vec{x}_b)\right\}$$

is non-empty. Let $\vec{x} = \text{cat}((x_1), \vec{x}_{-1})$.

First we want to show that, again without loss of generality,

$$W^2 = \left\{(\vec{x}_a, \vec{x}_b) \mid f^1(\vec{x}_a) \geq f^1(\vec{x}_b) \text{ and } f^2(\vec{x}_a) < f^2(\vec{x}_b)\right\}$$

is non-empty. Assume not. Then f^1 and f^2 induce the same ordering, or preference relation in utility theory. The "only if" direction of Theorem 1.2 in Jehle and Reny (2000) shows that there must exist some positive, strictly monotonic function m such that $f^1(\vec{x}_a) = m \circ f^2(\vec{x}_b)$ over the range of values taken on by f^2. As this contradicts Assumption 3, W^2 must be non-empty. The proof in Jehle and Reny is left as an exercise for the reader. Here is a quick sketch: pick an arbitrary point $\vec{x}_c$, let $y_c = f^2(\vec{x}_c)$ and $m(y_c) = f^1(\vec{x}_c)$. Repeat for all points. m is strictly increasing over the values f^2 takes on because if $f^2(\vec{x}_c) > f^2(\vec{x}_d)$ then $f^1(\vec{x}_c) > f^1(\vec{x}_d)$, under the contradiction, and $f^1(\vec{x}) = m \circ f^2(\vec{x})$.

We have shown W^2 is non-empty. Take a point $(\vec{x}_a, \vec{x}_b) \in W^2$. Then add $\delta_1 > 0$ to the x_1 element of $\vec{x}_a$. Because f^1 is strictly increasing in x_1, $f^1(\vec{x}_a + e_1\delta_1) > f^1(\vec{x}_b)$, where $e_1 = (1,0,0,\ldots,0)$ is a vector of length equal to the length of $\vec{x}_a$. Because f^2 is continuous, there exists a $\delta_2 > 0$ where $f^2(\vec{x}_a - e_1\delta_2) < f^2(\vec{x}_b)$ is preserved. Let $\delta = \min\{\delta_1, \delta_2\}$. The points $\vec{x}_a + e_1\delta$ and $\vec{x}_b$ satisfy the requirements of the lemma.

A.6 Theorem 4: Ordinal identification, group characteristics

Let $f^0, f^1 \in \mathscr{F}$, where f^0 is the production function to be identified and f^1 is an alternative where $f^1(\vec{x}) \neq m \circ f^0(\vec{x})$ for all $\vec{x}$ any for any positive monotonic function m. The goal is to show there exists a continuum of X and two assignments A_1 and A_2 where $\Pr(A_1 \mid X; f^0, S^0) > \Pr(A_2 \mid X; f^0, S^0)$ while $\Pr(A_1 \mid X; f^1, S^1) < \Pr(A_2 \mid X; f^1, S^1)$ for any $S_1 \in \mathscr{S}$.

Lemma 3 produces $\vec{x}_1$ and $\vec{x}_2$ such that $f^0(\vec{x}_1) > f^0(\vec{x}_2)$ and $f^1(\vec{x}_1) < f^1(\vec{x}_2)$ or $f^0(\vec{x}_1) < f^0(\vec{x}_2)$ and $f^1(\vec{x}_1) > f^1(\vec{x}_2)$. Focus on the first case. An inequality such as $f^0(\vec{x}_1) > f^0(\vec{x}_2)$ considers a group of matches centered around an upstream firm on the left and another group of matches centered around an upstream firm on the right. This is not a local production maximization inequality (Definition 3), which would require at least two groups, each centered on an upstream firm, on both the left and the right.

Consider a third set of characteristics, $\vec{x}_3$. The exact value of $\vec{x}_3$ will not matter for the case of group-specific characteristics. Add its production to both sides of the inequality $f(\vec{x}_1) > f(\vec{x}_2)$ to give

$$f(\vec{x}_1) + f(\vec{x}_3) > f(\vec{x}_2) + f(\vec{x}_3). \tag{35}$$

This inequality is satisfied for $f = f^0$; the opposite direction is satisfied for $f = f^1$.

I will now argue that this is a local production maximization inequality, Definition 3. Let $B_1 = \{\langle a,i\rangle, \langle b,j\rangle\}$ and $B_2 = \{\langle b,i\rangle, \langle a,j\rangle\}$. Also let there be sufficiently large (to handle $\vec{y}$) sets C_i^u and C_j^u where $a \in C_i^u$, $a \notin C_j^u$, $b \in C_j^u$, and $b \notin C_i^u$. Also define $\vec{x}(i, C_i^u) = \vec{x}_1$, $\vec{x}\left(j, C_j^u\right) = \vec{x}_3$, $\vec{x}(i, (C_i^u \setminus \{a\}) \cup \{b\}) = \vec{x}_2$ and $\vec{x}\left(j, \left(C_j^u \setminus \{b\}\right) \cup \{a\}\right) = \vec{x}_3$. Using the group-specific characteristics notation $\vec{x}(i, C^u) = \left(x^{\text{group}}_{(i,C^u),1}, \ldots, x^{\text{group}}_{(i,C^u),K}\right)$, these four distinct groups can have the four different covariate vectors listed. With $\pi\langle a,i\rangle = b$ and $\pi\langle b,j\rangle = a$, inspection shows (35) satisfies Definition 3 for the case of match-specific characteristics.

We can embed these four groups into a larger matching market. Let B_3 be a larger set of matches that includes the matches corresponding to the downstream firms in $C_i^u \setminus \{a\}$ and $C_j^u \setminus \{b\}$. The exact choice of B_3 plays no role in the proof, other than to ensure $C_i^u \setminus \{a\}$ and $C_j^u \setminus \{b\}$ are large enough given the number of non-empty elements in an arbitrary $\vec{x}$. Then set $A_1 = B_1 \cup B_3$ and $A_2 = B_2 \cup B_3$. Likewise, let there be some collection X of characteristics for all potential matches, including but not limited to the matches in $A_1 \cup A_2$. The choice of X plays no role in the proof, except that X must contain $\vec{x}(i, C_i^u)$, $\vec{x}\left(j, C_j^u\right)$, $\vec{x}(i, (C_i^u \setminus \{a\}) \cup \{b\})$ and $\vec{x}\left(j, \left(C_j^u \setminus \{b\}\right) \cup \{a\}\right)$.

Definition 4 requires that we find a set of X with positive probability and two assignments A_1 and A_2 where $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$. Unlike for the cardinal identification theorems, where I assumed all observable characteristics are continuous for transparency, here we assume that there need be only one component of a group-specific characteristics vector, $\vec{x}$, x_1, with continuous and full support. Partition X into (X', X''), where X' has all of the continuously varying characteristics in X (variables whose marginal density conditional on X'' has support equal to a connected subset of the real line) and X'' is the collection of all variables with discrete support in X. We will fix X'' and vary only the elements in X'. Likewise, for each $\vec{x}$, partition it into $\vec{x} = (\vec{x}', \vec{x}'')$, where $\vec{x}'$ has all of the continuous variables in $\vec{x}$ and $\vec{x}''$ has all of the discrete variables. Assume the arguments of f are appropriately reordered to respect this partition. Further, x_1 is the first element of $\vec{x}'$.

Let
$$g\left(\vec{z}_1', \vec{z}_2', \vec{z}_3', \vec{z}_4'; f\right) \equiv f\left(\left(\vec{z}_1', \vec{x}_1''\right)\right) + f\left(\left(\vec{z}_2', \vec{x}_3''\right)\right) - f\left(\left(\vec{z}_3', \vec{x}_2''\right)\right) - f\left(\left(\vec{z}_4', \vec{x}_3''\right)\right),$$
where $\left(\vec{z}_1', \vec{z}_2', \vec{z}_3', \vec{z}_4'\right)$ are arbitrary continuous arguments and $\vec{x}_1''$, $\vec{x}_2''$, and $\vec{x}_3''$ are the discrete components of the $\vec{x}_1$, $\vec{x}_2$ and $\vec{x}_3$ in (35). As each $f \in \mathscr{F}$ is continuous in its arguments with continuous support, then so is g. Let v' be the continuous elements of $\left(\vec{x}\left(i, C_i^u\right), \vec{x}\left(j, C_j^u\right), \vec{x}\left(i, \left(C_i^u \backslash \{a\}\right) \cup \{b\}\right), \vec{x}\left(j, \left(C_j^u \backslash \{b\}\right) \cup \{a\}\right)\right)$, or, in notation, $v' = \left(\vec{x}'\left(i, C_i^u\right), \vec{x}'\left(j, C_j^u\right), \vec{x}'\left(i, \left(C_i^u \backslash \{a\}\right) \cup \{b\}\right), \vec{x}'\left(j, \left(C_j^u \backslash \{b\}\right) \cup \{a\}\right)\right)$. Let V^0 be an open set around $g\left(v'; f^0\right)$ where $p > 0$ for $p \in V^0$ and let V^1 be an open set around $g\left(v'; f^1\right)$ where $p < 0$ for $p \in V^1$. Such sets exist by the continuity of g. By the continuity of g and the fact that topologies are closed under finite intersections, $W = g^{-1}\left(V^0; f^0\right) \cap g^{-1}\left(V^1; f^1\right)$ is a nonempty, open neighborhood of v' where $g\left(z; f^0\right) > 0$ and $g\left(z; f^1\right) < 0$ for $z \in W$.

