

NBER WORKING PAPER SERIES

IDENTIFICATION IN MATCHING GAMES

Jeremy T. Fox

Working Paper 15092

<http://www.nber.org/papers/w15092>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

June 2009

Previous versions of these theorems circulated as part of the paper “Estimating Matching Games with Transfers”, which now focuses only on estimation, not identification. I thank the National Science Foundation, the NET Institute, the Olin Foundation, and the Stigler Center for generous funding. 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.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2009 by Jeremy T. Fox. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Identification in Matching Games
Jeremy T. Fox
NBER Working Paper No. 15092
June 2009
JEL No. C14,C3,C35,C71,D4,L22

ABSTRACT

I study a many-to-many, two-sided, transferable-utility matching game. Consider data on matches or relationships between agents but not on the choice set of each agent. I investigate what economic parameters can be learned from data on equilibrium matches and agent characteristics. Features of a production function, which gives the surplus from a match, are nonparametrically identified. In particular, the ratios of complementarities from multiple pairs of inputs are identified. Also, the ordering of production levels is identified.

Jeremy T. Fox
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
fox@uchicago.edu

1 Introduction

Matching games are a new and important area of empirical interest. Consider the classic example of marriage. A researcher may have data on a set of marriages in each of a set of independent matching markets, say a set of towns. The researcher observes characteristics of each man and each woman in each town, as well as the sets of marriages that occurred. The researcher observes equilibrium outcomes, here marriages, and not choice sets, so identification in this type of model will not be able to rely trivially on the analysis of single-agent demand models. What type of parameters can be identified from these data?

Economists have studied nonparametric or semiparametric identification in auction games of private information (Elyakime, Laffont, Loisel and Vuong, 1994) as well as discrete games of complete information (Berry and Tamer, 2006) and incomplete information (Bajari, Chernozhukov, Hong and Nekipelov, 2009). This previous literature is unified in using noncooperative Nash equilibrium as the solution concept. Matching games are cooperative games and use pairwise stability instead of Nash equilibrium as the main solution concept. This is the first paper to study identification in a new and empirically important class of games.

I follow the classic works of Koopmans and Beckmann (1957), Shapley and Shubik (1972), and Becker (1973) and model the formation of matches (say marriages) as the outcome to a competitive market. Agents have preferences over partners and agents can exchange monetary transfers with their spouse. The equilibrium concept is pairwise stability: in part, at an outcome to a marriage market, no man would prefer to pay the transfer required to be able to marry any woman other than his actual wife in the proposed equilibrium. I assume the researcher does not have access to data on the transfers. For marriage, transfers are a modeling abstraction. For interfirm relationships, transfers may be private contractual details. Therefore, this paper studies point identification under partial observability of the outcome variables in the model.

Match data come from the outcome to a market, which intermingles the preferences of all participating agents and finds an equilibrium. An agent may not match with its most preferred partner because that partner is taken. Because agents on the same side of the market are rivals to match with potential partners, the failure for a match to form does not mean that the match gives low production. Given this rivalry for partners, it is not obvious what types of economic parameters are identified from having equilibrium outcome data from matching markets. Identification asks the question of just what economic parameters can be learned from data on who matches with whom? A production function gives the total output of a match. I prove that aspects of match production functions can be identified in a transferable-utility setting using data on only equilibrium matches. Identification relies on inequalities implied by the equilibrium concept, pairwise stability.

I first study what I label derivative-based identification, as the features of production functions one can learn about may involve derivatives. Derivative-based identification using qualitative match data arises because certain derivatives of match production functions govern sorting patterns in transferable-utility matching games. For example, a cross-partial derivative of the production function represents the importance of complementarities between a pair of characteristics, each from a different agent. Becker (1973) shows that complementarities result in assortative matching. I extend the results of Becker in several dimensions. For example, I show how to identify the ratio of complementarities in two pairs of agent-specific characteristics in a match production function, say the relative importance of wealth and schooling. This allows a researcher to measure the relative importance of complementarities on different pairs of characteristics. This is equivalent to a multivariate (multiple pairs of characteristics) analysis, while Becker's analytical characterization of the sorting pattern requires that each agent is distinguished by only a single characteristic.

Second, I ask whether a researcher can identify the relative ordering of match production for different types of matches. I learn whether match production is higher at one set of characteristics for the matched parties than another set of characteristics. I extend identification results from the single-agent, multinomial choice literature by Matzkin (1993). The extension is non-trivial because one cannot freely vary the choice set of a single agent when using data that are the equilibria to matching games. In matching with transfers, one must pay a potential partner to match with you, and the required payment involves the characteristics of rival agents. I prove the identification of match production functions, up to a positive monotonic transformation, by varying the exogenous distribution of the types of agents in a matching market.

My identification arguments do not require data on objects that are not found in many datasets but are important in matching models: the endogenous prices, the number of physical matches that an agent can make (quotas), or continuous outcomes such as production levels, revenues and profits. Quotas are often a modeling abstraction in many-to-many matching; not requiring data on such an abstraction is an advantage. Transfers, production levels, revenues and profits are often not recorded at all (marriage) or not disclosed (interfirm relationships).

The identification arguments are for a fairly general class of matching games. I study many-to-many, two-sided matching games. This means each potential agent may be involved in multiple matches with agents on the other side of the market. This generality is essential to applications in industrial organization where, for example, one supplier of goods may match to many retailers of those goods. I do not require a supplier's profit function to be additively separable across the characteristics of its multiple partners. I also prove separate results for three different types of observable characteristics that may enter the payoff of a group of matches: agent-specific characteristics, match-specific characteristics, and characteristics that vary for each group of matches. Importantly, each agent, match or group of matches may have a vector of characteristics.

In many-to-many matching games, I study the identification of production functions that can take as arguments the characteristics of many partners at once. In many-to-many matching, pairwise stability is a weaker solution concept than another solution concept, the core. One mathematical achievement of the paper is that all identification results use only the restrictions from pairwise stability, which as its name indicates, allows only a single pair of potential partners to consider deviating from the proposed equilibrium at once. This achievement is important because the communication volume necessary to believe an equilibrium is in the core is large: arbitrarily large groups of agents would need to coordinate their actions. Theorists are often comfortable specifying that a decentralized matching game's outcome is pairwise stable, but assuming that the outcome is in the core would be more controversial.

The identification arguments are completely nonparametric: I do not impose that production functions and the stochastic structure of the model are known up to a finite vector of parameters. The stochastic structure of the model uses a rank order assumption that is inspired by the maximum-score literature on single-agent multinomial choice (Manski, 1975; Matzkin, 1993; Fox, 2007). This maximum-score identification approach allows me to work with inequalities that are derived from pairwise stability, rather than working with high-dimensional integrals over match-specific unobservables. This allows me to focus on the matching-market configurations that lead to identification. In the single-agent, multinomial-choice papers, a similar rank order assumption is derived as a consequence of the payoff of each discrete choice having an i.i.d. unobservable component. I discuss in some detail why an i.i.d. unobservable component to each match's payoff does not give the necessary rank order property in matching. I discuss alternative sufficient conditions. I also provide

simulation evidence about how much the rank order property is violated in models with i.i.d. unobservable match-specific components.

