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, Bath, 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, Revised August 2009
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 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, production levels, or unmatched agents. 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. Koopmans and Beckmann (1957) and Shapley and Shubik (1972) theoretically analyze the same model of one-to-one, two-sided matching. 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 the characteristics of 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 assemblers. 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).

Match data come from the outcome to a market, which intermingles the preferences of all participating agents and finds an equilibrium. Agents on the same side of the market (say men) are rivals to match with agents on the other side of the market (say women). The equilibrium concept in a matching game is the cooperative solution concept known as pairwise stability. This paper studies matching games where matched parties can endogenously exchange monetary transfers in equilibrium. At a pairwise stable equilibrium, without loss of generality no man would strictly prefer to pay the transfer required for him to marry any woman other than his wife in the equilibrium outcome. In the equilibrium, the required transfers for deviating to another woman will be functions of the characteristics of all the agents in the market. I study the empirically relevant case where the researcher lacks data on the equilibrium transfers. Therefore, estimation cannot be primitively based on the individual rationality condition that says each agent takes the action that maximizes its payoffs. The cooperative analog to the non-nested Nash solution concept that says each agent picks a strategy to maximize payoffs will not suffice for estimation. Rather, estimation will need to be based on necessary conditions that are implied by an equilibrium outcome being pairwise stable.

My main empirical interest is in large matching markets. In the automotive supplier application, there are 1349 car parts (matching opportunities) in one particular car component category. For simplicity, I treat each component category as a separate matching market. There are 1349 opportunities for a cars parts supplier to match with an assembler in a single matching market. Further, there are 593 different component categories in the data; I use all of them in estimation. In Fox and Bajari (2009), we apply a related version of the estimator in this paper to the matching between bidders and items for sale in a FCC spectrum auction. There are 85 winning bidders and 480 items for sale in the auction application. Both the automotive supplier and auction datasets are rich. There is a lot of information on agent characteristics and a lot of unknown parameters that can be learned from the observed sorting of suppliers to assemblers or bidders to items for sale. However, to take advantage of these types of rich data sets, a researcher must propose an estimator that works around a computational curse of dimensionality and a data curse of dimensionality.

First consider the computational curse of dimensionality. A standard parametric approach would be to write down a likelihood function that gives the probability that the observed set of matches is part of a pairwise stable outcome to a matching market, conditional on the observed characteristics of all agents in the particular matching market. If the econometrician modeled heterogeneity as arising from unobserved, match-specific aspects of payoffs, then the likelihood would be an integral over the match-specific unobservables. The integrand to the likelihood would be the nested calculation of the equilibrium set of matches for each draw of the errors. The computational cost of this approach can be tremendous.

For estimation, I provide a computationally simple, maximum score estimator for match production functions (Manski, 1975, 1985; Horowitz, 1992; Matzkin, 1993; Fox, 2007). The estimator uses inequalities derived from necessary conditions for pairwise stability.[1] These necessary conditions involve only observable characteristics; there is no potentially high-dimensional integral over unobservable characteristics. Evaluating the statistical objective function is computationally simple: checking whether an inequality is satisfied requires only evaluating match production functions and conducting pairwise comparisons. The objective function is the number of inequalities that are satisfied for any guess of the structural parameters. The estimate is any parameter than maximizes the number of included necessary conditions. Because the set of necessary conditions can be large, I argue that the estimator will be consistent if the researcher samples from the set of necessary conditions. Numerically computing the global maximum of the objective function requires a global optimization routine, although estimation is certainly doable with software built into commercial packages such as MATLAB or Mathematica.

The maximum score estimator is consistent because of an assumed rank order property that relates the necessary conditions to the probabilities of different equilibrium assignments, conditional on all the agent characteristics in a market. Following the most straightforward identification argument using this rank-order property would lead to another curse of dimensionality in the size of the matching market (Fox, 2009). This time a data curse of dimensionality would arise from having to nonparametrically estimate the probabilities of different assignments conditional on all the agent characteristics in a market. Both the number of distinct arguments and the number of conditioning arguments are of quite high dimension. Maximum score bypasses this need to estimate equilibrium assignment probabilities nonparametrically, even though the estimator is consistent because of the properties of conditional assignment probabilities.

A typical dataset for a matching market lists observed matches and the characteristics of agents in those matches. In many-to-many applications, such as automotive suppliers and assemblers, other objects are often not recorded in the data available to academic researchers. As stated before, the transfers between upstream and downstream firms may be private contractual relationships. Likewise, the profit or revenue to each party from its set of matches can be but is typically not observed. In many-to-many matching, the quota of each agent is the maximum number of physical matches that they can make. Binding quotas make agents on the same side of the market, say automotive suppliers, rivals to match with the agents on the other side of the market, say automotive assemblers. In the car-parts application, quotas are a modeling abstraction. I do not have data on quotas. Finally, even though matching models allow quota slots to be unfilled or people in a marriage market to be single, often data on unmatched agents are not available. The maximum score estimator in this paper does not require data on these variables that are part of the model's data generating process but are typically unobserved in empirical applications: transfers, revenues, profits, quotas and the characteristics of unmatched agents. I make no statistical assumptions about these variables; they are not given some parametric distribution and then integrated out. Rather, the maximum score estimator uses necessary conditions that do not require the values of these variables. Thus, the estimator is practically oriented towards realistic datasets for the automotive supplier and other similar applications.

In order to borrow insights from Manski (1975) and the related literature on maximum score

[1] There is a tradition of using necessary conditions or inequalities to estimate complex games. See Haile and Tamer (2003) and Bajari, Benkard and Levin (2007) for applications to noncooperative, Nash games.

methods, this paper relies on one non-primitive assumption: the rank order property. The rank order property is an assumption about the stochastic structure of the model and will be discussed below. The upside of this assumption is that allows me to tap into previous results and to propose a computationally simple estimator for the large and rich matching markets that motivate my empirical interest. The resulting estimator is also semiparametric: it does not rely on parameterizing the distribution of unobservables with a finite vector of parameters.

The rank order property is developed in more detail in a companion paper on nonparametric identification (Fox, 2009). By nonparametric (as opposed to semiparametric in this paper), I mean that the object of interest, the production function, is not specified up to a finite vector of parameters. Neither is the distribution of unobservables. Fox (2009) is the first to study nonparametric identification in any sort of matching game. In that paper, I ask what economic objects can be learned using a population dataset on equilibrium matches and agent characteristics. As agents on one side of the market are rivals to match with agents on the other side, the identification results are not simple extensions of those for single-agent models or those for noncooperative Nash games. The identification paper does not talk about developing a practical estimator for large datasets. Altogether, the rank order property allows a unified approach to be taken to computationally simple semiparametric estimation and nonparametric identification in matching games.

After earlier versions of this paper were circulated, Fox and Bajari (2009), Akkus and Hortacsu (2007), Baccara, Imrohoroglu, Wilson and Yariv (2009), Levine (2008), Mindruta (2009), and Yang, Shi and Goldfarb (2009) have conducted empirical work using the matching maximum score estimator I develop here. Their applications are, respectively, matching between bidders and items for sale in a spectrum auction, mergers between banks after deregulation in the United States, matching between offices and employees with attention paid to several dimensions of social networks, matching between pharmaceutical developers and distributors, matching between individual research team members in the patent development process, and matching between professional athletes and teams with a focus on marketing alliances between players and teams. In addition to the empirical application to automotive suppliers, these disparate applications show the relevance of matching estimation in empirical work in economics and allied fields such as corporate finance, marketing and strategy.