Divide any continuous market characteristics X' into X_1' and X_2', where X_1' is the set comprised of the four elements of the tuple v' or the four arguments of the g function. Also, $X_2' = X' \backslash X_1'$, so X' is a one-to-one change of variables (or just notation) of $\left(X_1', X_2'\right)$. By Assumption 4, X_1' has support equal to the product of the marginal supports of its elements. With these notational additions, X is a change of variables from $\left(X_1', X_2', X''\right)$. For any X'', the probability that $\left(X_1', X_2'\right)$ lies in $W \times \mathbb{R}^s$ is strictly positive, where s is the number of elements of X_2'. All market characteristics $\left(X_1', X_2'\right)$ in $W \times \mathbb{R}^s$ and the fixed X'' satisfy $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$, by the key Assumption 1. By the arguments at the beginning of the proof, identification has been shown for the case under consideration. The case with $f^0\left(\vec{x}_1\right) < f^0\left(\vec{x}_2\right)$ and $f^1\left(\vec{x}_1\right) > f^1\left(\vec{x}_2\right)$ is similar: just reverse the local production maximization inequalities.

A.7 Theorem 5: Ordinal identification, match characteristics

Let $f^0, f^1 \in \mathscr{F}$, where f^0 is the production function to be identified and f^1 is an alternative where $f^1\left(\vec{x}\right) \neq m \circ f^2\left(\vec{x}\right)$ for all $\vec{x}$ and for any positive monotonic function m. The goal is to show there exists a continuum of X and two assignments A_1 and A_2 where $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$.

Lemma 3 produces $\vec{x}_1$ and $\vec{x}_2$ such that $f^0\left(\vec{x}_1\right) > f^0\left(\vec{x}_2\right)$ and $f^1\left(\vec{x}_1\right) < f^1\left(\vec{x}_2\right)$ or $f^0\left(\vec{x}_1\right) < f^0\left(\vec{x}_2\right)$ and $f^1\left(\vec{x}_1\right) > f^1\left(\vec{x}_2\right)$. Focus on the first case. An inequality such as $f^0\left(\vec{x}_1\right) > f^0\left(\vec{x}_2\right)$ considers a group of matches centered around an upstream firm on the left and another group of matches centered around an upstream firm on the right. This is not a local production maximization inequality (Definition 3), which would require at least two groups, each centered on an upstream firm, on both the left and the right. We will now construct a local production maximization inequality.

For match-specific characteristics, $\vec{x}(i,C^u) = \text{cat}\left(\left(x^{u,d}_{\langle a_1,i\rangle,1},\ldots x^{u,d}_{\langle a_1,i\rangle,K}\right),\ldots,\left(x^{u,d}_{\langle a_m,i\rangle,1},\ldots x^{u,d}_{\langle a_m,i\rangle,K}\right)\right)$. Let $m = |C^u|$ be the number of downstream firms in the set C^u, and hence the number of matches involving upstream firm i. Therefore, an alternative representation of $\vec{x}(i,C^u)$ is as a tuple of vectors rather than a concatenation of vectors (one long vector). For this proof only, let $\vec{x}_1 = \left(\vec{x}^{u,d}_{\langle 1,1\rangle},\ldots,\vec{x}^{u,d}_{\langle m,1\rangle}\right)$, where each $\vec{x}^{u,d}_{\langle 1,1\rangle}$ for $a = 1,\ldots m$ is itself potentially a vector. Likewise, let $\vec{x}_2 = \left(\vec{x}^{u,d}_{\langle 1,2\rangle},\ldots,\vec{x}^{u,d}_{\langle n,2\rangle}\right)$, where upstream firm 2 has n matches, each with a vector of characteristics. To further simplify notation, expand the shorter of the two characteristics collections $\vec{x}_1$ and $\vec{x}_2$ to have the same number of component matches by adding empty sets to the production vector. Call the common number of component matches $h = \max\{m,n\}$. If $m = 2$ and $n = 3$, $\vec{x}_1$ is expanded to be $\left(\vec{x}^{u,d}_{\langle 1,1\rangle},\vec{x}^{u,d}_{\langle 2,1\rangle},\emptyset\right)$.

Starting with an inequality $f(\vec{x}_1) > f(\vec{x}_2)$, we can construct a series $\{\vec{w}_c\}_{c=1}^{h-1}$ of coalition characteristics that add the same terms to both sides of $f(\vec{x}_1) > f(\vec{x}_2)$ to create a local production production maximization inequality of the form

$$f(\vec{x}_1) + \sum_{c=1}^{h-1} f(\vec{w}_c) > f(\vec{x}_2) + \sum_{c=1}^{h-1} f(\vec{w}_c) \tag{36}$$

This inequality will be satisfied for $f = f^0$, and will be satisfied with the $<$ direction for $f = f^1$.

A local production maximization inequality must satisfy Definition 3. The main challenge is that each group characteristic on the right side of the inequality must differ in only one vector of match-specific characteristics from a characteristics vector on the left side. This is because the equilibrium concept of pairwise stability does not allow more than one downstream firm to switch for each upstream firm. To show that (36) is indeed a local production maximization inequality, we need to show that we can pick $\{\vec{w}_c\}_{c=1}^{h-1}$ so that each term on the right side is only one match-specific characteristic vector separate from a term on the left side. The general construction of an $\vec{w}_c$ for $c \le h-1$ is

$$\vec{w}_c = \left(\vec{x}^{u,d}_{\langle 1,1\rangle},\ldots,\vec{x}^{u,d}_{\langle h-c,1\rangle},\vec{x}^{u,d}_{\langle 1,2\rangle},\ldots,\vec{x}^{u,d}_{\langle c,2\rangle}\right).$$

The construction is motivated as follows. From Definition 3, let $B_1 = \{\langle d_0,u_0\rangle,\langle d_1,u_1\rangle,\ldots,\langle d_{h-1},u_{h-1}\rangle\}$ and $B_2 = \{\langle d_1,u_0\rangle,\langle d_2,u_1\rangle,\ldots,\langle d_0,u_{h-1}\rangle\}$. The group centered around upstream firm u_0, with characteristics $\vec{x}\left(u_0,C_{u_0}\right) = \vec{x}_1$, replaces one downstream firm, $d_0 \in C_{u_0}$, with a new firm, $d_1 \in C_{u_1}$. A valid new match-specific value for u_0's new partner d_1 is, by intentional choice, $\vec{x}(u_0,\{d_1\}) = \vec{x}^{u,d}_{\langle 1,2\rangle}$, the first vector in $\vec{x}_2$.[83] This results in a group of matches with characteristics $\vec{w}_1 = \vec{x}\left(u_0,\left(C_{u_0}\setminus\{d_0\}\right)\cup\{d_1\}\right) = \left(\vec{x}^{u,d}_{\langle 1,1\rangle},\ldots,\vec{x}^{u,d}_{\langle h-1,1\rangle},\vec{x}^{u,d}_{\langle 1,2\rangle}\right)$ appearing on the right side of (36). Recall that we need to add the same terms on the left and right sides to move from $f(\vec{x}_1) > f(\vec{x}_2)$ to (36). So we add $f(\vec{w}_1) = f\left(\vec{x}\left(u_1,C_{u_1}\right)\right)$ on the left side. The group centered around upstream firm u_1 replaces one downstream firm, $d_1 \in C_{u_1}$, with $d_2 \in C_{u_2}$. On the right side, $\vec{w}_2 = \vec{x}\left(u_1,\left(C_{u_1}\setminus\{d_1\}\right)\cup\{d_2\}\right) = \left(\vec{x}^{u,d}_{\langle 1,1\rangle},\ldots,\vec{x}^{u,d}_{\langle h-2,1\rangle},\vec{x}^{u,d}_{\langle 1,2\rangle},\vec{x}^{u,d}_{\langle 2,2\rangle}\right)$. As before, $f(\vec{w}_2) = f\left(\vec{x}\left(u_2,C_{u_2}\right)\right)$ appears on the left side as well.