Another advantage of the maximum-score rank order assumption is that the identification arguments lead to a computationally simple estimator. Matching markets can have hundreds or thousands of agents in them. In Fox (2009), I present such a maximum-score estimator for matching games and show how it resolves two curses of dimensionality in the number of agents in a matching market: a computational curse of dimensionality from otherwise needing to compute or check equilibria and a data curse of dimensionality that might arise from the need to nonparametrically estimate matching probabilities as a function of all agent characteristics in a matching market. Simultaneously with this paper on identification, I have undertaken two empirical applications of the estimator. In Fox and Bajari (2009), we study matching between bidders and licenses for sale in a FCC spectrum auction. In Fox (2009), I study matching between automotive parts suppliers and automotive assemblers. In both cases, the datasets are fairly large and complementarities between multiple matches for the same agent are essential aspects of the empirical investigation. Also, the various types of characteristics (agent, match and group of matches) are all used in the empirical work. So the generality this paper strives for is used in my empirical applications. More recently, Akkus and Hortacsu (2007), Levine (2008), Mindruta (2009), and Yang, Shi and Goldfarb (2009) have conducted empirical work using the matching maximum-score estimator in Fox (2009). The identification results here are directly relevant for the above empirical papers.

No paper has performed a completely nonparametric analysis of any sort of matching game. Under many non-nested assumptions with this paper, Choo and Siow (2006) provide a logit-based estimator for one-to-one matching games using aggregate data. They identify surplus functions conditional on the parametric structure of the logit model for the error terms. Dagsvik (2000) also uses the logit model. A literature has explored parametric estimation in the non-nested Gale and Shapley (1962) matching games, i.e. games without endogenous transfers (Boyd et al., 2003; Gordon and Knight, 2009; Sørensen, 2007). These papers do not study identification.

One advantage of studying one-to-one, two-sided matching games (such as marriage) is that a pairwise stable match always exists and is unique with probability 1. Therefore, I present the intuition for the identification strategies and results first for the example of marriage, in Section 2. Once the intuition for the results has been presented, Section 3 introduces more general notation for many-to-many, two-sided matching games with several types of observable characteristics. Section 4 presents sufficient conditions for the maximum-score-like rank order property. These two sections mention existence and uniqueness issues that are not found in marriage games. The last two sections present the main identification theorems. Section 5 discusses derivative-based identification and Section 6 discusses the identification of orderings of production levels. The proofs of the theorems are in the appendix.

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

This section presents informal results for the case of one-to-one, two-sided matching. One-to-one matching was studied in Koopmans and Beckmann (1957), Shapley and Shubik (1972), and Becker (1973) and has been summarized in Roth and Sotomayor (1990, Chapter 8). The results in this section focus on intuition because there will be no formal definition of identification and hence no formal proofs. The formal details will be shown for the many-to-many matching case in Sections 5 and 6.

2.1 Model setup

This section studies a market with two men, a and b , and two women, i and j . Each agent has two observable characteristics: schooling and wealth. To simplify notation in this section, let agents be identified by their characteristics. So in a duplication of notation, $a = (a_1, a_2)$, where the vector of man a 's characteristics is equal to man a 's schooling a_1 and wealth a_2 . The same holds for man b and women i and j .

Each person can be married only once, to a person of the opposite sex. A key object of interest in this paper is the production from a match. If a marries i , they produce output $f(a, i) = f((a_1, a_2), (i_1, i_2))$, which is a function of the schooling of the man and woman and the wealth of both. Being single is allowed. If man a is unmatched, write his output as $f(a, 0)$, where we say he is matched to a dummy agent 0. The restriction to two men and two women (or equal numbers of men and women) is for expositional convenience; the restriction will not be present in the general results on many-to-many matching games.

Becker (1973) shows that if each man and each woman has only one characteristic, say schooling but not wealth, then, in a pairwise stable equilibrium, men and women will assortatively match if the schooling of men and women are complements in production

$$\frac{\partial^2 f(a, i)}{\partial a_1 \partial i_1} > 0.$$

Assortative matching means highly-schooled men match with highly-schooled women. Likewise, anti-assortative matching occurs if schooling levels of men and women are substitutes in production. My setup is already more general: each agent has a vector of characteristics $a = (a_1, a_2)$. Becker's result does not apply to this case. No previous matching theory paper provides an analytical characterization of the equilibrium sorting pattern using simple production-function properties such as complementarities when agents have vectors of characteristics and the exogenous distribution of agent characteristics is unrestricted. Numerical analysis is needed. One goal of this paper is to extend Becker's identification analysis to a host of cases where microeconomic theory has not analytically characterized the equilibrium sorting pattern based on simple production-function properties.

The focus on a production function $f(a, i)$ as an object of interest can be motivated by a model where men have utility functions over the characteristics of women, women have utility functions over the characteristics of men, and men and women exchange monetary transfers, which enter additively separably into utility. The researcher will not have data on the transfers. Details on this more primitive motivation will be given for the general case below.

2.2 Data on multiple matching markets

The econometrician has access to data on $\{A, a, b, i, j\}$ for each member of a population of matching markets. Think of each market as a very small town with two men and two women; the researcher observes their characteristics a, b, i and j . Likewise, the researcher observes the assignment A in each market. A is the set of matches that we observe. For example, if a and i as well as b and j marry, $A = \{\langle a, i \rangle, \langle b, j \rangle\}$, where $\langle a, i \rangle$ is the match between man a and woman i . If b and j are instead single, $A = \{\langle a, i \rangle, \langle b, 0 \rangle, \langle 0, j \rangle\}$.

Readers may be familiar with single-agent discrete-choice models, such as the parametric logit model or the semiparametric maximum-score model. It might be helpful to think of matching by analogy to single-agent choice: the independent variables are the agent characteristics a, b, i and j and the qualitative dependent variable is the assignment A . Of course, the underlying data-generating process for A is an equilibrium model and not a single-agent choice model.

For the purposes of intuitive identification, assume that all characteristics a , b , i and j vary across markets and are continuous random variables with (if needed) full support. Let the support be a product space for all eight characteristics (recall each of the four agents has two characteristics), so that there is some matching market with each combination of the eight characteristics. For conciseness, let $X = \{a, b, i, j\}$, or the eight observable characteristics. For identification, the researcher has access to the population data on i.i.d. market observations (X, A) and hence can identify the joint distribution of A and X . Let $\Pr(A | X)$ be the probability of observing assignment A given agent characteristics X . $\Pr(A | X)$ is identifiable given the joint distribution of A and X .

2.3 The rank order property and identification

For the case of one-to-one matching (this property will not generalize to many-to-many matching), Koopmans and Beckmann (1957), Shapley and Shubik (1972), and Becker (1973) prove that any pairwise stable assignment A will maximize the sum of production of all matches in the economy, or A will maximize $\sum_{\langle a, i \rangle \in A} f(a, i)$, where $\langle a, i \rangle$ is an arbitrary match in the feasible assignment A . An assignment is feasible for marriage if each agent has at most one spouse and the spouse is of the opposite gender. As is common in empirical work, often two markets with the same observable agent characteristics X will have different assignments A . The econometric model should ideally allow a positive probability for observing any feasible assignment, or $\Pr(A | X) > 0$ for all feasible A , given X . This is equivalent to the logit model's property of giving positive probability to all choices.