Recently, Dagsvik (2000), Choo and Siow (2006), and Weiss (2007) have introduced logit-based estimators for matching games with transfers. These authors study only one-to-one matching, namely marriage. Their estimators have not been extended to many-to-many matching, which is essential for most empirical applications outside of family economics, particularly applications in industrial organization and labor economics. Section 8 compares the maximum score estimator to the logit-based estimators in more detail. I briefly also compare my estimator to the parametric likelihood or method-of-moments estimators introduced for a non-nested class of matching games, those where agents cannot exchange monetary transfers (Gale and Shapley, 1962). Most empirical industrial organization and labor settings allow the matched agents, often firms or workers, to exchange monetary transfers in order to sell their services. Therefore, I believe matching games with transfers is the best model for many applications, even though the data on the monetary transfers are often unavailable.[2]

The paper is organized as follows. Section 2 provides a brief overview of some results from matching

[2]Matching with transfers is also related to models of hedonic equilibria, where typically features of the match in addition to price are endogeneously determined (Rosen, 1974; Ekeland, Heckman and Nesheim, 2004).

theory, discusses the role of matching estimation in empirical work and introduces and motivates the estimator. Section 3 outlines a many-to-many matching game. Section 4 discusses the aforementioned rank order property. Section 5 discusses estimation with the maximum score estimator and relevant asymptotic results. Section 6 presents a Monte Carlo study about the performance of the estimator. I focus on examples where the rank order property needed for consistency is not exactly satisfied, for reasons I go into below. Section 7 is the empirical application to automotive suppliers and assemblers. Having developed my approach to estimation in detail, Section 8 is the literature review.

2 Basic ideas for the case of marriage

Not all readers will be intimately familiar with matching theory. Using the simple example of one-to-one matching or marriage, this section introduces classic results from matching theory and then shows how identification, estimation and numerical analysis are needed to extend them. I then describe computational problems with traditional estimators and how maximum score solves those problems.

2.1 Why marriage is a particularly simple example

Let there be a set of M men. In a duplication of notation, let M be both the set of men and the number of men. Let the index a for a particular man also represent the vector of man a's observable characteristics. Let each man a be described by his schooling a_1 and his wealth a_2, so that the vector $a = (a_1, a_2)$. Likewise, there is a set W of W women. It is not necessarily the case that the number of men equals the number of women. Each woman i also two characteristics, schooling i_1 and wealth i_2. In a duplication of notation, $i = (i_1, i_2)$. Thus, the exogenous characteristics of agents in this matching market are $X = M \cup W$, a set of $M + W$ vectors of two scalars each.

If man a and woman i marry, the output of their match is given by the production function $f(a, i)$, which here is a function of observable agent characteristics: the schooling of the husband and wife and their respective wealth levels. Below, the focus on the production function will be motivated by a model where each man has preferences for female characteristics i, each woman has preferences for male characteristics a, and married couples can exchange transfers. The equilibrium concept of pairwise stability says no man would strictly prefer to pay the transfer needed to marry any woman other than his wife at an equilibrium outcome. For now, take the production function as the primitive object of interest.

An assignment is one part (the other part is the transfers) of an outcome to a matching market. An assignment A is a set of observed matches, say $\langle a, i\rangle$ for man a and woman i, $\langle b, j\rangle$ for man b and woman j, and so on. If man a is single, his match is recorded as $\langle a, 0\rangle$. An assignment is feasible if each man is married to at most one woman, each woman is married to at most one man, and men only marry women and women only marry men. A feasible assignment is an equilibrium assignment if the assignment is part of a pairwise stable outcome.

One-to-one, two-sided matching has three very convenient theoretical properties that do not generalize to many-to-many matching. First, a pairwise stable outcome is guaranteed to exist. Second, any pairwise stable outcome is in the core of the market: the assignment is efficient in that it maximizes the sum of production for all matches out of the set of all feasible assignments. The decentralized pairwise stable outcome solves a social planner's problem. Koopmans and Beckmann (1957) and Shapley and Shubik (1972) show the social planner's problem can be solved using linear programming techniques. Third, if agent characteristics such as a and i have distributions with continuous and product supports, then the probability that two different assignments solve the social planner's problem is zero. Thus, the third property ensures that the equilibrium assignment is unique with probability 1.

2.2 The role of matching estimation

Take a special case of the above model where $a = (a_1)$ and $i = (i_1)$, or each man and each woman has only a scalar characteristic, rather than a vector of two or more characteristics. Each man has a schooling level and each woman has a schooling level. In this special model, men with higher levels of schooling will marry women with higher levels of schooling when schooling levels are complements in the production of a match, or

$$\frac{\partial^2 f(a,i)}{\partial a_1 \partial i_1} > 0 \; \forall \, a_1 \in \mathbb{R}, i_1 \in \mathbb{R}. \tag{1}$$

Becker (1973) first proved the result that agents with scalar characteristics assortatively match when their characteristics are complements. Anti-assortative matching is the opposite: men with high levels of schooling marry women with low levels of schooling, and vice versa. Anti-assortative matching occurs when the two scalar characteristics are substitutes. Becker used these insights in empirical work: he analyzed whether married couples assortatively or anti-assortatively match on a variety of pairs of characteristics.

Unfortunately, Becker's sorting characterization that relates the sorting pattern to a high-level property of the match production function does not apply to a model where each agent has two or more characteristics, the case of the running example. Indeed, theorists have not analytically characterized the sorting patterns based on broad properties of production functions for cases other than the case of each agent having a single (scalar) characteristic. However, numerical analysis does apply to the case when agents have two or more characteristics. Given a particular choice of the production function $f(a,i)$ and values for X, the characteristics of all men and all women in a matching market, researchers can compute the socially optimal assignment, which (with probability 1) is the unique equilibrium assignment. Likewise, comparative statics in f can be undertaken by choosing different production functions and seeing how the assignments change. Alternatively, different values for the market characteristics X could be chosen, and the equilibrium computed for each choice of X.

The upshot is that numerical analysis is needed to analyze the predictions of matching models. A researcher must take a stand on the exact production function f and the exact characteristics in X in order to make predictions. Given the need for numerical analysis, the question becomes which f and X to pick? An obvious place to start is the X from the data and the f that is estimated from data on X and A in real-life matching markets. This is what this paper shows how to do: estimate f using i.i.d. observations on pairs (A, X) across many matching markets. One can look at how assignments (matches) of men to women vary with the set of exogenous characteristics of men and women in X and use that variation to estimate f.

As an assignment is a qualitative outcome to a market, it is helpful to keep in one's mind an analogy to discrete choice estimation. Consider the well-known multinomial logit model. The analog to an observation on a decision maker from the logit model is a matching market. The analog to a single agent's discrete choice from the logit model is A, the assignment in a market. The analog to the characteristics of the products from the logit model is X, the set of male and female characteristics in a market.

2.3 Two curses of dimensionality in the number of men and women in simulation estimation

Let me explain the issues with combinatorics in matching markets alluded to in the introduction. Let there be $M = 3$ men and $W = 3$ women in a marriage market, so $M = W$. None of the agents can be single, for expositional 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 $M = 100$ men and $W = 100$ 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.

For the purposes of estimation, let the production function $f_\beta(a, i)$ be specified up to a finite vector of parameters β. One could think of estimating the matching model using a standard parametric procedure. The method of simulated moments (MSM) is the most commonly used simulation estimator because the estimator is consistent as the number of observations, here the number of markets, goes to infinity, while holding the number of simulation draws constant. Recall the analogy to the logit model for single-agent choice. To estimate the logit using simulation, one would draw error terms and simulate the choices of each agent.

Let the set Γ collect the error terms. To be concrete, let Γ contain i.i.d. match-specific unobservables for all feasible matches, making the total production of a match be $f(a, i) + \epsilon_{\langle a,i\rangle}$ and $\Gamma = \left\{\epsilon_{\langle a,i\rangle}, \epsilon_{\langle a,k\rangle}, \ldots\right\}$, the set of unobserved payoffs for all $M \cdot W$ error terms. Let F be the distribution of $\epsilon_{\langle a,i\rangle}$. The natural MSM estimator for the matching model is based on the conditional moment equalities, for a pair (A, X)

$$m_{X,A}(\beta, F) = E_{\Gamma;F}\left[1\left[\alpha\left(X, \Gamma; f_\beta\right) = A\right] - \Pr\left(A \mid X\right) \mid X\right] = 0.$$

Here $\alpha(X, \Gamma; f_\beta)$ is a set-valued function that returns the equilibrium assignment to a matching market with observable agent characteristics X, simulation draws Γ, and production function $f_\beta(\cdot, \cdot)$. So the function $\alpha(X, \Gamma; f_\beta)$ is a nested computational procedure, in one-to-one matching a linear programming problem, that solves the social planning problem. Evaluating $\alpha(X, \Gamma; f_\beta)$ once can be time consuming if M and W are large. The moment is formed by computing the fraction of time the nested linear program computes A to be the assignment and comparing this probability to the empirical probability of observing assignment A when male and female characteristics are X, which is $\Pr(A \mid X)$. The unknown distribution F enters the moment through the expectation $E_{\Gamma;F}$.