This iterative process truncates. A hypothetical $\vec{w}_h$ equals $\vec{x}_2$, one of the original two vectors from the beginning of the proof. Also, $\vec{x}_1$ equals a hypothetical $\vec{w}_0$, the beginning of the iterative process. The above construction shows that each $\vec{x}(u_c,C_{u_c}) = \vec{w}_c$ on the left side exchanges one downstream firm d_c to yield $\vec{x}(u_c,(C_{u_c}\setminus\{d_c\})\cup\{d_{c+1}\}) = \vec{w}_{c+1}$ on the right side. By inspection, each collection of characteristics $\vec{x}(u_c,(C_{u_c}\setminus\{d_c\})\cup\{d_{c+1}\})$ is different from $\vec{x}(u_c,C_{u_c})$ by the characteristics of one match: $\vec{x}(u_c,\{d_{c+1}\}) = \vec{x}^{u,d}_{\langle c+1,2\rangle}$ instead of $\vec{x}(u_c,\{d_c\}) = \vec{x}^{u,d}_{\langle h-c,1\rangle}$.

[83] Keep in mind that the characteristics are match-specific, so there is no requirement that the characteristics of a firm be the same on the left and right sides.

Therefore, (36) is a valid local production maximization inequality according to Definition 3.

We can embed these groups into a larger matching market. Let B_3 be a larger set of matches that includes the matches corresponding to the downstream firms in $C_0^u \setminus \{d_0\}, \ldots, C_{h-1}^u \setminus \{d_{h-1}\}$. The exact choice of B_3 plays no role in the proof, other than to ensure $C_0^u \setminus \{d_0\}, \ldots, C_{h-1}^u \setminus \{d_{h-1}\}$ are large enough given the number of non-empty elements in an arbitrary $\vec{x}$. Then set $A_1 = B_1 \cup B_3$ and $A_2 = B_2 \cup B_3$. Likewise, let there be some collection X of characteristics for all potential matches, including but not limited to the matches in $A_1 \cup A_2$. The choice of X plays no role in the proof, except that X must contain the h elements $\vec{x}\left(u_0, C_{u_0}\right), \ldots, \vec{x}\left(u_{h-1}, C_{u_{h-1}}\right)$ and the h elements $\vec{x}\left(u_0, \left(C_{u_0} \setminus \{d_0\}\right) \cup \{d_1\}\right), \ldots, \vec{x}\left(u_{h-1}, \left(C_{u_{h-1}} \setminus \{d_{h-1}\}\right) \cup \{d_0\}\right)$, as constructed previously.

Definition 4 requires that we find a set of X with positive probability and two assignments A_1 and A_2 where $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$. Unlike for the cardinal identification theorems, where I assumed all observable characteristics are continuous for transparency, here we assume that there need be only one component of a match-specific characteristics vector, $\vec{x}$, x_1, with continuous and full support. Partition X into (X', X''), where X' has all of the continuously varying characteristics in X (variables whose marginal density conditional on X'' has support equal to a connected subset of the real line) and X'' is the collection of all variables with discrete support in X. We will fix X'' and vary only the elements in X'. Likewise, for each $\vec{x}$, partition it into $\vec{x} = (\vec{x}', \vec{x}'')$, where $\vec{x}'$ has all of the continuous variables in $\vec{x}$ and $\vec{x}''$ has all of the discrete variables. Assume the arguments of f are appropriately reordered to respect this partition. Further, x_1 is the first element of $\vec{x}'$.

Let

$$g\left(\vec{z}_1', \ldots, \vec{z}_{2h}'; f\right) \equiv f\left(\left(\vec{z}_1', \vec{x}_1''\right)\right) + \sum_{c=1}^{h-1} f\left(\left(\vec{z}_{c+1}', \vec{w}_c''\right)\right) - f\left(\left(\vec{z}_{h+1}', \vec{x}_2''\right)\right) - \sum_{c=1}^{h-1} f\left(\left(\vec{z}_{h+1+c}', \vec{w}_c''\right)\right),$$

where $\left(\vec{z}_1', \ldots, \vec{z}_{2h}'\right)$ are arbitrary continuous arguments and $\vec{x}_1''$, $\vec{x}_2''$, and $\vec{w}_1'', \ldots, \vec{w}_{h-1}''$ are the discrete components of the $\vec{x}_1$, $\vec{x}_2$ and $\vec{w}_1, \ldots, \vec{w}_{h-1}$ in (36). As each $f \in \mathscr{F}$ is continuous in its arguments that have continuous support, then so is g. Let v' be the continuous elements of the h elements $\vec{x}\left(u_0, C_{u_0}\right), \ldots, \vec{x}\left(u_{h-1}, C_{u_{h-1}}\right)$ and the h elements $\vec{x}\left(u_0, \left(C_{u_0} \setminus \{d_0\}\right) \cup \{d_1\}\right), \ldots, \vec{x}\left(u_{h-1}, \left(C_{u_{h-1}} \setminus \{d_{h-1}\}\right) \cup \{d_0\}\right)$. Let V^0 be an open set around $g\left(v'; f^0\right)$ where $p > 0$ for $p \in V^0$ and let V^1 be an open set around $g\left(v'; f^1\right)$ where $p < 0$ for $p \in V^1$. Such sets exist by the continuity of g. By the continuity of g and the fact that topologies are closed under finite intersections, $W = g^{-1}\left(V^0; f^0\right) \cap g^{-1}\left(V^1; f^1\right)$ is a nonempty, open neighborhood of v' where $g\left(z; f^0\right) > 0$ and $g\left(z; f^1\right) < 0$ for $z \in W$.

Divide any collection of continuous market characteristics X' into X_1' and X_2', where X_1' is the set comprised of the $2h$ elements of the tuple v' or the $2h$ arguments of the g function. Also, $X_2' = X' \setminus X_1'$, so X' is a one-to-one change of variables (or just notation) of $\left(X_1', X_2'\right)$. By Assumption 4, X_1' has support equal to the product of the marginal supports of its elements. With these notational additions, X is a change of variables from $\left(X_1', X_2', X''\right)$. For any X'', the probability that $\left(X_1', X_2'\right)$ lies in $W \times \mathbb{R}^s$ is strictly positive, where s is the number of elements of X_2'. All market characteristics $\left(X_1', X_2'\right)$ in $W \times \mathbb{R}^s$ and the fixed X'' satisfy $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$, by the key Assumption 1. By the arguments at the beginning of the proof, identification has been shown for the case under consideration. The case with $f^0\left(\vec{x}_1\right) < f^0\left(\vec{x}_2\right)$ and $f^1\left(\vec{x}_1\right) > f^1\left(\vec{x}_2\right)$ is similar: just reverse the local production maximization inequalities.

A.8 Theorem 6: Ordinal identification, firm characteristics

Let $f^0, f^1 \in \mathscr{F}$, where f^0 is the production function to be identified and f^1 is an alternative where $f^1(\vec{x}) \neq m \circ f^0(\vec{x})$ for all $\vec{x}$ and for any positive monotonic function m. The goal is to show there exists a continuum of X and two assignments A_1 and A_2 where $\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$.

Lemma 3 produces $\vec{x}_1$ and $\vec{x}_2$ such that $f^0(\vec{x}_1) > f^0(\vec{x}_2)$ and $f^1(\vec{x}_1) < f^1(\vec{x}_2)$ or $f^0(\vec{x}_1) < f^0(\vec{x}_2)$ and $f^1(\vec{x}_1) > f^1(\vec{x}_2)$. Focus on the first case. An inequality such as $f^0(\vec{x}_1) > f^0(\vec{x}_2)$ considers a group of matches centered around an upstream firm on the left and another group of matches centered around an upstream firm on the right. This is not a local production maximization inequality (Definition 3), which would require at least two groups, each centered on an upstream firm, on both the left and the right. We will now construct a local production maximization inequality.

We need to add the same terms to both sides of the inequality and then argue that the resulting inequality is a local production maximization inequality, where each coalition on the left side is different from a coalition on the right side only in the identity of one downstream firm. The challenge with firm-specific characteristics is that the characteristics of firms remain the same on both sides of the inequality, and different characteristics are in $\vec{x}_1$ and $\vec{x}_2$.