The main property identification will be based upon is the rank order property. This is inspired by related conditions in the literature on maximum-score estimation of the single-agent, multinomial choice model (Manski, 1975; Matzkin, 1993; Fox, 2007). Let A_1 and A_2 be two different feasible assignments for the same matching market. A strong version of the rank order property states that $\Pr(A_1 | X) > \Pr(A_2 | X)$ if and only if $\sum_{\langle a, i \rangle \in A_1} f(a, i) > \sum_{\langle a, i \rangle \in A_2} f(a, i)$. For example, focus on the two assignments where no agent is single. Let $A_1 = \{\langle a, i \rangle, \langle b, j \rangle\}$ and $A_2 = \{\langle a, j \rangle, \langle b, i \rangle\}$. Given observable agent characteristics in X and knowledge of f , the property states that assignment A_1 is more frequently observed if and only if

$$f(a, i) + f(b, j) > f(a, j) + f(b, i). \quad (1)$$

For marriage, the rank order property is a stochastic extension of the deterministic idea from the theory literature that the equilibrium assignment maximizes production and assignments that do not maximize production do not occur. Under the rank order property, all assignments can occur, and their frequencies are rank ordered by their sums of production.

Let f^0 be the true production function and let f^1 be some other production function. Identification will require us to find a set of market characteristics X where (1) holds for $f = f^0$ and

$$f(a, i) + f(b, j) \leq f(a, j) + f(b, i) \quad (2)$$

holds for $f = f^1$. The inequality (1) for $f = f^0$ implies that $\Pr(A_1 | X) > \Pr(A_2 | X)$ in the population data, while (2) for $f = f^1$ implies $\Pr(A_1 | X) \leq \Pr(A_2 | X)$ if f^1 happened to generate the data, which it does not. $\Pr(A_1 | X) > \Pr(A_2 | X)$ and $\Pr(A_1 | X) \leq \Pr(A_2 | X)$ are exclusive possibilities. Therefore, the production function

f^0 is (informally) identified if for any $f^1 \neq f^0$ we can find a set of eight market observables $X = \{a, b, i, j\}$ where (1) holds for $f = f^0$ and (2) holds for $f = f^1$. This particular X will be decisive because f^0 and f^1 give different implications for the population data $\Pr(A | X)$.

2.4 Derivative-based identification

Qualitative data on who matches with whom in equilibrium will not be enough to identify production functions in full generality. At a minimum, multiplying each f by a positive constant will preserve the inequality in (1). Thus, this paper focuses on identifying features of f^0 .

2.4.1 Are two inputs complements or substitutes at a point?

The first feature of f^0 that that will be identifiable is the sign of $\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1}$, or whether the schooling levels of men and women are complements or substitutes in production. Here a and i should be take as any arbitrary male and female characteristics. This extends Becker (1973) in two ways: each a has two characteristics schooling and wealth, not just schooling, and the signs of $\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1}$ can be positive for some couple characteristics (a,i) and negative for other (a,i) . The sign of $\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1}$ will be learned for each a and i separately, and so the signs will be known for all points of the support of f . Thus, the analysis does not rely on inputs being complements or substitutes for all values (a,i) .

For the sake of argument, let $\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1} > 0$ and so, to show identification, let f^1 be some other production function where $\frac{\partial^2 f^1(a,i)}{\partial a_1 \partial i_1} < 0$. We will need to find some characteristics $X = \{a, b, i, j\}$, eight characteristics as each element is a vector, where (1) holds for $f = f^0$ and (2) holds for $f = f^1$.

In what follows, assume each f is three-times differentiable, so that cross-partial derivatives are symmetric. Then a cross-partial derivative can be expressed as the limit of a middle-difference quotient:

$$\frac{\partial^2 f(a,i)}{\partial a_1 \partial i_1} = \lim_{h \rightarrow 0} \frac{f((a_1 + h, a_2), (i_1 + h, i_2)) - f((a_1 + h, a_2), (i_1, i_2)) - f((a_1, a_2), (i_1 + h, i_2)) + f((a_1, a_2), (i_1, i_2))}{h^2}, \quad (3)$$

where h is the limit argument in the expression for the cross-partial derivative. The value of (a,i) where we wish to identify the sign of $\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1}$ is given. We will work with markets with observables X of the form $X = \{(a_1, a_2), (a_1 + h, a_2), (i_1, i_2), (i_1 + h, i_2)\}$. In the previous notation for the two men and two women, $a = (a_1, a_2)$, $b = (a_1 + h, a_2)$, $i = (i_1, i_2)$, and $j = (i_1 + h, i_2)$. Men a and b have identical observable characteristics except that b has h more units of schooling than a . Likewise, women i and j are identical except that j has h more units of schooling.

The numerator of the middle-difference quotient in (3) for $f = f^0$ will be positive for sufficiently small h , because $\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1} > 0$. Likewise for sufficiently small $h > 0$, the numerator of the middle-difference quotient in (3) for $f = f^1$ will be negative. Let h be sufficiently small so that both the previous statements hold. For $f = f^0$, we can rearrange the positive numerator to give

$$f((a_1, a_2), (i_1, i_2)) + f((a_1 + h, a_2), (i_1 + h, i_2)) > f((a_1, a_2), (i_1 + h, i_2)) + f((a_1 + h, a_2), (i_1, i_2)).$$

This is just a special case of (1). Likewise, the opposite inequality will hold for $f = f^1$, which is just a special case of (2). Now let there be two hypothetical assignments, $A_1 = \{ \langle a, i \rangle, \langle b, j \rangle \}$ and $A_2 = \{ \langle a, j \rangle, \langle b, i \rangle \}$. So,

by the above rank order property, f^0 implies $\Pr(A_1 | X) > \Pr(A_2 | X)$ and f^1 implies the reverse, so we have identification of f^0 and hence we learn the sign of $\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1}$. We can use data on the equilibrium outcomes to matching markets to learn whether two inputs or complements or substitutes in production.

What is the economic intuition? Given the true f^0 and the alternative f^1 , we were able to find a set of matching market observables X where f^0 and f^1 gave different predictions about the relative frequencies of two assignments. Here, f^0 predicted that, in markets with this X , agents would assortatively match on schooling in more markets than they would anti-assortatively match, while f^1 predicted anti-assortatively matching would occur in more markets. We can look at the population data on $\Pr(A | X)$ to see which is more common.

Note that this result does not allow a researcher to tell whether a pair of inputs are “more” complementary at (a,i) than some other point (c,k) . For example, $\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1} = 5$ and $\frac{\partial^2 f^0(c,k)}{\partial c_1 \partial k_1} = 7$ cannot be distinguished from any other pair of positive values.