Empirically, one needs to estimate the moment conditions. Say the researcher has first-stage estimates of assignment probabilities, as in $\hat{\Pr}(A \mid X)$. Let the empirically implemented moments in the MSM be

$$\hat{m}_{X,A}(\beta, F) = \frac{1}{S} \sum_{s=1}^{S} 1\left[\alpha\left(X, \Gamma_s; f_\beta\right) = A\right] - \hat{\Pr}(A \mid X).$$

Let us evaluate the computational cost of this procedure. Here the researcher uses S simulation draws in each moment. Each draw corresponds to a guess of Γ_s, the set of $M \cdot W$ match-specific unobservables. The size of Γ_s is 10,000 in the example with $M = W = 100$. To evaluate the moment condition, the researcher has to solve S linear programming problems. The computational cost of each linear programming problem will suffer from a computational cost in the number of men and women.

A further computational cost arises from the number of moments. There is one conditional moment for each pair (A, X). As already described, if $M = W = 100$, there are more values A than atoms in the universe, for a given X. The number of moments is thus unfathomable. One could try to reduce the computational cost by dropping some of the moments, but then the model will make less use of the data for pairs that are dropped. Simulated maximum likelihood would eliminate the need to choose moments, but would require more simulation draws for a low finite-sample bias.

There is also a data curse of dimensionality in forming this moment condition $\hat{m}_{X,A}(\beta, F)$. The assignment probabilities $\Pr(A \mid X)$ need to be estimated in a first stage. This step can only be done nonparametrically, as any functional form assumptions primitively specified on $\Pr(A \mid X)$ will likely be inconsistent with the model being estimated in the second stage, as $\Pr(A \mid X)$ is the outcome to an economic model. Any attempt to estimate $\Pr(A \mid X)$ nonparametrically using i.i.d. observations on (A, X) across markets will result in a data curse of dimensionality, as the arguments included in X often involve thousands of agent characteristics, and the number of possible values of A could exceed the number of atoms in the universe. If there are 100 men and 100 women in the market and each agent has a vector of four characteristics, this would require estimating the conditional probability of each of $100! = 9.33 \times 10^{157}$ distinct assignments, using conditioning arguments equal to the $(100 + 100) \cdot 4 = 800$ scalar agent characteristics. Nonparametrically estimating 9.33×10^{157} functions of 800 arguments each is not feasible.

2.4 Semiparametric estimation using maximum score

Let me now explain the simplicity of the maximum score estimator that I introduce in this paper. Consider observing a set of M towns, each an independent matching market. Let M also refer to the number of towns. Let the assignment A_m in town or market m be a finite set of observed matches of the form $\langle a, i \rangle$, where a is the characteristics of a particular a man and i is the characteristics of a particular woman. Again, let $f_\beta(a, i)$ be the production function, known up to a finite vector of parameters β. Assume the researcher uses data on only matched couples. Then the maximum score objective function is

$$H_M(\beta) = \frac{1}{M} \sum_{m \in M} \sum_{\{\langle a,i \rangle, \langle b,j \rangle\} \in A_m} 1 \left[f_\beta(a,i) + f_\beta(b,j) > f_\beta(a,j) + f_\beta(b,i) \right]. \tag{2}$$

The indicator functions $1[\cdot]$ are equal to 1 when the inequality in brackets is true and 0 otherwise. Each inequality says that the total sum of deterministic production of two matches will not be increased if the husbands exchange their wives. I will derive this inequality as an implication of pairwise stability soon. The score of correct predictions increases by 1 when a "local 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.

The objective function (2) is computationally simple. Evaluation of the objective function involves only evaluation of production functions, addition and checking of inequalities. Software is available on my website to numerically maximize (2) and to conduct statistical inference via subsampling, as I briefly discuss below (Santiago and Fox, 2007).[3]

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.

Some theoretical results on one-to-one, two-sided matching with transferable utility have been generalized 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 (2008) 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. The estimator in this paper can be extended to the cases studied by Kovalenkov and Wooders (2003) for one-sided matching, Ostrovsky (2008) 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.[4] 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.

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). Let $a \in D$ be the characteristics of an assembler or downstream firm, where D is the set of characteristics for all downstream firms. Let

[3] My website is http://home.uchicago.edu/~fox .

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

$i \in U$ be a supplier or upstream firm, with U the set of characteristics of all upstream firms. Let $X = D \cup U$ be the set of characteristics of all firms in a market. I will assume that X is observable in each market.

In many-to-many matching, each firm has a quota, a number of physical matches that it can have at once. Let $Q : U \cup D \to \mathbb{N}_+$ be the set of quotas, where $p_a \in Q$ is the quota of a downstream firm a and $q_i \in Q$ is the quota of the upstream firm i.

Let $\langle a, i \rangle$ be a match between downstream firm or automobile assembler a and upstream firm or car parts supplier i. If $p_a > 1$, a downstream firm, a, say, may be part of multiple matches. 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. The space of individual matches is $(U \cup \{0\}) \times (D \cup \{0\})$.

A matching-market outcome is a tuple (A, T). An assignment A, or a finite collection of matches for all agents in the market, is an element of the power set of $(U \cup \{0\}) \times (D \cup \{0\})$. For any assignment A with N matches, $A = \{\langle a_1, i_1 \rangle, \langle a_2, i_2 \rangle, \ldots, \langle a_N, i_N \rangle\}$, $T = \left\{t_{\langle a_1, i_1 \rangle}, t_{\langle a_2, i_2 \rangle}, \ldots, t_{\langle a_N, i_N \rangle}\right\}$ is a set of transfers for all matches in A. Each $t_{\langle a_1, i_1 \rangle} \in \mathbb{R}$ and represents a payment for a downstream firm to an upstream firm. In a supplier market with 100 consummated relationships, A is a finite set of 100 matches and T is a finite set of 100 transfers between each of the matched firms. Altogether, the combination of the exogenous and endogenous elements of a matching market form the tuple (D, U, Q, A, T).

For the purposes of semiparametric estimation, I will assume that the production function $f_\beta(\cdot, \cdot)$ is known up to a finite vector of parameters, β. The object of estimation will be β. A production function takes the characteristics of one upstream firm i and a set of n downstream firms $d \subseteq D$, such that $d = \{a_1, \ldots, a_n\}$ and $n \leq q_i$. In other words, $f_\beta(i, d)$ is the production of the set d of downstream firms in matches involving upstream firm i. Let β lie in some real space $\mathcal{B}$.

While the theoretical model allows many-to-many matching, the production function $f_\beta(i, d)$ involves nonlinearities across only multiple matches involving the same upstream firm, not multiple matches involving the same downstream firm $a \in d$. One could add a second production function $f_\beta^2(a, u)$ that takes as arguments the characteristics of a downstream firm or assembler a and a set u of n upstream firms $u = \{i_1, \ldots, i_n\}$. Then the production function $f_\beta^2(a, u)$ would capture nonlinearities across multiple matches involving the same assembler a. The ability to distinguish the role of $f_\beta(i, d)$ and $f_\beta^2(a, u)$ might involve functional form or exclusion restrictions. The nonparametric identification analysis in Fox (2009) requires $f_\beta^2(a, u) \equiv 0$; I keep the assumption here for compatibility with the identification results in the other paper.