The characteristics are firm specific: $\vec{x}(i, C^u) = \text{cat}\left(\left(x^u_{i,1}, \ldots, x^u_{i,K^u}\right), \left(x^d_{a_1,1}, \ldots, x^d_{a_1,K^d}\right), \ldots, \left(x^d_{a_l,1}, \ldots, x^d_{a_l,K^d}\right)\right)$. In this proof only, I will use the notation $f(\vec{x}_1) = f\left(\vec{x}^u_1, \vec{x}^{d,1}_1, \ldots, \vec{x}^{d,1}_l\right)$ to represent the production of a group of matches with firm characteristics $\vec{x}_1$. Here, $\vec{x}^u_1 = \left(x^u_{1,1}, \ldots, x^u_{1,K^u}\right)$ and $\vec{x}^{d,1}_a = \left(x^d_{a,1}, \ldots, x^d_{a,K^d}\right)$. In other words, each argument of $f\left(\vec{x}^u_1, \vec{x}^{d,1}_1, \ldots, \vec{x}^{d,1}_l\right)$ is a vector of firm-specific characteristics. I put the 1 superscript on these downstream firms to remind us that their characteristics are part of $\vec{x}_1$. Also, let l be the maximum of the number of downstream firms whose characteristics are in $\vec{x}_1$ and $\vec{x}_2$; vectors of empty sets can be added as arguments if the numbers of downstream firms in $\vec{x}_1$ and in $\vec{x}_2$ are not equal. Altogether, $f(\vec{x}_1) = f\left(\vec{x}^u_1, \vec{x}^{d,1}_1, \ldots, \vec{x}^{d,1}_l\right)$ and $f(\vec{x}_2) = f\left(\vec{x}^u_2, \vec{x}^{d,2}_1, \ldots, \vec{x}^{d,2}_l\right)$.

The proposed rewriting of $f(\vec{x}_1) > f(\vec{x}_2)$ to make it a local production maximization inequality by adding the same terms to both sides of the inequality is

$$\begin{aligned} f(\vec{x}_1) + \sum_{a=1}^{l-1} f\left(\vec{x}^u_1, \vec{x}^{d,1}_1, \ldots, \vec{x}^{d,1}_a\right) + f(\vec{x}^u_1) + \sum_{a=1}^{l} f\left(\vec{x}^{d,1}_a\right) + \sum_{a=1}^{l-1} f\left(\vec{x}^u_2, \vec{x}^{d,2}_1, \ldots, \vec{x}^{d,2}_a\right) + f(\vec{x}^u_2) + \sum_{a=1}^{l} f\left(\vec{x}^{d,2}_a\right) > \\ \sum_{a=1}^{l-1} f\left(\vec{x}^u_1, \vec{x}^{d,1}_1, \ldots, \vec{x}^{d,1}_a\right) + f(\vec{x}^u_1) + \sum_{a=1}^{l} f\left(\vec{x}^{d,1}_a\right) + \sum_{a=1}^{l-1} f\left(\vec{x}^u_2, \vec{x}^{d,2}_1, \ldots, \vec{x}^{d,2}_a\right) + f(\vec{x}^u_2) + \sum_{a=1}^{l} f\left(\vec{x}^{d,2}_a\right) + f(\vec{x}_2). \quad (37) \end{aligned}$$

The inequality holds for $f = f^0$ and holds with the opposite sign ($<$) for $f = f^1$.

By inspection, one can loosely verify that (37) is almost, but not quite, a local production maximization inequality, Definition 3, with firm-specific characteristics. The term $\vec{x}_1$ on the left exchanges $\vec{x}^{d,1}_l$ for the option of being unmatched, 0, to add $f\left(\vec{x}^u_1, \vec{x}^{d,1}_1, \ldots, \vec{x}^{d,1}_{l-1}\right) + f\left(\vec{x}^{d,1}_l\right)$ on the right side. Following a pattern, each term $f\left(\vec{x}^u_1, \vec{x}^{d,1}_1, \ldots, \vec{x}^{d,1}_a\right)$ on the left side splits away the term $\vec{x}^{d,1}_a$ to leave a $f\left(\vec{x}^u_1, \vec{x}^{d,1}_1, \ldots, \vec{x}^{d,1}_{a-1}\right) + f\left(\vec{x}^{d,1}_a\right)$ on the right. Each term on the left involving the characteristics originally from $\vec{x}_2$, for example $f\left(\vec{x}^u_2, \vec{x}^{d,2}_1, \ldots, \vec{x}^{d,2}_a\right)$, combines with an unmatched $\vec{x}^{d,1}_{a+1}$ to form $f\left(\vec{x}^u_2, \vec{x}^{d,2}_1, \ldots, \vec{x}^{d,2}_{a+1}\right)$ on the right side.

The inequality (37) is not a local production maximization inequality. For example, look at the terms

$f\left(\vec{x}_1^u\right)+\sum_{a=1}^l f\left(\vec{x}_a^{d,1}\right)$ on the left side. These unmatched firms do not combine with other firms to make pairings on the right side of (37). Therefore, as written (37) is not a local production maximization inequality according to Definition 3. However, the statement of the theorem imposes a non-innocuous localization normalization, which gives $f\left(\vec{x}_1^u\right)+\sum_{a=1}^l f\left(\vec{x}_a^{d,1}\right)=0$ on the left and $f\left(\vec{x}_2^u\right)+\sum_{a=1}^l f\left(\vec{x}_a^{d,2}\right)=0$ on the right. With this change, (37) becomes

$$f(\vec{x}_1)+\sum_{a=1}^{l-1} f\left(\vec{x}_1^u,\vec{x}_1^{d,1},\ldots,\vec{x}_a^{d,1}\right)+\sum_{a=1}^{l-1} f\left(\vec{x}_2^u,\vec{x}_1^{d,2},\ldots,\vec{x}_a^{d,2}\right)+f\left(\vec{x}_2^u\right)+\sum_{a=1}^{l} f\left(\vec{x}_a^{d,2}\right)>$$
$$\sum_{a=1}^{l-1} f\left(\vec{x}_1^u,\vec{x}_1^{d,1},\ldots,\vec{x}_a^{d,1}\right)+f\left(\vec{x}_1^u\right)+\sum_{a=1}^{l} f\left(\vec{x}_a^{d,1}\right)+\sum_{a=1}^{l-1} f\left(\vec{x}_2^u,\vec{x}_1^{d,2},\ldots,\vec{x}_a^{d,2}\right)+f(\vec{x}_2), \quad (38)$$

which by the above informal arguments is a local production maximization inequality. I intentionally do not remove from (37) all production functions with zero production. Even though the production $f\left(\vec{x}_1^u\right)$ of singleton matches is zero, these production functions are needed to show (38) satisfies the definition of a local production maximization inequality, Definition 3.

The above arguments were informal. I will now formally show that (38) satisfies Definition 3. There are $3l$ terms on the left side of (38). The number $3l$ explains the statement in the theorem, "Further, let there be assignments A that contain as many matched coalitions as three times the maximum quota of an upstream firm." Let $B_1=\{\langle d_1,u_1\rangle,\ldots,\langle d_{3l},u_{3l}\rangle\}$, where the indexing $\langle d_c,u_c\rangle$ follows the order on the left side of (38), from left to right. As I will show, many of these match partners will be 0, representing being unmatched. Now let

$$B_2=\{\langle d_{l+1},u_1\rangle,\langle d_{l+2},u_2\rangle,\ldots,\langle d_{2l},u_l\rangle,\langle d_{2l+2},u_{l+1}\rangle,\ldots,\langle d_{3l},u_{2l-1}\rangle,\langle d_{2l+1},u_{2l}\rangle,\langle d_1,u_{2l+1}\rangle,\ldots,\langle d_l,u_{3l}\rangle\}.$$

The match $\langle d_{l+1},u_1\rangle\in B_2$ means that the upstream firm u_1, which on the left side has characteristics $\vec{x}\left(u_1,C_{u_1}\right)=\vec{x}_1$, exchanges a downstream firm d_1 for the downstream firm $d_{l+1}\in C_{u_{l+1}}$. In this case, downstream firm d_1 has the characteristics $\vec{x}_l^{d,1}$ while d_{l+1} is actually a dummy partner, 0, representing being unmatched. For each index $c=1,\ldots,3l$, Table 5 lists the upstream firm characteristics, the characteristics for the group of all firms u_c and C_{u_c}, downstream firm d_c's characteristics for the match in B_1, the downstream firm partner in the permutation π creating B_2, the characteristics of that downstream firm partner, and the characteristics of the entire group of all firms u_c and its downstream firm partners after the switch. One can verify that the characteristics of the firm $\pi\langle d_c,u_c\rangle$ in the fifth column are always the same as the characteristics of that downstream firm in the third column. This is the key idea behind showing that (38) is a local production maximization inequality with firm-specific characteristics: the characteristics of downstream firms remain the same after the permutation of partners between B_1 and B_2.