2.4.2 How important are complementarities for one pair of inputs compared to another pair?

Becker (1973) does not allow agents to have a vector of characteristics, here schooling and wealth. A natural question to ask is how much more important are complementarities between schooling levels of men and women compared to complementarities between wealth levels of men and women? This is a horse race analysis: the multivariate model has multiple characteristics and we wish to identify the relative importance of the different pairs of characteristics in match production. If the previous analysis is analogous to regressing a dependent variable on one regressor, this analysis will be equivalent to regressing a dependent variable on two regressors.

Say both schooling and wealth are complements: $\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1} > 0$ and $\frac{\partial^2 f^0(a,i)}{\partial a_2 \partial i_2} > 0$. We will identify the ratio of the complementarities of schooling to the complementarities of wealth, or

$$\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1} / \frac{\partial^2 f^0(a,i)}{\partial a_2 \partial i_2}.$$

As before, the analysis is local: for a given value of male and female characteristics (a,i) . We can establish these ratios globally by varying (a,i) . Also note that we are identifying the ratio of complementarities, which is an actual numerical value. This will be harder than identifying whether $\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1} > \frac{\partial^2 f^0(a,i)}{\partial a_2 \partial i_2}$, which is a qualitative comparison instead of a quantitative value.¹

Let $f^1 \neq f^0$ be some other production function. Because we can use the previous arguments to identify whether any pair of inputs are complements or substitutes, we can restrict attention to the case where $\frac{\partial^2 f^1(a,i)}{\partial a_1 \partial i_1} > 0$ and $\frac{\partial^2 f^1(a,i)}{\partial a_2 \partial i_2} > 0$ but, without loss of generality,

$$\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1} / \frac{\partial^2 f^0(a,i)}{\partial a_2 \partial i_2} > \frac{\partial^2 f^1(a,i)}{\partial a_1 \partial i_1} / \frac{\partial^2 f^1(a,i)}{\partial a_2 \partial i_2}.$$

We will need to embellish the running example and allow each matching market to have three men and three women. Let the men be a, b , and c and let the women be i, j , and k . Let all men start at the baseline characteristics a . Man $a = (a_1, a_2)$ is the baseline man. Let man b have h_1 extra units of schooling, $b = (a_1 + h_1, a_2)$. Let

¹The identification of $\frac{\partial^2 f^0(a,i)}{\partial a_1 \partial i_1} / \frac{\partial^2 f^0(a,i)}{\partial a_2 \partial i_2}$ seems parallel to the identification of marginal rates of substitution in single-agent choice. The ratio of marginal utilities is preserved under positive monotonic transformations. However, the ratio of cross-partial derivatives is not preserved under positive monotonic transformations.

man c have h_2 extra units of wealth, $c = (a_1, a_2 + h_2)$. Likewise, i is the baseline for the three women. Woman j has h_1 extra units of schooling and woman k has h_2 extra units of wealth. Now $X = \{a, b, c, i, j, k\}$.

The formal argument using the limits of middle-difference quotients is somewhat technical and will appear in the proof to Theorem 5.1 for the many-to-many matching case, below. For now, this omitted argument will give particular values of h_1 and h_2 , the extra schooling and the extra wealth, where key inequalities hold. In particular, under $f = f^0$ and these choices of h_1 and h_2 ,

$$\begin{aligned} & f((a_1 + h_1, a_2), (i_1 + h_1, i_2)) + f((a_1, a_2 + h_2), (i_1, i_2)) + f((a_1, a_2), (i_1, i_2 + h_2)) \\ & > f((a_1, a_2 + h_2), (i_1, i_2 + h_2)) + f((a_1 + h_1, a_2), (i_1, i_2)) + f((a_1, a_2), (i_1 + h_1, i_2)), \end{aligned}$$

or, equivalently,

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

Meanwhile, under $f = f^1$ the reverse inequality will hold. What is going on? On the left side, there is the total production from an assignment $A_1 = \{\langle b, j \rangle, \langle c, i \rangle, \langle a, k \rangle\}$ where the man and woman with h_1 extra units of schooling (man b and woman j) marry. Also, the man and woman with h_2 extra units of wealth (c and k) each marry baseline individuals (i and a). On the right side, there is the total production from an assignment $A_2 = \{\langle c, k \rangle, \langle b, i \rangle, \langle a, j \rangle\}$ where the couple where both have h_2 of extra wealth (c and k) marry but the man and woman with h_1 of extra schooling each marry a baseline person (b marries i and j marries a).

We have found a set of observable characteristics $X = \{a, b, c, i, j, k\}$ and two corresponding assignments $A_1 = \{\langle b, j \rangle, \langle c, i \rangle, \langle a, k \rangle\}$ and $A_2 = \{\langle c, k \rangle, \langle b, i \rangle, \langle a, j \rangle\}$ where f^0 and f^1 give different implications for the comparison of total, deterministic production. So f^0 implies $\Pr(A_1 | X) > \Pr(A_2 | X)$ and f^1 implies the reverse. The economic intuition is easy to understand. Assignment A_1 has assortative matching on schooling but anti-assortative matching on wealth and A_2 has assortative matching on wealth but anti-assortative matching on schooling. If, at these choices for X , assortative matching on schooling and anti-assortative matching on wealth occur in more markets than assortative matching on wealth and anti-assortative matching on schooling, then the true ratio of complementarities is $\frac{\partial^2 f^0(a, i)}{\partial a_1 \partial i_1} / \frac{\partial^2 f^0(a, i)}{\partial a_2 \partial i_2}$ instead of $\frac{\partial^2 f^1(a, i)}{\partial a_1 \partial i_1} / \frac{\partial^2 f^1(a, i)}{\partial a_2 \partial i_2}$. The value of $\frac{\partial^2 f^0(a, i)}{\partial a_1 \partial i_1} / \frac{\partial^2 f^0(a, i)}{\partial a_2 \partial i_2}$ is identified because f^1 was arbitrary.

2.5 Identification of orderings of production levels

The derivative-based identification analysis is sufficient for many matching empirical applications. However, it is impossible to use the derivative-based analysis to tell apart two production functions with the same cross-partial derivatives. Briefly let schooling be only the agent characteristic. The production functions $f^1(a_1, i_1) = -(a_1 - i_1)^2$ and $f^2(a_1, i_1) = 2a_1 \cdot i_1$ both have $\frac{\partial^2 f(a_1, i_1)}{\partial a_1 \partial i_1} = 2$. For f^1 , $f^1(1, 1) = 0$ and $f^1(2, 1) = -1$. For f^2 , $f^2(1, 1) = 2$ and $f^2(2, 1) = 4$. So $f^1(1, 1) > f^1(2, 1)$ yet $f^2(1, 1) < f^2(2, 1)$. Under f^1 , each man's marriage's production is highest when the man matches with a woman with the same level of schooling. There is no single ideal female schooling attainment for all men. Under f^2 , any man's marriage's surplus will be maximized by matching with the most-educated woman.

Now return to there being two characteristics, wealth and schooling, for each man and each woman. Identifying cross-partial derivatives does not tell us whether production is higher at one argument (a, i) than another argument. For any two sets of characteristics for a pair of matches (a, i) and (b, j), we identify whether

$f^0(a, i) > f^0(b, j)$ or the reverse. Ordering production function levels is helpful in distinguishing whether an individual match characteristic such as schooling is actually a “good” that raises output.