3.2 Motivating production functions and pairwise stability

In empirical work, one typically primitively specifies the functional form $f_\beta(\cdot, \cdot)$. However, in economic theory the production function can arise as the sum of payoffs involving assemblers and suppliers. For the purposes of understanding where production functions and pairwise stability arise from, this section provides some more primitive background.

Given an outcome (A, T) in which supplier i is matched to the firms in the set d_i, the payoff of

$i \in U$ is

$$r_{\beta}^{1}(i, d_i) + \sum_{a \in d_i} t_{\langle a,i \rangle}. \tag{3}$$

Here, $r_{\beta}^{1}(\cdot, \cdot)$ is the structural revenue function for upstream firms. Likewise, $r_{\beta}^{2}(\cdot, \cdot)$ is the structural revenue function for downstream firms. The payoff at (A, T) for $a \in D$ for the match $\langle a, i \rangle \in A$ is $r_{\beta}^{2}(a, i) - t_{\langle a,i \rangle}$. 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. Given this notation, the production function $f_{\beta}(i, d) \equiv r_{\beta}^{1}(i, d) + \sum_{a \in d} r_{\beta}^{2}(a, i)$. Note that any transfers $t_{\langle a,i \rangle}$ would cancel if one sought to include transfers into the definition of a production function.

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 1. *Given an outcome (A, T) in which supplier i is matched to the assembler firms in the set d_i and assembler a is matched to the supplier firms in the set u_a, the 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 $\langle a, j \rangle \notin A$,*

$$r_{\beta}^{1}(i, d_i) + \sum_{c \in d_i \setminus \{a\}} t_{\langle c,i \rangle} + t_{\langle a,i \rangle} \geq r_{\beta}^{1}(i, (d_i \setminus \{a\}) \cup \{b\}) + \sum_{c \in d_i \setminus \{a\}} t_{\langle c,i \rangle} + \tilde{t}_{\langle b,i \rangle}, \tag{4}$$

where $\tilde{t}_{\langle b,i \rangle} \equiv r_{\beta}^{2}(b, i) - \left(r_{\beta}^{2}(b, j) - t_{\langle b,j \rangle}\right)$.

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

$$r_{\beta}^{1}(i, d_i) + \sum_{c \in d_i \setminus \{a\}} t_{\langle c,i \rangle} + t_{\langle a,i \rangle} \geq r_{\beta}^{1}(i, d_i \setminus \{a\}) + \sum_{c \in d_i \setminus \{a\}} t_{\langle c,i \rangle}.$$

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

$$r_{\beta}^{2}(a, i) - t_{\langle a,i \rangle} \geq 0.$$

4. *For all $\langle a, i \rangle \notin A$ where $|d_i| < q_i$ and $|u_a| < p_a$, there exists no $\tilde{t}_{\langle a,i \rangle} \in \mathbb{R}$ such that*

$$r_{\beta}^{1}(i, d_i) + \sum_{c \in d_i} t_{\langle c,i \rangle} < r_{\beta}^{1}(i, d_i \cup \{a\}) + \sum_{c \in d_i} t_{\langle c,i \rangle} + \tilde{t}_{\langle a,i \rangle}$$

and

$$r_{\beta}^{2}(a, i) - \tilde{t}_{\langle a,i \rangle} \geq 0.$$

Part 1 of the definition of pairwise stability says that upstream firm i prefers its current assembler a instead of some alternative assembler b at the transfer $\tilde{t}_{\langle b,i \rangle}$ that makes 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. Part 1 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 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.[5] The counterexamples mean that general existence theorems do not exist.[6]

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.[7]

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 the inequality that states that upstream firm j does not want to exchange its matched assembler b for a new assembler partner a:

$$r^1_\beta\left(j, d_j\right) + \sum_{c \in d_j \backslash \{b\}} t_{\langle c,j\rangle} + t_{\langle b,j\rangle} \geq r^1_\beta\left(j, \left(d_j \backslash \{b\}\right) \cup \{a\}\right) + \sum_{c \in d_i \backslash \{b\}} t_{\langle c,j\rangle} + \tilde{t}_{\langle a,j\rangle}, \tag{5}$$

where $\tilde{t}_{\langle a,j\rangle} \equiv r^2_\beta\left(a, j\right) - \left(r^2_\beta\left(a, i\right) - t_{\langle a,i\rangle}\right)$. Substituting in $\tilde{t}_{\langle b,i\rangle}$ and $\tilde{t}_{\langle a,j\rangle}$, adding (4) and (5),

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

[6] 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.

[7] In the non-nested-with-matching literature on estimating normal-form Nash games, Ciliberto and Tamer (2009) throw out a particular realization of the error term's contribution to the likelihood if no pure-strategy equilibrium exists. Bajari, Hong and Ryan (2009) 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 uses monotonic operators and hence requires structure on preferences to ensure this monotonicity.

canceling the transfers $t_{\langle a,i\rangle}$, $t_{\langle b,j\rangle}$, $\sum_{c\in d_j\setminus\{b\}} t_{\langle c,j\rangle}$ and $\sum_{c\in d_i\setminus\{a\}} t_{\langle c,i\rangle}$ that now are the same on both sides of the inequality, and substituting in the definition of production function gives

$$f_\beta\left(i,d_i\right)+f_\beta\left(j,d_j\right)\geq f_\beta\left(i,\left(d_i\setminus\{a\}\right)\cup\{b\}\right)+f_\beta\left(j,\left(d_j\setminus\{b\}\right)\cup\{a\}\right). \tag{6}$$

Given two sets of downstream firms d_i and d_j that may or may not be related to an assignment A, I call a strict ($>$ instead of $\geq$) version of this inequality 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 is the key inequality that will form the basis for estimation.

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 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 $f_\beta\left(i,a\right)=\beta_1'i+\beta_2'a$, then a local production maximization inequality is

$$\beta_1'i+\beta_2'a+\beta_1'j+\beta_2'b>\beta_1'i+\beta_2'b+\beta_1'j+\beta_2'a, \tag{7}$$

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.[8]

Fox (2009) shows that exchanges of downstream firm partner each between three or more upstream firms can provide additional sets of valid local production maximization inequalities. Fox uses those extra inequalities in some nonparametric identification theorems. As the extra inequalities will often not be needed for some common functional form choices for $f_\beta\left(i,d\right)$, for conciseness I will not discuss the extra inequalities here. Note that inequalities based on exchanges of two or more downstream firms per upstream firm are not motivated by the definition of pairwise stability. Pairwise stability implies local production maximization inequalities based on exchanges of only one downstream firm per upstream firm. This paper does not consider adding inequalities from stronger solution concepts, such as the core. Fox explores nonparametric identification using only inequalities from pairwise stability.

4 The rank order property

The previous section does not discuss econometric unobservables. Let us ponder a model where the production of each set d of downstream firms matched to a supplier i is $f_\beta\left(i,d\right)+\sum_{a\in d}\epsilon_{\langle a,i\rangle}$, with each $\epsilon_{\langle a,i\rangle}$ being having some common distribution F, which lies in some space of distributions $\mathcal{F}$. The assignment probability function $\Pr\left(A\mid X;\,\beta,F\right)$ is a property of the matching game and

[8] 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.

the distribution of errors F. The assignment probability function integrates out the unobservables, conditioning on the observables, and gives the probability of observing assignment A as part of the pairwise stable outcome (A, T) to a market with observables X. The assignment probability function also involves equilibrium assignment selection rules for matching games with multiple equilibrium assignments as well as the distribution of the unobserved quotas Q conditional on the observed firm characteristics in X, if quotas are unobserved. Fox (2009) contains many more details on these complications and how they affect the rank order property.