For those readers who have read the earlier proofs, the following steps follow similar logical arguments as before. We can embed these four groups into a larger matching market. Let B_3 be a larger set of matches that includes the matches corresponding to the downstream firms in $C_0^u\backslash\{d_0\},\ldots,C_{h-1}^u\backslash\{d_{h-1}\}$. The exact choice of B_3 plays no role in the proof, other than to ensure $C_0^u\backslash\{d_0\},\ldots,C_{h-1}^u\backslash\{d_{h-1}\}$ are large enough given the number of non-empty elements in an arbitrary $\vec{x}$. Then set $A_1=B_1\cup B_3$ and $A_2=B_2\cup B_3$. Likewise, let there be some collection X of characteristics for all the firms with matches in A_1, and, by implication, all firms with matches in A_2. The choice of X plays no further role in the proof.

Definition 4 requires that we find a set of X with positive probability and two assignments A_1 and A_2 where

Table 5: Proof of Theorem 6: Demonstrating That (38) Is a Local Production Maximization Inequality

	(1)	(2)	(3)	(4)	(5)	(6)
Index c	$\vec{x}(u_c,\{0\})$	$\vec{x}(u_c,C_{u_c})$	$\vec{x}(0,\{d_c\})$	$\pi\langle d_c,u_c\rangle$	$\vec{x}(0,\{\pi\langle d_c,u_c\rangle\})$	$\vec{x}(u_c,(C_{u_c}\setminus\{d_c\})\cup\{\pi\langle d_c,u_c\rangle\})$
1	$\vec{x}^u_1$	$\vec{x}_1$	$\vec{x}^{d,1}_l$	d_{l+1}	$\emptyset$	$\left(\vec{x}^u_1,\vec{x}^{d,1}_1,\ldots,\vec{x}^{d,1}_{l-1}\right)$
2	$\vec{x}^u_1$	$\left(\vec{x}^u_1,\vec{x}^{d,1}_1\right)$	$\vec{x}^{d,1}_1$	d_{l+2}	$\emptyset$	$\left(\vec{x}^u_1\right)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
l	$\vec{x}^u_1$	$\left(\vec{x}^u_1,\vec{x}^{d,1}_1,\ldots,\vec{x}^{d,1}_{l-1}\right)$	$\vec{x}^{d,1}_{l-1}$	d_{2l}	$\emptyset$	$\left(\vec{x}^u_1,\vec{x}^{d,1}_1,\ldots,\vec{x}^{d,1}_{l-2}\right)$
$l+1$	$\vec{x}^u_2$	$\left(\vec{x}^u_2,\vec{x}^{d,2}_1\right)$	$\emptyset$	d_{2l+2}	$\vec{x}^{d,2}_2$	$\left(\vec{x}^u_2,\vec{x}^{d,2}_1,\vec{x}^{d,2}_2\right)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2l-1$	$\vec{x}^u_2$	$\left(\vec{x}^u_2,\vec{x}^{d,2}_1,\ldots,\vec{x}^{d,2}_{l-1}\right)$	$\emptyset$	d_{3l}	$\vec{x}^{d,2}_l$	$\vec{x}_2=\left(\vec{x}^u_2,\vec{x}^{d,2}_1,\ldots,\vec{x}^{d,2}_l\right)$
$2l$	$\vec{x}^u_2$	$\left(\vec{x}^u_2\right)$	$\emptyset$	d_{2l+1}	$\vec{x}^{d,2}_1$	$\left(\vec{x}^u_2,\vec{x}^{d,2}_1\right)$
$2l+1$	$\emptyset$	$\left(\vec{x}^{d,2}_1\right)$	$\vec{x}^{d,2}_1$	d_1	$\vec{x}^{d,1}_l$	$\left(\vec{x}^{d,1}_l\right)$
$2l+2$	$\emptyset$	$\left(\vec{x}^{d,2}_2\right)$	$\vec{x}^{d,2}_2$	d_2	$\vec{x}^{d,1}_1$	$\left(\vec{x}^{d,1}_1\right)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$3l$	$\emptyset$	$\left(\vec{x}^{d,2}_l\right)$	$\vec{x}^{d,2}_l$	d_l	$\vec{x}^{d,1}_{l-1}$	$\left(\vec{x}^{d,1}_{l-1}\right)$

$\Pr\left(A_1 \mid X; f^0, S^0\right) > \Pr\left(A_2 \mid X; f^0, S^0\right)$ while $\Pr\left(A_1 \mid X; f^1, S^1\right) < \Pr\left(A_2 \mid X; f^1, S^1\right)$ for any $S_1 \in \mathscr{S}$. Unlike for the cardinal identification theorems, where I assumed all observable characteristics are continuous for transparency, here we assume that there need be only one component of a group-specific characteristics vector, $\vec{x}$, x_1, with continuous and full support. Partition X into (X', X''), where X' has all of the continuously varying characteristics in X (variables whose marginal density conditional on X'' has support equal to a connected subset of the real line) and X'' is the collection of all variables with discrete support in X. We will fix X'' and vary only the elements in X'. Likewise, for each $\vec{x}$, partition it into $\vec{x} = (\vec{x}', \vec{x}'')$, where $\vec{x}'$ has all of the continuous variables in $\vec{x}$ and $\vec{x}''$ has all of the discrete variables. Assume the arguments of f are appropriately reordered to respect this partition. Further, x_1 is the first element of $\vec{x}'$.

There is a total of $6l$ firms, $3l$ upstream and $3l$ downstream, in the matches in the set B_1, and hence B_2 as well. As in previous proofs, I can define a function $g\left(\vec{z}'_1, \ldots, \vec{z}'_{6l}; f\right)$ to be a function of the $6l$ vectors of corresponding continuous firm characteristics $\vec{z}'_1$. By convention, let the first $3l$ vectors of continuous characteristics correspond to downstream firms and let the second $3l$ vectors correspond to upstream firms. Because of the complexity of (38), to save space I will not introduce new notation to explicitly write out $g\left(\vec{z}'_1, \ldots, \vec{z}'_{6l}; f\right)$. It is clear from previous proofs that the following can be done. I can let $g\left(\vec{z}'_1, \ldots, \vec{z}'_{6l}; f\right)$ be inspired by the right side of (38) subtracted from the left side. Each firm-specific characteristics vector for a downstream firm $a \in \{1, \ldots 3l\}$ is evaluated at some $\left(\vec{z}'_a, \vec{x}^{d,''}_a\right)$, where $\vec{z}'_a$ is one of the first $3l$ arguments of $g\left(\vec{z}'_1, \ldots, \vec{z}'_{6l}; f\right)$ and $\vec{x}^{d,''}_a$ is the non-continuous characteristics of the corresponding firm in (38). Likewise, each upstream firm vector i is evaluated at some $\left(\vec{z}'_i, \vec{x}^{u,''}_i\right)$, where $\vec{z}'_i$ is one of the second $3l$ arguments of $g\left(\vec{z}'_1, \ldots, \vec{z}'_{6l}; f\right)$ and $\vec{x}^{u,''}_i$ is the vector of non-continuous characteristics for the corresponding upstream firm in (38). As in previous proofs, as each $f \in \mathscr{F}$ is continuous in its arguments that have continuous support, then so is g. Let v' be the continuous elements of the $6l$ firm-specific characteristics in (38). Let V^0 be an open set around $g\left(v'; f^0\right)$ where $p > 0$ for $p \in V^0$ and let V^1 be an open set around $g\left(v'; f^1\right)$ where $p < 0$ for $p \in V^1$. Such sets exist by the continuity of g. By the continuity of g and the fact that topologies are closed under finite intersections, $W = g^{-1}(V^0; f^0) \cap g^{-1}(V^1; f^1)$ is a nonempty, open neighborhood of v' where $g(z; f^0) > 0$ and $g(z; f^1) < 0$ for $z \in W$.