Matzkin (1993) proves the identification of utility functions for the single-agent multinomial-choice model. Following Matzkin, restrict attention to a class of production functions where no two members are related by a positive monotonic transformation. Let f^0 be the true production function and f^1 some alternative not related to f^0 by a positive monotonic transformation. Matzkin proves there exists two points (a, i) and (b, j) where, without loss of generality, $f^0(a, i) > f^0(b, j)$ and $f^1(a, i) < f^1(b, j)$. For the single-agent model, identification is then proved using a rank order property: a consumer with utility function f^0 picks the product with characteristics (a, i) more frequently than the product with characteristics (b, j) . A consumer with utility function f^1 does the reverse. Data on the frequency of choice will show f^1 is not the correct utility function.

In the single-agent model, one can vary the characteristics of the choices facing the single agent. In a matching market, an agent must pay the appropriate transfer to match with a partner, and that transfer is both an outcome of the game and assumed to not be in the data. Therefore, I extend the mathematical arguments in Matzkin to show the identification of the production function f by using only exogenous information on X , the collection of characteristics of all agents and potential matches in a matching market. In other words, I work with the equilibrium structure of the game and variation in the exogenous market-level characteristics of matches to show identification.

We need to transform the inequality $f(a, i) > f(b, j)$ for $f = f^0$, or

$$f((a_1, a_2), (i_1, i_2)) > f((b_1, b_2), (j_1, j_2)), \quad (4)$$

into an inequality that compares assignments A_1 and A_2 for the same market with observable characteristics X . The problem is that agents with different characteristics appear on the left and right sides of $f(a, i) > f(b, j)$. Focusing on single people will resolve this dilemma. Add the payoff for each agent being single to each side of (4). Adding identical terms to each side preserves the inequality:

$$\begin{aligned} f((a_1, a_2), (i_1, i_2)) + f((a_1, a_2), 0) + f(0, (i_1, i_2)) + f((b_1, b_2), 0) + f(0, (j_1, j_2)) > \\ f((b_1, b_2), (j_1, j_2)) + f((a_1, a_2), 0) + f(0, (i_1, i_2)) + f((b_1, b_2), 0) + f(0, (j_1, j_2)). \end{aligned} \quad (5)$$

This inequality almost involves two assignments A_1 and A_2 to the same market, X . An issue is that a man with characteristics $a = (a_1, a_2)$ appears on the left side twice, as does a woman $i = (i_1, i_2)$. Similarly, man b and woman j appear twice on the right side. One commonly-used assumption in the theory literature, for example Koopmans and Beckmann (1957) and Shapley and Shubik (1972), is that the payoff to being single is 0. If this is the case, we can choose to set the production of certain unmarried agents in (5) to 0, giving

$$f((a_1, a_2), (i_1, i_2)) + f((b_1, b_2), 0) + f(0, (j_1, j_2)) > f((b_1, b_2), (j_1, j_2)) + f((a_1, a_2), 0) + f(0, (i_1, i_2)).$$

We could set all single matches to 0, but doing so would return us to (4). On the left side, we have an assignment $A_1 = \{(a, i), \langle b, 0 \rangle, \langle 0, j \rangle\}$ and on the right side we have an assignment $A_2 = \{\langle b, j \rangle, \langle a, 0 \rangle, \langle 0, i \rangle\}$. The production function f^0 implies $\Pr(A_1 | X) > \Pr(A_2 | X)$ and f^1 implies the reverse, for $X = \{a, b, i, j\}$.

The economic intuition here is simple. If we wish to know whether or not $f^0(a, i) > f^0(b, j)$, we merely need to see the relative frequency of a and i being married and b and j being single compared to b and j being

married and a and i being single. The couple with the higher marital production will be single less often. Thus, we can learn the orderings of production functions: we can identify production functions up to positive monotonic transformations.

The need to use single people arises because the example used agent-specific characteristics. Identification of orderings of production functions will be proved below for match-specific and group-of-matches-specific characteristics. Those results will not rely on single people.

2.6 Uses of matching estimation in empirical work

Before starting with the many-to-many notation and results, I should stop and discuss the role of matching estimation (and hence identification) in empirical work. One goal is to distinguish the role of the distribution of exogenous agent characteristics from the role of the production function in the sorting pattern we see in the data. For example, Choo and Siow (2006) find changes in the sorting patterns between broad types of men and women across decades in the United States, and in part ask whether changes in match production functions or changes in agent characteristics are behind the differences in sorting patterns.

The identification results in this paper are specifically referenced in Fox and Bajari (2009). In that paper, we estimate the production function for companies matching to geographic mobile-phone licenses in a FCC spectrum auction. We use our estimates to measure the efficiency of the assignment in the auction (we do not use the above production maximization arguments to motivate the inequalities). We use qualitative data on which bidders win what packages of licenses. First, we use the derivative-based identification arguments to cardinally measure the relative complementarities of higher-value bidders sorting to packages with greater scale and the complementarities of packages of geographically nearby licenses being grouped together. Second, we use the results on the orderings of production levels to argue that both of these types of bidder and license characteristics are “goods” that raise output. This lets us use our production functions to measure (up to scale) the total output from counterfactual assignments of bidders to licenses and to measure how much efficiency is lost from the actual assignment in the auction.

3 Many-to-many matching games

The rest of the paper studies the general case of 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 my empirical work elsewhere has illustrated (Fox and Bajari, 2009; Fox, 2009). This section outlines a two-sided, many-to-many matching game without econometric errors. Simpler models such as marriage are special cases. In the next section, I discuss how to extend these models to introduce econometric error terms. The running example will be downstream firms (think automotive manufacturers such as General Motors and Toyota) matching with upstream firms (for automotive parts suppliers, think Bosch and Johnson Controls).

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 identification strategy used 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.² 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

Notationally, I drop the equivalence between agent indices and their characteristics because I will allow for all of firm-, match- and group-specific characteristics.

Several exogenous objects define a matching market. Let U be a finite set of upstream firms, indexed by i . Let D be a finite set of downstream firms, indexed by a . Let $Q : U \cup D \rightarrow \mathbb{N}_+$ be the set of quotas, where $q_a^d \in Q$ is the quota of a downstream firm a and $q_i^u \in Q$ is the quota of the upstream firm i . A quota represents the maximum number of physical matches that a firm can have. Let X be the collection of all payoff-relevant exogenous characteristics. I will be specific about the elements of X below. A matching market also has the exogenous preferences of agents, which I will also discuss below.