The discussion in Section 2.3 showed that a straightforward simulation estimator will be computationally infeasible in many empirical applications. This paper proposes a maximum score estimator that is computationally feasible. The assumption that is needed for the consistency of maximum score and for the nonparametric identification theorems in Fox (2009) is called the rank order property. The rank order property is

Assumption 1. *Let A_1 be a feasible assignment for a market with characteristics X. Let*

$$A_2 = (A_1 \backslash \{\langle a, i \rangle, \langle b, j \rangle\}) \cup \{\langle a, j \rangle, \langle b, i \rangle\}$$

be another feasible assignment, where $\{\langle a, i \rangle, \langle b, j \rangle\} \subseteq A_1$. Let d_i be the assemblers matched to supplier i at the assignment A_1. Let $F \in \mathcal{F}$ be any distribution of the error terms and let $\beta \in \mathcal{B}$ be any valid value of the parameters in the production function.

Assume that

$$f_\beta(i, d_i) + f_\beta(j, d_j) > f_\beta(i, (d_i \backslash \{a\}) \cup \{b\}) + f_\beta(j, (d_j \backslash \{b\}) \cup \{a\}) \tag{8}$$

if and only if

$$\Pr(A_1 \mid X; \beta, F) > \Pr(A_2 \mid X; \beta, F).$$

The rank order property states that if a local production maximization equality is satisfied *when the error terms are ignored*, the probability of observing the market-wide assignment A_1 where the deterministic matching game ($\epsilon_{\langle a,i \rangle} \equiv 0$) may satisfy pairwise stability is greater than the probability of observing the market-wide assignment A_2 where the deterministic matching game is known not to satisfy pairwise stability. Given X, 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 firms in A_2, say a and j, would prefer to match with each other instead of their assigned partners, leading to A_1. 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.

Again, Fox (2009) has an in-depth discussion of the rank order property and its validity. The rank order property does not hold, exactly, if the output to a match is $f_\beta(i, d) + \sum_{a \in d} \epsilon_{\langle a,i \rangle}$ and each $\epsilon_{\langle a,i \rangle}$ is i.i.d. However, Fox includes simulation evidence that the rank order property is often not seriously violated when the output to a match is $f_\beta(i, d) + \sum_{a \in d} \epsilon_{\langle a,i \rangle}$ and each $\epsilon_{\langle a,i \rangle}$ is i.i.d. Further, if the unobservables occur at the assignment A level (each ϵ_A is i.i.d. or exchangeable), then Fox shows that the rank order property holds, exactly.

This paper uses the rank order property because its leads to a computationally tractable estimator while simulation estimators, which are explicit about how the errors enter production, do not. Further, the rank order property allows the discussion of nonparametric identification in Fox (2009). Perhaps the most practical way of judging the usefulness of the estimator when the true output of a match is indeed $f_\beta(i,d) + \sum_{a \in d} \epsilon_{\langle a,i \rangle}$ is to perform a Monte Carlo study of the estimator's finite sample performance under this misspecification (the rank order property is not satisfied). Section 6 presents such a Monte Carlo experiment.

5 The maximum score estimator

I now discuss how maximum score can form the basis for a practical estimator. The maximum score estimator avoids a computational curse of dimensionality by eliminating all nested calculations. Further, all inequalities do not need to be included with probability 1 to maintain the consistency of the estimator. It avoids a data curse of dimensionality by avoiding the need to estimate the very high-dimensional $\Pr(A \mid X)$ nonparametrically. Maximum score estimation was introduced by Manski (1975, 1985) for the single-agent model.

I assume the researcher has access to i.i.d. observations on distinct matching markets (A_m, X_m), for $m = 1, \ldots, M$. The number of observations is M, the number of markets. Each observation on a large matching market will contain much more information than an observation on, say, an agent making a binary choice. Still, the asymptotics will be in the number of markets. Fox and Bajari (2009) consider the case of asymptotics in the number of agents in a single matching market.

5.1 The matching maximum score estimator

There are a variety of inequalities that could be included for each market. 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 $\{\langle a,i \rangle, \langle b,j \rangle\} \subseteq A_m$ on the left side. The maximum score estimator is any parameter vector $\hat{\beta}_M$ that maximizes

$$H_M(\beta) = \frac{1}{M} \sum_{m \in M} \sum_{\{\langle a,i \rangle, \langle b,j \rangle\} \in I_m} 1\left[f_\beta(i,d_i) + f_\beta(j,d_j) > f_\beta(i,(d_i \setminus \{a\}) \cup \{b\}) + f_\beta(j,(d_j \setminus \{b\}) \cup \{a\})\right]. \tag{9}$$

Evaluating $H_M(\beta)$ 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 $\epsilon_{\langle a,i \rangle}$ in (9), even though the estimator may perform well if the data are generated from a model with such errors. Not all inequalities will be satisfied, even at the maximizer $\hat{\beta}_M$ and even at the probability limit of the objective function.[9]

Manski and Thompson (1986) and Pinkse (1993) present optimization algorithms for the maximum score objective function where the parameters enter linearly into the payoff function. In the

[9]This distinguishes maximum score from a moment-inequality approach (Pakes, Porter, Ho and Ishii, 2006).

empirical application, I numerically maximize the maximum score objective function using the global optimization routine known as differential optimization (Storn and Price, 1997).

5.2 Choosing inequalities

The set of inequalities I_m included in estimation for market m does not need to include all theoretically valid inequalities. If all inequalities were included, the estimator would suffer from a computational curse of dimensionality in the number of firms in a matching market, as the number of valid inequalities grows rapidly with the size of the market. Luckily, inequalities only need to be included with some positive probability for the estimator to be consistent.[10] 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 $\{\langle a, i\rangle, \langle b, j\rangle\}$, given assignment A and observable firm characteristics X. Let $I(\{\langle a, i\rangle, \langle b, j\rangle\} \mid X)$ be the probability, conditional on X, that a researcher includes an inequality when $\{\langle a, i\rangle, \langle b, j\rangle\} \in N(A, X)$. Hence, $I(\{\langle a, j\rangle, \langle b, i\rangle\} \mid X)$ is the probability of sampling $\{\langle a, j\rangle, \langle b, i\rangle\}$ when $\{\langle a, j\rangle, \langle b, i\rangle\} \in N(A_2, X)$, for some other assignment A_2.

Assumption 2. *For all $\{\langle a, i\rangle, \langle b, j\rangle\} \in N(A, X)$ and for any feasible pair (A, X),*

1. $I(\{\langle a, i\rangle, \langle b, j\rangle\} \mid X) = I(\{\langle a, j\rangle, \langle b, i\rangle\} \mid X)$.
2. $I(\{\langle a, i\rangle, \langle b, j\rangle\} \mid X) > 0$.

The assumption means that the probability of including an inequality when it is valid must be equal to the probability of including the reverse inequality when it is valid. 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. Note that issues such as including inequalities with only positive probability do not arise in the previous literature on maximum score, which mainly considered computationally tractable single-agent choice problems.

Often a researcher will not have a good idea of the boundaries in space and time of a matching market. By defining a market conservatively, so that the market definition used in estimation is weakly smaller than the true market, consistency will be maintained if the discarded inequalities are not necessary for point identification. By contrast, other simulation estimators will be inconsistent if the market is defined incorrectly.

5.3 Consistency

Some other sufficient conditions for consistency follow.

Assumption 3.

1. *The production function parameters β lie in a compact set $\mathcal{B} \subseteq \mathbb{R}^{|\beta|}$, $|\beta| < \infty$.*
2. *Identification: Let $\beta^0 \in \mathcal{B}$ and $F^0 \in \mathcal{F}$ be the true primitives that generate the data. For any $\beta^1 \neq \beta^0$, $\beta_1 \in \mathcal{B}$, and for any $F^1 \in \mathcal{F}$, there exists a set of market characteristics $\mathcal{X}$ with positive*

[10] This estimator will not have a normal distribution. Therefore, I will avoid discussing how the choice of inequalities relates to statistical efficiency.

probability and two assignments A_1 and A_2 such that $\Pr\left(A_1 \mid X;\, \beta^0, F^0\right) > \Pr\left(A_2 \mid X;\, \beta^0, F^0\right)$ *while* $\Pr\left(A_1 \mid X;\, \beta^1, F^1\right) < \Pr\left(A_2 \mid X;\, \beta^1, F^1\right)$ *for any* $X \in \mathcal{X}$.