Divide any collection of continuous market characteristics X' into X_1' and X_2', where X_1' is the set comprised of the $6l$ elements of the tuple v' or the $6l$ arguments of the g function. Also, $X_2' = X' \backslash X_1'$, so X' is a one-to-one change of variables (or just notation) of (X_1', X_2'). By Assumption 4, X_1' has support equal to the product of the marginal supports of its elements. With these notational additions, X is a change of variables from (X_1', X_2', X''). For any X'', the probability that (X_1', X_2') lies in $W \times \mathbb{R}^s$ is strictly positive, where s is the number of elements of X_2'. All market characteristics (X_1', X_2') in $W \times \mathbb{R}^s$ and the fixed X'' satisfy $\Pr(A_1 \mid X; f^0, S^0) > \Pr(A_2 \mid X; f^0, S^0)$ while $\Pr(A_1 \mid X; f^1, S^1) < \Pr(A_2 \mid X; f^1, S^1)$ for any $S_1 \in \mathscr{S}$, by the key Assumption 1. By the arguments at the beginning of the proof, identification has been shown for the case under consideration. The case with $f^0(\vec{x}_1) < f^0(\vec{x}_2)$ and $f^1(\vec{x}_1) > f^1(\vec{x}_2)$ is similar: just reverse the local production maximization inequalities.

A.9 Theorem 7: Consistency as $M \to \infty$

A.9.1 Constructive identification

Constructive identification follows from the identification theorems. By a law of large numbers and the law of iterated expectations, the probability limit of the maximum score objective function is

$$Q_\infty(\beta) = E_X \left\{ \sum_{A \in \mathscr{A}(X)} \sum_{(B_1,B_2) \in N(A,X)} \Pr(A \mid X) I(B_1, B_2 \mid X) 1\left[\sum_{\langle a,i \rangle \in B_1} f_\beta(\vec{x}(i, C_i^u(A))) > \sum_{\langle a,i \rangle \in B_2} f_\beta(\vec{x}(i, C_i^u((A \backslash B_1) \cup B_2))) \right] \right\},$$

where $\mathscr{A}(X)$ is the set of feasible assignments given X and $\Pr(A \mid X) = \Pr\left(A_1 \mid X; f_{\beta^0}, S^0\right)$.

For each pair of an assignment $A_1 \in \mathscr{A}(X)$ and a $(B_1, B_2) \in N(A_1, X)$ in the integrand above, there is an assignment $A_2 \in \mathscr{A}(X)$ that is $A_2 = (A_1 \backslash B_1) \cup B_2$. An inequality for A_1 and (B_1, B_2) is mutually exclusive with a paired inequality for A_2 and (B_2, B_1). As ties occur with probability 0, with probability 1 either the indicator with A_1 or the indicator with A_2 will be 1, and the other 0. By Assumption 6, $I(B_1, B_2 \mid X) = I(B_2, B_1 \mid X)$. The ranking of the weights on the indicators reduces to comparing $\Pr\left(A_1 \mid X; f_{\beta^0}, S^0\right)$ and $\Pr\left(A_2 \mid X; f_{\beta^0}, S^0\right)$. By the rank order property, all parameters in the identified set make the inequality (of the pair) with the highest weights satisfied and therefore globally maximize $Q_\infty(\beta)$.

Let $\beta^1 \in \mathscr{B}$ be some parameter vector where $\beta^1 \neq \beta^0$. By Assumption 5, there exist a set $\mathscr{X}$ of X with positive probability and two assignments $A_1$and A_2 such that $\Pr\left(A_1 \mid X; f_{\beta^0}, S^0\right) > \Pr\left(A_2 \mid X; f_{\beta^0}, S^0\right)$ while $\Pr\left(A_1 \mid X; f_{\beta^1}, S^1\right) < \Pr\left(A_2 \mid X; f_{\beta^1}, S^1\right)$ for any $X \in \mathscr{X}$, for any $S^1 \in \mathscr{S}$. Considering all the $X \in \mathscr{X}$, $Q_\infty(\beta^1) < Q_\infty(\beta^0)$ because β^1 causes inequalities with the lower of $\Pr\left(A_1 \mid X; f_{\beta^0}, S^0\right)$ and $\Pr\left(A_2 \mid X; f_{\beta^0}, S^0\right)$ to enter the objective function.

A.9.2 Continuity of the limiting objective function and uniform convergence

Lemma 2.4 from Newey and McFadden (1994) can be used to prove continuity of $Q_\infty(\beta)$ as well as uniform in probability convergence of $Q_M(\beta)$ to $Q_\infty(\beta)$. Remember that the asymptotics are in the number of markets. The conditions of Lemma 2.4 are that the data (across markets) are iid, which can hold even if we view the number of upstream and downstream firms as random; that the parameter space $\mathscr{B}$ is compact (Assumption 5), that the terms for each market are continuous with probability 1 in β; and that the terms for each market are bounded by a function whose mean is not infinite. While the terms for each market are not continuous in β because of the indicator functions, they are continuous with probability 1 by Assumption 5, as each $\vec{x}$ in X has some elements with continuous support. The value of the objective function for a given market is bounded by the number of inequalities, which is finite.

A.10 Theorem 8: Consistency as $H \to \infty$

The consistency proof is very similar to the proof of Theorem 7 in Appendix A.9. Rather than repeat all the arguments, I will focus on proving that β^0 is a global maximum of the probability limit of (17),

$$Q(\beta) = \int 1\left[f_\beta(\vec{x}_1^m, \vec{x}_2^w) + f_\beta(\vec{x}_3^m, \vec{x}_4^w) > f_\beta(\vec{x}_1^m, \vec{x}_4^w) + f_\beta(\vec{x}_3^m, \vec{x}_2^w)\right] \cdot g_{m,w}^{\beta^0, S^0}(\langle \vec{x}_1^m, \vec{x}_2^w \rangle) \cdot g_{m,w}^{\beta^0, S^0}(\langle \vec{x}_3^m, \vec{x}_4^w \rangle)\, d(\vec{x}_1^m, \vec{x}_2^w, \vec{x}_3^m, \vec{x}_4^w).$$

Let V be all $(\vec{x}_1^m, \vec{x}_2^w, \vec{x}_3^m, \vec{x}_4^w)$ such that $f_{\beta^0}(\vec{x}_1^m, \vec{x}_2^w) + f_{\beta^0}(\vec{x}_3^m, \vec{x}_4^w) > f_{\beta^0}(\vec{x}_1^m, \vec{x}_4^w) + f_{\beta^0}(\vec{x}_3^m, \vec{x}_2^w)$. For each $(\vec{x}_1^m, \vec{x}_2^w, \vec{x}_3^m, \vec{x}_4^w) \in V$ there is a mutually exclusive inequality $(\vec{x}_1^m, \vec{x}_4^w, \vec{x}_3^m, \vec{x}_2^w) \notin V$. By Assumption 7,

$$Q\left(\beta^0\right) = \int_V g_{m,w}^{\beta^0, S^0}(\langle \vec{x}_1^m, \vec{x}_2^w \rangle) \cdot g_{m,w}^{\beta^0, S^0}(\langle \vec{x}_3^m, \vec{x}_4^w \rangle)\, d(\vec{x}_1^m, \vec{x}_2^w, \vec{x}_3^m, \vec{x}_4^w) > \int_V g_{m,w}^{\beta^0, S^0}(\langle \vec{x}_1^m, \vec{x}_4^w \rangle) \cdot g_{m,w}^{\beta^0, S^0}(\langle \vec{x}_3^m, \vec{x}_2^w \rangle)\, d(\vec{x}_1^m, \vec{x}_2^w, \vec{x}_3^m, \vec{x}_4^w).$$

by a pointwise comparison of each of the terms in the two integrands. Let V^C be all inequalities not in V. If a positive measure of inequalities are exchanged between V and V^C to create $\bar{V}$ and $\bar{V}^C$, then by the mutual exclusivity and Assumption 7,

$$\int_V g_{m,w}^{\beta^0, S^0}(\langle \vec{x}_1^m, \vec{x}_2^w \rangle) \cdot g_{m,w}^{\beta^0, S^0}(\langle \vec{x}_3^m, \vec{x}_4^w \rangle)\, d(\vec{x}_1^m, \vec{x}_2^w, \vec{x}_3^m, \vec{x}_4^w) > \int_{\bar{V}} g_{m,w}^{\beta^0, S^0}(\langle \vec{x}_1^m, \vec{x}_2^w \rangle) \cdot g_{m,w}^{\beta^0, S^0}(\langle \vec{x}_3^m, \vec{x}_4^w \rangle)\, d(\vec{x}_1^m, \vec{x}_2^w, \vec{x}_3^m, \vec{x}_4^w).$$

Therefore, β^0 is a global maximum of $Q(\beta)$. The identification portion of Assumption 8 can be used to show that β^0 is the unique global maximum, as in the proof of Theorem 7.