Let $\mu = \langle a, i \rangle$ be a match between downstream firm a and upstream firm i . If $q_a^d > 1$, a downstream firm, a , say, may be part of multiple matches. As before, $\mu = \langle a, 0 \rangle$ refers to an unfilled quota slot for a downstream firm and $\mu = \langle 0, i \rangle$ refers to an unfilled quota slot for an upstream firm. 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, $A = \{\mu_1, \mu_2, \dots, \mu_N\}$, $T = \{t_{\mu_1}, t_{\mu_2}, \dots, t_{\mu_N}\}$ is a set of transfers for all matches in A . Each $t_{\mu} \in \mathbb{R}$ and represents a payment by a downstream firm to an upstream firm. I use the convention that the downstream firm is sending positive transfers to the upstream firm, but the notation allows transfers to be negative. In a market with 100 upstream-downstream relationships, A is a finite set of 100 relationships and T is a finite set of 100 transfers between each of the matched partners. Altogether, the combination of the exogenous (D, U, Q, X) and endogenous (A, T) elements of a matching market is the tuple (D, U, Q, X, A, T) .

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

$$r^u(\bar{x}(i, C_i^u(A))) + \sum_{a \in C_i^u(A)} t_{\langle a, i \rangle}. \quad (6)$$

Here, $C_i^u(A)$ is a collection of downstream firms i is matched to in assignment A and $r^u(\cdot)$ is the structural revenue function for upstream firms. The payoff at (A, T) for $a \in D$ for the match $\langle a, i \rangle \in A$ is $r^d(\bar{x}(i, \{a\})) - t_{\langle a, i \rangle}$. I will now explain each of the elements of the notation that enters these payoffs.

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

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

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

²Lucas (1978) and Rosen (1982) are predecessors to Garicano and Rossi-Hansberg (2006).

similar interpretations for downstream firms.

A feasible assignment A is one that is under quotas for all agents. This means $C_i^u(A) \leq q_i^u \forall i \in U$ and $C_a^d(A) \leq q_a^d \forall a \in D$. Quotas ensure that firms are rivals to match with the most attractive partners on the other side, as opposed to all firms choosing the most attractive partner.³

This paper studies matching in characteristic space. Let $\vec{x}(i, C^u)$ be the vector of characteristics corresponding to the set of matches involving firm $i \in U$ and the set of potential downstream firm partners in C^u .⁴ I will consider three types of characteristics. First consider the case studied in Section 2, where each agent has a fixed type, a vector \vec{x}_a^d for downstream firm a and a vector \vec{x}_i^u for upstream firm i . For example, \vec{x}_a^d could be the geographic location of a 's assembly plant, information on the cars manufactured by a , the markets a sells to, etc. Likewise, \vec{x}_i^u could be the geographic location of upstream firm i , the past experience of the upstream firm, etc. Allowing types to be vectors is an important extension of existing theoretical work. In this case, $\vec{x}(i, C^u) = \text{cat}(\vec{x}_i^u, \vec{x}_{a_1}^d, \dots, \vec{x}_{a_n}^d)$, where $C^u = \{a_1, \dots, a_n\}$.⁵

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

X is the set of characteristics for all firms for firm-specific characteristics, all potential matches for match-specific characteristics, and all potential groups of matches in a market, whether the matches are part of a particular assignment or not. Formally,

$$X = \left(\bigcup_{i \in U \cup \{0\}} \{\vec{x}_i^u\} \right) \cup \left(\bigcup_{a \in D \cup \{0\}} \{\vec{x}_a^d\} \right) \cup \left(\bigcup_{i \in U \cup \{0\}} \bigcup_{a \in D \cup \{0\}} \{\vec{x}_{(a, i)}^{u, d}\} \right) \cup \left(\bigcup_{i \in U \cup \{0\}} \bigcup_{C^u \in \mathcal{P}(D \cup \{0\})} \{\vec{x}_{(i, C^u)}^{\text{group}}\} \right),$$

where $\mathcal{P}(D \cup \{0\})$ is the power set of downstream firms. This set notation will allow me to hold one element of X constant while varying other elements, which will be important in some proofs. All elements of X may not be present in many applications. For example, an application may lack match-specific characteristics. In that case, just treat the corresponding terms as the empty set, \emptyset , in the definition of X .⁶

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

At a matching market outcome (A, T) , the total profits of $i \in U$ are given by (6). The fact that transfers enter

³Quotas are not necessary; some other explanation such as decreasing returns to scale may explain why all matches do not occur.

⁴I use the vector notation \vec{x} for characteristics only. Later I will refer to the individual, scalar elements of \vec{x} as, say, x_k .

⁵The concatenation operator makes one long vector out of a set of shorter vectors. I use the concatenation operator because discussing the properties of a production function using familiar ideas from the econometrics literature will be easier if a production function takes a single vector of arguments, rather than a number of distinct vectors, as arguments.

⁶This definition of X does not require knowledge of quotas Q , which will later be said to be unmeasured. However, if quotas are known, the researcher can disregard including in X any $\vec{x}_{(i, C^u)}^{\text{group}}$ for a $|C^u| > q_i^u$. The definition of X requires dummy arguments for characteristics involving the partner of being unmatched, 0, for notational conciseness later.

additively separably for both upstream and downstream firms allows us to focus on the following production function.

Definition 3.1. The production function for (i, C^u) for $i \in U$ and $C^u \subseteq D$ is

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

For the Section 2 example of one-to-one matching with fixed types, $f(\bar{x}(i, \{a\})) \equiv r^u(\text{cat}(\bar{x}_i^u, \bar{x}_a^d)) + r^d(\text{cat}(\bar{x}_i^u, \bar{x}_a^d))$.

I assume that the maximum quota for all upstream firms, $\max_{i \in U} q_i^u$, is known and finite. This means that $\bar{x}(i, C^u)$ has a known maximum number of elements. For many-to-many matching, a maximum quota and the additive separability of $r^d(\bar{x}(i, \{a\}))$ across multiple upstream-firm partners makes the set of arguments of f finite.⁷ Additive separability for one side of the market is restrictive. Unfortunately, I know of no other way to define a production function without relying on parametric assumptions. In an empirical application, a researcher might be willing to make parametric assumptions and choose a functional form for f so that nonlinearities in an upstream firm's profits across its downstream-firm partners are distinguished from a downstream firm's nonlinearities across its upstream-firm partners.

Sometimes I will view $f(\cdot)$ as an abstract function to be identified and estimated. In this case, I write $f(\bar{x})$, where the argument \bar{x} is an arbitrary vector of characteristics. When upstream firm i does not use all of its quota, null arguments can be included in the argument vector \bar{x} of $f(\bar{x})$ to refer to the unfilled match slots.

3.2 Pairwise stability

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

Definition 3.2. A feasible outcome (A, T) is a **pairwise stable equilibrium** when:

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

$$r^u(\bar{x}(i, C_i^u(A))) + \sum_{c \in C_i^u(A) \setminus \{a\}} t_{\langle c, i \rangle} + t_{\langle a, i \rangle} \geq r^u(\bar{x}(i, (C_i^u(A) \setminus \{a\}) \cup \{b\})) + \sum_{c \in C_i^u(A) \setminus \{a\}} t_{\langle c, i \rangle} + \tilde{t}_{\langle b, i \rangle}, \quad (7)$$

where $\tilde{t}_{\langle b, i \rangle} \equiv r^d(\bar{x}(i, \{b\})) - (r^d(\bar{x}(j, \{b\})) - t_{\langle b, j \rangle})$.