3. *Each vector of supplier characteristics i or assembler characteristics a in X has one or more elements with continuous support.*

4. *X is independently and identically distributed across markets.*

Assumption 3 assumes identification rather than proving it for a given functional form choice. The paper by Fox (2009) proves the nonparametric identification of various features of match production functions; the identification theorems can be used to determine what parametric functional forms can be identified using equilibrium data on who matches with whom. Also, Fox (2007) provides an easy-to-follow consistency proof of a single-agent maximum score estimator when the single agent makes a multinomial choice and the utility of each choice j is the linear index $x_j'\beta$.

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

Theorem 1. *As $M \to \infty$, any $\hat{\beta}_M \in \mathcal{B}$ that maximizes the matching maximum score objective function is a consistent estimator of $\beta^0 \in \mathcal{B}$, the parameter vector in the data generating process.*

There is a simple proof in the appendix. 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.[11] 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 $\Pr(A \mid X;\, \beta, F)$ does not have to be manually computed for each guess of β and F 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,

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

where $1\left[A_m = A\right]$ equals 1 if assignment A occurs in market m.[12]

5.4 Inference, estimators with faster rates of convergence and set inference

Kim and Pollard (1990) show that the binary 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

[11] Like the work on single-agent choice by Manski, the matching maximum score estimator does not allow the distribution of unobservables, F, to be estimated. Indeed, F is not identified under the weakest assumptions needed for the identification of β, even in single-agent choice.

[12] The proof shows that the true parameter vector β^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 β. 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.

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 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.

Jun, Pinkse and Wan (2009) present a new estimator for models such as maximum score. The estimator is a Chernozhukov and Hong (2003) Laplace type estimator (LTE), although the nonstandard asymptotics give the estimator somewhat different properties. Like smoothed maximum score, the LTE can converge at a rate close to $\sqrt{M}$; inference does not require a resampling procedure such as subsampling.

In private conversations, Manski suggests using set inference procedures for maximum score, even if the model is perhaps point identified. Point identification in maximum score is not equivalent to to identification at infinity (Andrews and Schafgans, 1998). Rather, point identification involves finding firm characteristics such that

$$f_{\beta^0}(i,d_i) + f_{\beta^0}(j,d_j) - f_{\beta^0}(i,(d_i \backslash \{a\}) \cup \{b\}) - f_{\beta^0}(j,(d_j \backslash \{b\}) \cup \{a\}) \approx \\ f_{\beta^1}(i,d_i) + f_{\beta^1}(j,d_j) - f_{\beta^1}(i,(d_i \backslash \{a\}) \cup \{b\}) - f_{\beta^1}(j,(d_j \backslash \{b\}) \cup \{a\})$$

for the true parameter vector β^0 and some alternative $\beta^1 \neq \beta^0$. As β^0 is not known to the researcher, typically a full support condition ensures that any needed values of (i, d_i, j, d_j) will be in the support of the data. A failure of this assumption results in set rather than point identification. Set identification is robust to the failure of support conditions for point identification. In a sense, set inference makes more use of the data. Bajari, Fox and Ryan (2008) explore set inference in maximum score, motivated by an industrial organization demand application. The matching estimation software available on my website conducts subsampling inference both for point- and set-identified maximum score (Santiago and Fox, 2007).

6 Monte Carlo experiments

This section presents evidence that the maximum score estimator works well in finite samples and with i.i.d. match-specific errors. The Monte Carlo study examines games of one-to-one, two-sided matching. Section 2 provides background on this class of games and Fox (2009) argues that the rank order property holds only approximately under i.i.d. match-specific errors. Fox presents an alternative sufficient condition involving a social planner's errors, but in the Monte Carlo study I restrict attention to the i.i.d. match-specific errors case. This section reports a Monte Carlo study for an estimator that is not formally consistent: the rank-order property does not hold.

Each agent is distinguished by two characteristics, for upstream firm i, i_1 and i_2, and for downstream firm a, a_1 and a_2. The total output from a match of i to a is

$$f_{\beta_1,\beta_2}(a,i) + \epsilon_{\langle a,i\rangle} = \beta_1 a_1 i_1 + \beta_2 a_2 i_2 + \epsilon_{\langle a,i\rangle}.$$

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,

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 10 \\ 10 \end{bmatrix}, \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}\right).$$

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 $\epsilon_{\langle a,i\rangle}$'s are i.i.d. with a normal distribution with a standard deviation of either 1 or 5. In the second set of experiments, the match-specific $\epsilon_{\langle a,i\rangle}$'s are i.i.d. 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 a linear programming problem described in Roth and Sotomayor (1990). The linear programming formulation ensures that all consummated matches provide non-negative surplus. Few of the agents are unmatched in the fake data, as the means of both characteristics are high.

While not shown, I have generated 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.[13]

Table 1 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

[13]For each replication, the Monte Carlo study reports the maximum provided by the optimization routine, which is a consistent estimator under the conditions in the Monte Carlo experiment. 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.

firms on each side increases to 200. Fox and Bajari (2009) investigate the asymptotic properties of the estimator using data on only one large matching market.

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 errors in the data generating process. In these experiments, the estimator is not very biased when there are i.i.d. match-specific errors. This supports the use of the maximum score estimator even when it may be formally misspecified, as when there are i.i.d. match-specific errors.[14]

7 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.[15]

The data group car parts into component categories, and I treat each component category as a statistically independent matching market.[16] 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

[14] Note that the misspecification is analogous to estimating a logit when the true model is probit much more than not correcting for selection bias or omitted variable bias.

[15] 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.

[16] The same firm may appear in multiple component categories, and so a researcher might want to model spillovers and hence statistical dependence in the outcomes across component categories. Pooling all component categories into one matching market would require asymptotics in the number of agents in a single matching market, which is discussed in Fox and Bajari (2009) but not here.

in some other papers on estimating matching games, which are discussed in Section 8. 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 Fox and Bajari (2009), 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.

7.1 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 and Korea), Europe, and North America.[17] The management literature has suggested that supplier specialization may be a key driver of assembler performance (Dyer, 1996, 1997).[18] Here I focus on how specialization can affect the production from a set of car-part relationships centered around a single supplier.[19]

In a slight extension of the notation from the earlier part of the paper, let d 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

$$f_\beta (i, d) = \beta_{\text{Cont.}} x^{\text{Continent}} (d) + \beta_{\text{PG}} x^{\text{ParentGroup}} (d) + \beta_{\text{Brand}} x^{\text{Brand}} (d) + \beta_{\text{Car}} x^{\text{Car}} (d) . \tag{10}$$

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 $x^{\text{ParentGroup}} (d)$ 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

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

[17] 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.

[18] 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.

[19] 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 are recorded and used as endogenous matching outcomes. Vertical integration in automobile manufacturing has been studied previously (Monteverde and Teece, 1982; Novak and Eppinger, 2001; Novak and Stern, 2008, 2009).

More generally,

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

where $|d|$ is the number of parts in d, D^{PG} is the set of parent groups, L_a^{PG} is the set of car parts for parent group a, and $\left| L_a^{\text{PG}} \cap d \right|$ 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.

It should be clear that the four match-specific characteristics in (10) 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 $\beta_{\text{PG}} x^{\text{ParentGroup}}(d)$ 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}.

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.[20] 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.[21]

I use the maximum score estimator, (9), to compute point estimates, and subsampling to pro-

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

[21] 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.

duce confidence intervals. I use local-production-maximization inequalities with the left and right side matches being of the form $\{\langle a,i,l\rangle, \langle b,j,h\rangle\}$ and $\{\langle b,i,l\rangle, \langle a,j,h\rangle\}$. I include two suppliers per inequality, and they exchange one car part each.[22] 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 2.

Table 2 presents point estimates and subsampled confidence intervals for the four HHI specialization measures.[23] We see that all four estimates are positive, meaning as expected specialization on these dimensions 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 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. 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 2 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 special-

[22] 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.