B Empirical: Allowing the Asian indicator to be recomputed in inequalities

In Section 9.2 and the results in Table 4, I did not recompute the measure of being a supplier to Asian assemblers on the right side of the inequalities. This section explores specifications that do recompute the Asian supplier measure for counterfactual sets of matches. I also explore why the point estimates differ between the specifications where the measure of being a supplier to Asian assemblers is not and is recomputed.

The previous production function specification was (16), where $x^{\text{SupplierToAsian}}\left(C_i^u(A_m)\right)$ is a function of only the actual assignment, A_m, and is not recomputed when C^u changes on the right side of the inequality. Table 6 reports estimates of the matching model using the production function

$$f_\beta(\vec{x}(i, C^u)) = \beta_{\text{Cont.}} x^{\text{Continent}}(C^u) + \beta_{\text{PG}} x^{\text{ParentGroup}}(C^u) + \beta_{\text{Brand}} x^{\text{Brand}}(C^u) + \beta_{\text{Car}} x^{\text{Car}}(C^u) + \beta_{\text{AsianCont.}} x^{\text{Continent}}(C^u)\, x^{\text{SupplierToAsian}}(C^u),$$

Table 6: Supplier competitive edge from supplying Asian assemblers: Supplier to Asian assembler measure recomputed for counterfactual matches

HHI Measure	Estimate	95% CI	Estimate	95% CI
Continent	+1	Superconsistent	+1	Superconsistent
Parent Group	6.69	(6.31, 9.89)	6.35	(5.59, 9.28)
Brand	8.59	(1.28, 12.6)	9.67	(5.66, 14.1)
Model	116	(128, 172)	95.0	(87.7, 139)
Continent * Asian Dummy	-0.0519	(-0.0811, 0.552)		
Continent * Asian %			-0.0356	(-1.55, 0.0169)
# Inequalities	532,939		532,939	
% Satisfied	0.753		0.753	

where now $x^{\text{SupplierToAsian}}(C^u)$ is recomputed each time $C^u \subseteq D$ changes. When I do allow the Asian supplier measure to be recomputed, the coefficient $\beta_{\text{AsianCont.}}$ is zero in terms of its economic magnitude, for both the dummy and market share measures of being an Asian supplier. In the previous Table 4, the point estimates for the Asian supplier measures were -1.09 for the indicator and -5.30 for the continuous market share measure. Compared to these, the point estimates of -0.0519 and -0.0356 in Table 6 are economically small and have confidence regions that lead to the rejection of null hypotheses of large in absolute value, negative coefficients for $\beta_{\text{AsianCont.}}$.[84]

I spent some time investigating why the point estimates for $\beta_{\text{AsianCont.}}$ varied across so much across Tables 4 and 6. Here I focus on the specification with the indicator variable measure of being a supplier to an Asian assembler. A local production maximization inequality used in estimation looks like

$$f_\beta\left(\vec{x}\left(i, C_i^u(A_m)\right)\right) + f_\beta\left(\vec{x}\left(j, C_j^u(A_m)\right)\right) > f_\beta\left(\vec{x}\left(i, C_i^u\left((A_m \setminus \{a\}) \cup \{b\}\right)\right)\right) + f_\beta\left(\vec{x}\left(j, C_j^u\left((A_m \setminus \{b\}) \cup \{a\}\right)\right)\right),$$

for the matches of car parts and suppliers $\langle a, i\rangle$ and $\langle b, j\rangle$. On the left side are actual matches from the data; the counterfactual matches are on the right. The indicator variable $x^{\text{SupplierToAsian}}(C^u)$ is either 0 or 1 for each of the four matches, so the values of $x^{\text{SupplierToAsian}}(C^u)$ for an inequality can be written as, for example, $\{1,1\} \to \{1,0\}$.[85] This notation means that, in the data, both upstream firms i and j supply at least one Asian assembler each. After the exchange of partners, one of i and j does not serve an Asian assembler any more. Incidentally, this can only occur if one of i and j produces only one Asian car part in component category m, in the data. By contrast, the specification without recomputing the Asian dummy would be $\{1,1\} \to \{1,1\}$, as a firm's Asian supplier status is a fixed firm characteristic. The two main types of possibilities for an inequality with some change in $x^{\text{SupplierToAsian}}(C^u)$ are $\{1,1\} \to \{1,0\}$ and $\{1,0\} \to \{1,1\}$. The case $\{1,0\} \to \{1,1\}$ occurs when a supplier with two or more Asian-assembler car parts exchanges one of those car parts with a supplier that supplies, in the data, no car parts to Asian assemblers.

Through some exploratory empirical work, I found that exchanges of the form $\{1,1\} \to \{1,0\}$ were driving the differences in the point estimates.[86] To confirm this, I created an artificial set of inequalities equal to the

[84] The confidence regions for the coefficient on the HHI specialization measure at the model level for the Asian dummy specification do not include the point estimate for β_{Car}. This can occur with subsampling, the method used for inference here.

[85] The notation uses sets instead of tuples because the order of the production functions is not recorded.

[86] I changed each of several types of inequalities from their Table 6 back to their Table 4 forms, and evaluated the objective function at the Table 4 estimates. I then looked at the number of satisfied inequalities (a measure of statistical fit), and found that the $\{1,1\} \to \{1,0\}$ inequalities were the most instrumental in increasing the statistical fit.

Table 7: Artificial inequalities: Reconciling different point estimates between Tables 4 and 6

HHI Measure	Estimate	95% CI
Continent	+1	Superconsistent
Parent Group	4.55	(3.79, 6.54)
Brand	6.94	(5.14, 10.5)
Model	73.1	(78.1, 105)
Continent * Asian Dummy	-1.03	(-1.05, -0.947)
Continent * Asian %		
# Inequalities	532,939	
% Satisfied	0.761	

inequalities used in Table 6, except that 10,358 inequalities (out of the 532,939 total inequalities) of the form $\{1,1\} \to \{1,0\}$ were replaced by the corresponding inequalities from Table 4, where the Asian indicator is not recomputed.[87] The estimates are in Table 7. We can see that the point estimates for the HHI specialization measures are in between those in Tables 4 and 6. Further, the point estimate for $\beta_{\text{AsianCont.}}$ on the interaction $x^{\text{Continent}}(C^u)\,x^{\text{SupplierToAsian}}(C^u)$, the key variable being altered, is -1.03, similar to the value of -1.09 in Table 4.

The exercise in Table 7 confirms the proposition that inequalities where two suppliers exchange car parts and one supplier ceases to be an Asian supplier drive the drop in the estimate of $\beta_{\text{AsianCont.}}$ from -1.09 in Table 4 to -0.000641 in Table 6. I will now take a speculative stab at offering an economic story to explain the point estimates. When $x^{\text{SupplierToAsian}}(C_i^u(A_m))$ is not recomputed for counterfactual matches, the inequalities answer the questions discussed in Section 9.2: do suppliers to Asian brands have some competitive edge with non-Asian assemblers? When $x^{\text{SupplierToAsian}}(C_i^u)$ is recomputed for counterfactual C_i^u's, in addition the inequalities ask why more suppliers are not supplying Asian assemblers if there is some competitive advantage from doing so? The tendency would be for the terms $\{1,1\}$ in the key inequalities $\{1,1\} \to \{1,0\}$ to be given a positive weight $\beta_{\text{AsianCont.}}$, which counteracts the -1.09 coefficient for $\beta_{\text{AsianCont.}}$ found in Table 4. In Table 6, the model deals with these two opposing forces by setting the coefficient on $\beta_{\text{AsianCont.}}$ to be near zero. The inequalities in Table 4 are easier to understand and interpret because the estimate for the parameter $\beta_{\text{AsianCont.}}$ reflects a fixed firm-specific characteristic $x^{\text{SupplierToAsian}}(C_i^u(A_m))$ that represents only one economic phenomenon, the competitive edge of suppliers to Asian brands.[88]

[87] My goal was to find the minimum set of inequalities that could change and restore the point estimates of Table 4. There were 10,603 inequalities of the form $\{1,1\} \to \{1,0\}$ in the dataset behind Table 6. I modified only 10,358 inequalities. When evaluated at the point estimates from Table 4 (except for $\beta_{\text{AsianCont.}}$), 5992 of the 10,358 inequalities in question switch from providing a lower bound for $\beta_{\text{AsianCont.}}$ (as in $\beta_{\text{AsianCont.}} > z$) to providing an upper bound for $\beta_{\text{AsianCont.}}$ (as in $\beta_{\text{AsianCont.}} < z$). The remaining 4366 of the inequalities keep a lower bound for $\beta_{\text{AsianCont.}}$, even after the switch. The value of the lower bound z does change. I did not modify the 245 other inequalities of the form $\{1,1\} \to \{1,0\}$ where some other sort of change in whether an inequality provided a lower or upper bound for $\beta_{\text{AsianCont.}}$ occurred, when the inequalities are evaluated at the point estimates from Table 4.