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

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

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

$$r^d(\bar{x}(i, \{a\})) - t_{\langle a, i \rangle} \geq 0.$$

⁷Consider an example with matches $\langle a, i \rangle$, $\langle b, i \rangle$ and $\langle b, j \rangle$. If the model allowed arbitrary nonlinearities in both upstream and downstream firms' structural revenue functions, there would be a set of firms $\{a, b, i, j\}$ with production $f(\bar{x}(\langle \{i, j\}, \{a, b\} \rangle))$, even though a and b , a and j as well as i and j have no direct links.

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

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

and

$$r^d(\bar{x}(i, \{a\})) - \tilde{t}_{\langle a, i \rangle} \geq 0.$$

Part 1 of the definition of pairwise stability says that upstream firm i prefers its current downstream firm a instead of some alternative downstream firm b at the transfer $\tilde{t}_{\langle b, i \rangle}$ that makes downstream firm b switch to sourcing its supplies from i instead of its equilibrium upstream firm, j . Because of transferable utility, upstream firm 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.

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.⁸ The counterexamples mean that general existence theorems do not exist.⁹

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.¹⁰ I discuss multiple equilibrium assignments below.

3.3 Using matches only: local production maximization

A matching-game outcome (A, T) has two components: the assignment, sorting or matching A and the equilibrium transfers T . I consider using data on only A . This is because researchers often lack data on transfers, even when the agents use transfers. Upstream and downstream firms 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 . For upstream firms i and j , consider an example where $C_i^u(A) = \{a\}$ and $C_j^u(A) = \{b\}$. The inequality (7)

⁸Pycia (2007) has both existence and nonexistence results for matching markets without endogenous prices (Gale and Shapley, 1962).

⁹The possibility 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.

¹⁰In the non-nested-with-matching literature on estimating normal-form Nash games, Ciliberto and Tamer (2008) throw out a particular realization of the error term's contribution to the likelihood if no pure-strategy equilibrium exists. Bajari, Hong and Ryan (2008) 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.

becomes

$$r^u(\bar{x}(i, \{a\})) + t_{\langle a, i \rangle} \geq r^u(\bar{x}(i, \{b\})) + r^d(\bar{x}(i, \{b\})) - \left(r^d(\bar{x}(j, \{b\})) - t_{\langle b, j \rangle} \right), \quad (8)$$

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

$$r^u(\bar{x}(j, \{b\})) + t_{\langle b, j \rangle} \geq r^u(\bar{x}(j, \{a\})) + r^d(\bar{x}(j, \{a\})) - \left(r^d(\bar{x}(i, \{a\})) - t_{\langle a, i \rangle} \right). \quad (9)$$

Adding (8) and (9), cancelling the transfers $t_{\langle a, i \rangle}$ and $t_{\langle b, j \rangle}$ that now are the same on both sides of the inequality, and substituting the definition of a production function, Definition 3.1, creates the new inequality

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

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

The local-production-maximization inequality suggests that interactions between the characteristics of agents in production functions drive the equilibrium pattern of sorting in a market. As the same set of firms 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 $f(\bar{x}(i, \{a\})) = \beta'_u \bar{x}_i^u + \beta'_d \bar{x}_a^d$, then a local production maximization inequality is

$$\beta'_u \bar{x}_i^u + \beta'_d \bar{x}_a^d + \beta'_u \bar{x}_j^u + \beta'_d \bar{x}_b^d \geq \beta'_u \bar{x}_i^u + \beta'_d \bar{x}_b^d + \beta'_u \bar{x}_j^u + \beta'_d \bar{x}_a^d, \quad (10)$$

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.¹¹

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

Lemma 3.1. Given a pairwise stable outcome (A, T) , let $B_1 \subseteq A$, let π be a permutation of the downstream-firm partners in B_1 , and let

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

Then the inequality

$$\sum_{\langle a, i \rangle \in B_1} f(\bar{x}(i, C_i^u(A))) \geq \sum_{\langle a, i \rangle \in B_2} f(\bar{x}(i, C_i^u((A \setminus B_1) \cup B_2))) \quad (11)$$

holds.

All proofs are found in the appendix.¹² The definition of a local-production-maximization inequality is similar

¹¹For 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.

¹²A permutation π of the downstream firm partners applied to a set of matches $\{ \langle a, i \rangle, \langle b, j \rangle, \langle c, k \rangle \}$ gives each upstream firm a new downstream-firm partner. An example of a permutation is $\{ \langle c, i \rangle, \langle a, j \rangle, \langle b, k \rangle \}$. For simplicity of notation, I let $\pi \langle a, i \rangle = c$ give the index of the new downstream firm partner c of the upstream firm i .

to (11), except that no particular outcome (A, T) needs to be stated. This definition will be used formally in the identification proofs.

Definition 3.3. Let there be a set of matches B_1 and let B_2 be a permutation π of B_1 , $B_2 = \{\langle \pi \langle a, i \rangle, i \rangle \mid \langle a, i \rangle \in B_1\}$. For each i where $\langle a, i \rangle \in B_1$, let there be a set of downstream firms C_i^u such that $\langle a, i \rangle \in B_1$ implies $a \in C_i^u$. Call

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

a local-production-maximization inequality.¹³

The definition of pairwise stability is powerful: the condition that no upstream firm wants to swap a single downstream-firm partner for a single new partner at the equilibrium transfers implies local-production-maximization inequalities involving large sets of matches B_1 and B_2 . The potential large size of B_1 and B_2 in the lemma will be important for some of the nonparametric identification theorems below.¹⁴

4 Adding econometric error terms

4.1 Data on many independent matching markets

I will consider identification using data on the population of different matching markets. As before, a matching market is described by (D, U, Q, X, A, T) . Data on the transfers T are often not available. Similarly, quotas, Q , are often an abstraction of the matching model and are usually not found in datasets on upstream and downstream firms. Therefore, I will explore identification using data on (D, U, X, A) . From now on, I subsume D and U into X in order to use more concise notation. The researcher then observes (A, X) across markets.¹⁵

With data on the population of statistically independent and identically distributed as well as economically unrelated matching markets, the researcher is able to identify $\Pr(A \mid X)$, the probability of observing assignment A given that the market has characteristics X , as defined previously. To ensure that the model gives full support to the data, I wish that $\Pr(A \mid X) > 0$ for any physically feasible (matches of each agent under that agent's quota) assignment A .¹⁶ The probability $\Pr(A \mid X)$ will be induced by a stochastic structure S . In a model with match- $\langle a, i \rangle$ -specific error terms, $S \in \mathcal{S}$ will represent the distribution of the error terms. Then

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

where $\Pr(A \mid X, Q; f^0, S^0)$ is the probability of an assignment A being observed given the exogenous character-

¹³The notation B_2 is not strictly speaking needed for Definition 3.3. Later I use B_1 and B_2 when showing that an inequality satisfies Definition 3.3.