[23] 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 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 2 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.

ization 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 relationships.[24]

7.2 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.[25] 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 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?[26]

[24] 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, $2.0 \times 10^{-15} > 1.0 \times 10^{-15}$.

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

[26] In the non-causal interpretation, one should not use the production function to explore counterfactuals where $x^{\text{SupplierToAsian}}(d)$ changes because the equilibrium set of downstream partners in d changes. In this interpretation, $x^{\text{SupplierToAsian}}(d)$ 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

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 (10) to be

$$f_\beta(i,d) = \beta_{\text{Cont.}} x^{\text{Continent}}(d) + \beta_{\text{PG}} x^{\text{ParentGroup}}(d) + \beta_{\text{Brand}} x^{\text{Brand}}(d) + \beta_{\text{Car}} x^{\text{Car}}(d) + \beta_{\text{AsianCont.}} x^{\text{Continent}}(d)\, x^{\text{SupplierToAsian}}(d_i^m), \quad (11)$$

where d_i^m is the set of downstream firms matched to supplier i in market m. The new term

$$x^{\text{Continent}}(d)\, x^{\text{SupplierToAsian}}(d_i^m)$$

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 $\left(\beta_{\text{Cont.}} + \beta_{\text{AsianCont.}} x^{\text{SupplierToAsian}}(d_i^m)\right) \cdot x^{\text{Continent}}(d)$. 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 $x^{\text{SupplierToAsian}}(d_i^m)$ that do not impose any mechanical relationship between $x^{\text{SupplierToAsian}}(d_i^m)$ and the previous HHI-specialization measures. In other words, $x^{\text{SupplierToAsian}}(d_i^m)$ 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 d_i^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

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

where again d_i^m is a set of car parts for supplier i in a component-category market m with equilibrium assignment A_m. $x^{\text{SupplierToAsian,2}}(d_i^m)$ 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 $x^{\text{SupplierToAsian}}(d_i^m)$ 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. Section 7.3 explores the alternative specification, where $x^{\text{SupplierToAsian}}(d)$ is recomputed with counterfactual matches d.

Table 3 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

factual patterns in the data, just like the non-causal interpretation of production functions summarizes facts about sorting patterns.

Table 2, the lower point estimates for the non-normalized specialization parameters coincide with the normalized parameter, here $\beta_{\text{Cont.}}$, being relatively more important.[27] 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 3 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 $x^{\text{SupplierToAsian},2}\left(d_i^m\right)$ 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 similar to the specification with the dummy, except for the fact that the point estimates on the other three HHI specialization 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 3 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.

7.3 Recomputing the Asian indicator in inequalities

Previously, 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.

In the previous production function specification, $x^{\text{SupplierToAsian}}\left(d_i^m\right)$ is a function of only the actual assignment, A_m and hence d_i^m, and is not recomputed when d changes on the right side of the

[27] With the interaction term included in (11), 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, $x^{\text{SupplierToAsian}}\left(d_i^m\right) = 0$. Suppliers with no Asian contracts have a 0 value for the interaction term, $x^{\text{Continent}}(d)\, x^{\text{SupplierToAsian}}\left(d_i^m\right)$. 48% of supplier / matching market combinations do not supply any assembler brand with its headquarters in Asia.

inequality. Table 4 reports estimates of the matching model using the production function

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

where now $x^{\text{SupplierToAsian}}(d)$ is recomputed each time $d \subseteq D$ 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 Table 3, 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 4 are economically small and have confidence regions that lead to the rejection of null hypotheses of large in absolute value, negative coefficients for $\beta_{\text{AsianCont.}}$.[28]

I spent some time investigating why the point estimates for $\beta_{\text{AsianCont.}}$ varied across so much across Tables 3 and 4. 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

$$f_\beta(i, d_i^m) + f_\beta\left(j, d_j^m\right) > f_\beta\left(i, (d_i^m \setminus \{a\}) \cup \{b\}\right) + f_\beta\left(j, \left(d_j^m \setminus \{b\}\right) \cup \{a\}\right),$$

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 $x^{\text{SupplierToAsian}}(d)$ is either 0 or 1 for each of the four matches, so the values of $x^{\text{SupplierToAsian}}(d)$ for an inequality can be written as, for example, $\{1,1\} \to \{1,0\}$.[29] 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 $x^{\text{SupplierToAsian}}(d)$ 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.[30] To confirm this, I created an artificial set of inequalities equal to the inequalities used in Table 4, 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 3, where the Asian indicator is not recomputed.[31] The estimates are in Table 5. We can see

[28] 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.

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

[30] I changed each of several types of inequalities from their Table 4 forms back to their Table 3 forms, and evaluated the objective function at the main text Table 3 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.

[31] My goal was to find the minimum set of inequalities that could change and restore the point estimates of Table 3. There were 10,603 inequalities of the form $\{1,1\} \to \{1,0\}$ in the dataset behind Table 4. I modified only 10,358

that the point estimates for the HHI specialization measures are in between those in Table 3 and Table 4. 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 3.

The exercise in Table 5 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 3 to -0.000641 in Table 4. I will now take a speculative stab at offering an economic story to explain the point estimates. When $x^{\text{SupplierToAsian}}(d_i^m)$ is not recomputed for counterfactual matches, the inequalities answer the questions discussed in Section 7.2: do suppliers to Asian brands have some competitive edge with non-Asian assemblers? When $x^{\text{SupplierToAsian}}(d)$ is recomputed for counterfactual d'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 the Table 3. In Table 4, the model deals with these two opposing forces by setting the coefficient on $\beta_{\text{AsianCont.}}$ to be near zero. The inequalities in Table 3 are easier to understand and interpret because the estimate for the parameter $\beta_{\text{AsianCont.}}$ reflects a fixed firm-specific characteristic $x^{\text{SupplierToAsian}}(d_i^m)$ that represents only one economic phenomenon, the competitive edge of suppliers to Asian brands.[32]

8 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.

8.1 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.[33] These estimators are computationally simple because they exploit the mathematics behind the aggregate-data, multinomial-choice logit model (McFadden, 1973; Berry, 1994). However, these estimators have not been expanded to the case

inequalities. When evaluated at the point estimates from Table 3 (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 3.

[32] More mechanically, when the interaction term is $x^{\text{Continent}}(d)\,x^{\text{SupplierToAsian}}(d_i^m)$, 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, $x^{\text{Continent}}(d)$, from one car part being exchanged. For a company that supplies several car parts in this matching market, $x^{\text{Continent}}(d)$ and hence the interaction change by a relatively small amount. When $x^{\text{SupplierToAsian}}(d)$ is recomputed for counterfactual downstream firm partners d, then firms that supply only one or two car parts to Asian assemblers have a relatively large change in $x^{\text{SupplierToAsian}}(d)$ and hence in the interaction $x^{\text{Continent}}(d)\,x^{\text{SupplierToAsian}}(d)$. It is unsurprising that the Table 3 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 4 specification with relatively large changes in the interaction.

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

of many-to-many matching.

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 the same is in the current paper: for each value of unobservables, a physically feasible assignment is not formed. 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 i.i.d. 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 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 that 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 this paper has the error terms enter a social planner's (linear programming) problem that ensures, for every realization of the errors for all agents, that the resulting assignment is physically feasible.[34]

8.2 Estimators for games without transfers

Recently, Boyd, Lankford, Loeb and Wyckoff (2003), Sørensen (2007), and Gordon and Knight (2009) estimate Gale and Shapley (1962) matching games, which do not use transfers as part of the equilibrium concept.[35] 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. 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.[36]

[34] 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 this paper a full matching game is solved for each realization of the error terms. In this model, the distinction between error terms $\epsilon_{\langle a,i\rangle}$ and the characteristics in X is only whether the exogenous variable in question is recorded in the data or not. Equilibrium monetary transfers given by the model, even if not recorded in the data, will be a function of the error terms $\epsilon_{\langle a,i\rangle}$ for all potential matches $\langle a,i\rangle$.

[35] Hitsch, Hortaçsu and Ariely (2009) 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 (2008) examines testable restrictions on the lattice of equilibrium assignments of the Gale and Shapley (1962) model.