[88] More mechanically, when the interaction term is $x^{\text{Continent}}(C^u)\,x^{\text{SupplierToAsian}}(C_i^u(A_m))$, the only variation in this term between the left and right sides of a local production maximization inequality comes from changes in the HHI specialization measure, $x^{\text{Continent}}(C^u)$, from one car part being exchanged. For a company that supplies several car parts in this matching market, $x^{\text{Continent}}(C^u)$ and hence the interaction change by a relatively small amount. When $x^{\text{SupplierToAsian}}(C^u)$ is recomputed for counterfactual downstream firm partners C_i^u, then firms that supply only one or two car parts to Asian assemblers have a relatively large change in $x^{\text{SupplierToAsian}}(C^u)$ and hence in the interaction $x^{\text{Continent}}(C^u)\,x^{\text{SupplierToAsian}}(C^u(A_m))$. It is unsurprising that the Table 4 specification with relatively small changes in the interaction has a correspondingly large (in absolute value) estimate of $\beta_{\text{AsianCont.}}$ compared to the estimate for the Table 6 specification with relatively large changes in the interaction.

References

Abrevaya, Jason and Jian Huang, "On the bootstrap of the maximum score estimator," *Econometrica*, July 2005, *73* (4), 1175–1204.

Andrews, Donald, Steven T. Berry, and Panle Jia, "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Store Locations," 2004. working paper.

Atakan, Alp E., "Assortative matching with explicit search costs," *Econometrica*, May 2006, *74* (3), 667–680. forthcoming.

__ , "Efficient matching with explicit search frictions," February 2007. working paper.

Bajari, Patrick and Jeremy T. Fox, "Measuring the Efficiency of an FCC Spectrum Auction," 2007. working paper.

__ , C. Lanier Benkard, and Jonathan Levin, "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, September 2007, *75* (5), 1331–1370.

__ , Han Hong, and Stephen Ryan, "Identification and Estimation of a Discrete Game of Complete Information," May 2007. working paper.

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

Beresteanu, Arie, Ilya Molchanov, and Francesca Molinari, "Sharp Identification Regions in Games," June 2008. working paper.

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

__ , "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 1994, *25* (2), 242–262.

__ , 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.

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

Camiña, Ester, "A Generalized Assignment Game," *Mathematical Social Sciences*, 2006, *52* (2), 152–161.

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

Ciliberto, Federico and Elie Tamer, "Market Structure and Multiple Equilibria in Airline Markets," October 2007. working paper.

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

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.

Dyer, Jeffrey H., "Specialized Supplier Networks as a Source of Competitive Advantage: Evidence from the Auto Industry," *Strategic Management Journal*, 1996, *17* (4), 271–291.

_ , "Effective interim collaboration: how firms minimize transaction costs and maximise transaction value," *Strategic Management Journal*, 1997, *18* (7), 535–556.

Echenique, Federico, "What Matchings Can Be Stable? The Testable Implications of Matching Theory," November 2006. working paper.

Fox, Jeremy T., "Semiparametric estimation of multinomial discrete-choice models using a subset of choices," *RAND Journal of Economics*, Winter 2007, *38* (4), 1002–1019.

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

Garicano, Luis and Esteban Rossi-Hansberg, "Organization and Inequality in a Knowledge Economy," *The Quarterly Journal of Economics*, November 2006, pp. 1383–1435.

Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey, "Regular Quantal Response Equilibrium," *Experimental Economics*, December 2005, *8*, 347–367.

Gordon, Nora and Brian Knight, "The Causes of Political Integration: An Application to School Districts," February 2006. working paper.

Gourieroux, Christian and Alain Monfort, *Statistics and Econometric Models*, Vol. 1 of *Themes in Modern Econometrics*, Cambridge, 1995.

Haile, Phil and Elie Tamer, "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, February 2003, *111* (1), 1–51.

Han, Aaron K., "Nonparametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 1987, *35*, 303–316.

Hatfield, John William and Fuhito Kojima, "Matching with Contracts: Comment," *American Economic Review*, June 2008, *98* (3), 1189–1194.

— **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,” March 2008. 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.

Jaume, Daniel, Jordi Massó, and Alejandro Neme, “The Multiple-partners Assignment Game with Heterogeneous Sells and Multi-unit Demands: Competitive Equilibria,” December 2007. working paper.

Jehle, Geoffrey A. and Philip J. Reny, *Advanced Microeconomic Theory*, Addison Wesley Longman, 2000.

Jia, Panle, “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry,” *Econometrica*, 2008. forthcoming.

Kamath, Rajan R. and Jeffrey K. Liker, “A Second Look at Japanese Product Development,” *Harvard Business Review*, November-December 1994.

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.

Langfield-Smith, K. and M.R. Greenwood, “Developing Co-operative Buyer-Supplier Relationships: A Case Study of Toyota,” *Journal of Management Studies*, 1998, *35* (3), 331–353.

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.

Liker, Jeffrey K. and Yen-Chun Wu, “Japanese Automakers, U.S. Suppliers and Supply-Chain Superiority,” *Sloan Management Review*, Fall 2000.

Lucas, Robert E., "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 1978, *9* (2), 508–523.

Magnac, Thierry and Eric Maurin, "Identification and information in monotone binary models," *Journal of Econometrics*, 2007.

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

_ , "Semiparametric Analysis of Discrete Response," *Journal of Econometrics*, 1985, *27*, 313–333.

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

_ **and T. Scott Thompson**, "Operational Characteristics of Maximum Score Estimation," *Journal of Econometrics*, 1986, pp. 85–108.

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

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

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

McFadden, Daniel L., "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press, 1973.

Milgrom, Paul and Chris Shannon, "Monotone Comparative Statics," *Econometrica*, 1994, *62* (1), 157–180.

Monteverde, Kirk and David J. Teece, "Supplier Switching Costs and Vertical Integration in the Automobile Industry," *The Bell Journal of Economics*, Spring 1982, *13* (1), 206–213.

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.

Novak, Sharon and Birger Wernerfelt, "On the Grouping of Tasks into Firms: Make-or-Buy with Interdependent Parts," 2007. working paper.

_ **and Scott Stern**, "Complementarity Among Vertical Integration Decisions: Evidence from Automobile Product Development," 2007. working paper.

_ **and** _ , "How Does Outsourcing Affect Performance Dynamics? Evidence from the Automobile Industry," 2007. working paper.

_ **and Steven D. Eppinger**, "Sourcing by Design: Product Complexity and the Supply Chain," *Management Science*, 2001, *47* (1), 189–204.

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

Pakes, Ariel, "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 1986, *54* (4), 755–784.

__ **, Jack Porter, Katherine Ho, and Joy Ishii**, "Moment Inequalities and Their Application," 2006. working paper.

Pinkse, C.A.P., "On the computation of semiparametric estimates in limited dependent variable models," *Journal of Econometrics*, 1993, pp. 185–205.

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.

Pycia, Marek, "Many-to-One Matching with Complementarities and Peer Effects," April 2007. working paper.

__ , "Non-Existence Result for Matching Games with Transfers," March 2008. working paper.

Rosen, Sherwin, "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, 1982, *13* (2), 311–323.

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

Rust, John, "Optimal Replacement of the GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 1987.

Sattinger, Michael, "Differential Rents and the Distribution of Earnings," *Oxford Economic Papers*, March 1979, *31* (1), 60–71.

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

Sherman, Robert P., "The Limiting Distribution of the Maximum Rank Correlation Estimation," *Econometrica*, January 1993, *61* (1), 123–137.

__ , "Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators," *The Annals of Statistics*, March 1994, *22* (1), 439–459.

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

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

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.

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

Tinbergen, Jan, "Some Remarks on the Distribution of Labour Incomes," in "Proceedings of the Royal Netherlands Academy of Sciences," International Economic Papers (1951, No 1), 1947.

Weiss, Yoram, "Sharing Rules," 2007. unpublished book chapter.