[36] A similar assumption for matching games with transfers would be that production functions are additively separable across multiple matches: $f_\beta(i,\{a,b\}) = f_\beta(i,\{a\}) + f_\beta(i,\{b\})$. This would rule out the study of spectrum auctions in

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 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 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, Fox (2009) argues that the maximum score necessary-conditions approach can be valid in the presence of multiple equilibrium assignments, under some moderately strong assumptions about the equilibrium selection rule.[37]

9 Conclusions

This paper discusses the 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.

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 re-

Fox and Bajari (2009) and the automotive-supplier specialization empirical example in this paper.

[37]Further, games 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.

quire numerical integrals over match-specific error terms. Third, inequalities need to be included only with some positive probability, which is important given that the number of necessary conditions from pairwise stability increases rapidly with the number of agents in a matching market. 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.

A Proofs

A.1 Theorem 1: Consistency

A.1.1 Constructive identification

By a law of large numbers and the law of iterated expectations, the probability limit of the maximum score objective function is

$$H_\infty(\beta) = E_X \left\{ \sum_{A \in \mathcal{A}(X)} \sum_{\{\langle a,i\rangle, \langle b,j\rangle\} \in N(A,X)} \Pr(A \mid X) \cdot I(\{\langle a,i\rangle, \langle b,j\rangle\} \mid X) \cdot \right.$$
$$\left. 1\left[f_\beta(i,d_i) + f_\beta(j,d_j) > f_\beta(i,(d_i \setminus \{a\}) \cup \{b\}) + f_\beta(j,(d_j \setminus \{b\}) \cup \{a\})\right] \right\},$$

where $\mathcal{A}(X)$ is the set of feasible assignments given X and $\Pr(A \mid X) = \Pr\left(A \mid X;\, \beta^0, F^0\right)$, where β^0 is the true parameter vector and F^0 is the true distribution of the unobservable terms. The sets of downstream matches d_i and d_j are functions of the assignment A, although the notation is suppressed.

For each pair of an assignment $A_1 \in \mathcal{A}(X)$ and a $\{\langle a,i\rangle, \langle b,j\rangle\} \in N(A_1,X)$ in the integrand above, there is an assignment $A_2 \in \mathcal{A}(X)$ that is $A_2 = (A_1 \setminus \{\langle a,i\rangle, \langle b,j\rangle\}) \cup \{\langle a,j\rangle, \langle b,i\rangle\}$. An inequality for A_1 and $\{\langle a,i\rangle, \langle b,j\rangle\}$ is mutually exclusive with a paired inequality for A_2 and $\{\langle a,j\rangle, \langle b,i\rangle\}$. 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 2, $I(\{\langle a,i\rangle, \langle b,j\rangle\} \mid X) = I(\{\langle a,j\rangle, \langle b,i\rangle\} \mid X)$. The ranking of the weights on the indicators reduces to comparing $\Pr\left(A_1 \mid X;\, \beta^0, F^0\right)$ and $\Pr\left(A_2 \mid X;\, \beta^0, F^0\right)$. 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 $H_\infty(\beta)$.

Let $\beta^1 \in \mathcal{B}$ be some parameter vector where $\beta^1 \neq \beta^0$. By Assumption 3, there exists a set $\mathcal{X}$ of X with positive probability and two assignments A_1 and A_2 such that $\Pr\left(A_1 \mid X;\, \beta^0, F^0\right) > \Pr\left(A_2 \mid X;\, \beta^0, F^0\right)$ while $\Pr\left(A_1 \mid X;\, \beta^1, F^1\right) < \Pr\left(A_2 \mid X;\, \beta^1, F^1\right)$ for any $X \in \mathcal{X}$, for any $F^1 \in \mathcal{S}$. Considering all the $X \in \mathcal{X}$, $H_\infty\left(\beta^1\right) < H_\infty\left(\beta^0\right)$ because β^1 causes inequalities with the lower of $\Pr\left(A_1 \mid X;\, \beta^0, F^0\right)$ and $\Pr\left(A_2 \mid X;\, \beta^0, F^0\right)$ to enter the objective function.

A.1.2 Continuity of the limiting objective function and uniform convergence

Lemma 2.4 from Newey and McFadden (1994) can be used to prove continuity of $H_\infty(\beta)$ as well as uniform-in-probability convergence of $H_M(\beta)$ to $H_\infty(\beta)$. Remember that the asymptotics are in the number of markets. The conditions of Lemma 2.4 are that the data (across markets) are i.i.d., which can hold even if we view the number of upstream and downstream firms as random; that the parameter space $\mathcal{B}$ is compact (Assumption 3), 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 3, as each firm's characteristic vector 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.

References

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

Akkus, Oktay and Ali Hortacsu, "The Determinants of Bank Mergers: A Revealed Preference Analysis," March 2007. University of Chicago working paper.

Andrews, Donald W.K. and Marcia M.A. Schafgans, "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies*, 1998, *65* (3), 497–517.

Baccara, Mariagiovanna, Ayse Imrohoroglu, Alistair Wilson, and Leeat Yariv, "A Field Study on Matching with Network Externalities," August 2009. New York University working paper.

Bajari, Patrick, 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," *Econometrica*, 2009. forthcoming.

_ , Jeremy T. Fox, and Stephen Ryan, "Evaluating Wireless Carrier Consolidation Using Semiparametric Demand Estimation," *Quantitative Marketing and Economics*, 2008, *6* (4), 299–338.

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

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

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

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

Chernozhukov, Victor and Han Hong, "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 2003, *115* (2), 293–346.

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," *Econometrica*, 2009. forthcoming.

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," *Mathematics of Operations Research*, August 2008, *33* (3), 757–768.

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

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

_ , "Identification in Matching Games," June 2009. University of Chicago working paper.

_ **and Patrick Bajari**, "Measuring the Efficiency of an FCC Spectrum Auction," 2009. University of Chicago working paper.

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, *121* (4), 1383–1435.

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

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

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, "Matching and Sorting in Online Dating Markets," *American Economic Review*, 2009.

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

_ , "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," May 2008. working paper.

Jun, Sung Jae, Joris Pinkse, and Yuanyuan Wan, "Cube-Root-N and Faster Convergence, Laplace Estimators, and Uniform Inference," 2009. Pennsylvania State University working paper.

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.

Levine, Anna A., "Licensing and Scale Economies in the Biotechnology Pharmaceutical Industry," June 2008. Harvard University working paper.

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.

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.

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

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

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

Mindruta, Denisa, "Value Creation in University-Firm Research Collaborations: A Matching Approach," February 2009. HEC Paris working paper.

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. MIT working paper.

_ **and Scott Stern**, "How Does Outsourcing Affect Performance Dynamics? Evidence from the Automobile Industry," *Management Science*, December 2008, *54* (12), 1963–1979.

_ **and** _ , "Complementarity Among Vertical Integration Decisions: Evidence from Automobile Product Development," *Management Science*, 2009, *55* (2), 311–332.

_ **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," *American Economic Review*, December 2008, *98* (3), 897–923.

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," November 2006. Harvard University 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. UCLA working paper.

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

Rosen, Sherwin, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 1974.

_ , "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.

Santiago, David and Jeremy T. Fox, "A Toolkit for Matching Maximum Score Estimation and Point and Set Identified Subsampling Inference," January 2007. University of Chicago working paper.

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

Sørensen, Morten, "How Smart is Smart Money? 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.

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.

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

Yang, Yupin, Mengze Shi, and Avi Goldfarb, "Estimating the Value of Brand Alliances in Professional Team Sports," *Marketing Science*, 2009.

Table 1: 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 $\epsilon_{\langle a,i \rangle}$ 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 2.1.

The distribution of the fixed agent types is given in the text. On the left table, the errors $\epsilon_{\langle a,i \rangle}$ 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.

Table 2: 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			

Table 3: 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	

Table 4: 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	

Table 5: Artificial inequalities: Reconciling different point estimates between Tables 3 and 4

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	