¹⁴I have no proof that satisfying (11) for all pairs (B_1, B_2) is a sufficient (as opposed to necessary) condition for A to be part of a pairwise stable equilibrium (A, T) in many-to-many matching games. Sotomayor (1999) implies that in a game where each firm's payoffs are additively separable across multiple matches, $f(\bar{x}(i, \{a, b\})) = f(\bar{x}(i, \{a\})) + f(\bar{x}(i, \{b\}))$, then checking all sets with two matches and their permutations, such as $B_1 = \{\langle a, i \rangle, \langle c, j \rangle\}$ and $B_2 = \{\langle c, i \rangle, \langle a, j \rangle\}$, should be enough. Additively separability across multiple matches rules out many interesting empirical applications.

¹⁵For the sake of brevity, I assume the researcher has data on all elements of X . By adding additional notation, one could extend the nonparametric-identification results to the case where some elements of X are missing. For example, all of the agents in the market may not be observed. See Fox (2007) for a related discussion on estimating the single-agent multinomial-choice model without data on all available choices.

¹⁶This focus on allowing errors to affect the realization of A distinguishes this paper's approach to matching games from the work on estimating Nash games by Pakes, Porter, Ho and Ishii (2006), which does not allow for these errors in general normal-form Nash games and so, except in a few cases such as ordered choice, does imply the analog to $\Pr(A \mid X) = 0$ for some physically possible A 's.

istics X , the exogenous quotas Q , the true match production function f^0 , and the true distribution of the error terms S^0 . The functions f^0 and S^0 are unknown to the econometrician and are arguments to the endogenous-variable data generating process $\Pr(A | X, Q; f^0, S^0)$, but they are fixed across markets and are not random variables. The matching model and any equilibrium-assignment selection rule together induce the distribution $\Pr(A | X, Q; f^0, S^0)$. I will discuss primitive formulations of error terms in detail below. The quotas in Q are unmeasured, so the econometrician observes data on $\Pr(A | X; f^0, S^0) \equiv E_{Q|X} [\Pr(A | X, Q; f^0, S^0)]$, where the expectation over Q is taken with respect to its distribution conditional on X .¹⁷

4.2 The rank order property

I will rely on a rank order property to add econometric randomness to the matching outcomes.¹⁸ I first describe a non-primitive rank order property for matching games. In Section 2, this was related to production maximization in the entire economy. The general model allows many-to-many matching, where pairwise stability does not give a link to economy-wide production, efficiency. The rank order property is stated as an assumption and can be seen as a stochastic version of local production maximization.

Assumption 4.1. Let A_1 be a feasible assignment for a market with characteristics X . Let $B_1 \subseteq A_1$ and let π be a permutation of the downstream firm partners in B_1 , giving $B_2 = \{\langle \pi(a, i), i \rangle | \langle a, i \rangle \in B_1\}$. Let $A_2 = (A_1 \setminus B_1) \cup B_2$. Let $S \in \mathcal{S}$ be any distribution of the error terms and let $f \in \mathcal{F}$ be any production function. Assume that

$$\sum_{\langle a, i \rangle \in B_1} f(\bar{x}(i, C_i^u(A_1))) > \sum_{\langle a, i \rangle \in B_2} f(\bar{x}(i, C_i^u(A_2))) \quad (12)$$

if and only if

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

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

To understand the rank order property, consider a situation where A_1 contains thousands of matches and $B_1 = \{\langle a, i \rangle, \langle b, j \rangle\}$ contains only two matches. Then $A_2 = (A_1 \setminus B_1) \cup B_2$ is equal to A_1 except that the matches $B_2 = \{\langle b, i \rangle, \langle a, j \rangle\}$ form. Given X and Q , neither A_1 or A_2 may be a stable assignment to the matching model without error terms. But A_1 might dominate A_2 in the deterministic model in that at least two agents in B_2 would prefer to match with each other instead of their assigned partners, leading to A_1 . More generally, if the local-production-maximization inequality (12) is satisfied, then some agents in B_1 want to deviate in the deterministic matching model. In a model with error terms, both A_1 and A_2 could be pairwise stable assignments to some realizations of the unobserved components in the matching model. The assumption says that A_1 will be more likely to be a pairwise stable assignment to some realized model than A_2 .

As the quotas in Q are not observed in many empirical applications, a slightly more primitive version of Assumption 4.1 is that (12) holds if and only if $\Pr(A_1 | X, Q; f, S) > \Pr(A_2 | X, Q; f, S)$, for any valid Q . Then taking expectations with respect to $Q | X$ gives Assumption 4.1. Even if Q is unobserved, for the most part I

¹⁷The transfers T do not need to be integrated out because T is a separate endogenous outcome from A .

¹⁸Maximum score is a partial-identification approach. In single-agent discrete-choice problems with the payoff structure $x'_{a,i}\beta + \varepsilon_{a,i}$ for agent i and choice a , S represents the distribution of error terms $\varepsilon_{a,i}$. S is not identified under the standard (rank order property) conditions for the identification of β in either binary- or multinomial-choice problems (Manski, 1975, 1988). In matching, I will focus on identifying f and not S .

have only considered inequalities where the total number of matches of each agent in A_1 and A_2 is kept the same.¹⁹ If unmatched agents are not considered in B_1 and B_2 and if A_1 is a feasible assignment for Q , A_2 is also a feasible assignment for that Q .

I feel the downside of the non-primitive nature of Assumption 4.1 is outweighed by the benefits it gives. The assumption allows me to tap into the results on the nonparametric identification of orderings of utility functions in Matzkin (1991, 1993). Rather than developing a completely new framework, Assumption 4.1 allows me to build on existing results on single-agent choice in order to focus on the novel complications from working with equilibrium data on matches. This is the first paper to study nonparametric identification in any type of matching game. Also and as mentioned in the introduction, this assumption leads to a maximum score estimator that resolves two curses of dimensionality (one computational, one data) in the number of agents in a matching market.

4.3 A sufficient condition for the rank order property

The remainder of Section 4 discusses the applicability of the rank order property, Assumption 4.1. It is logically coherent to skip the remainder of this section and to continue reading with the identification results in Section 5.

This subsection explores a sufficient condition for the rank order property in the context of models where assignments have unobserved components in production. In this subsection, I assume that the outcome (A, T) is in the core. The core is an equilibrium concept. A core outcome is robust to deviations by any group of firms. If the group of all firms cannot improve its joint payoff, a core assignment must maximize the sum of production for the entire matching economy. Consequently, the decentralized matching-market assignment can be restated as a social planning problem.²⁰

There is a finite, although potentially large, number of assignments. The social planning problem is a single-agent, unordered, discrete-choice problem of Manski (1975), where the single agent is the social planner. From Manski's work, we know the sufficient condition that will arise. For an assignment A , let its total production be $\sum_{(a,i) \in A} f(\bar{x}(i, C_i^u(A))) + \psi_A$, where ψ_A is an unobserved component of the production of assignment A . Let ψ be the vector of all ψ_A 's. Let ψ have the density S and let ψ be independent of Q and X .²¹ Let $\Pr(A | Q, X; f, S)$ be the probability A is the core assignment.

Lemma 4.1. Let the equilibrium concept be the core and let the density S exist, have full support and be exchangeable in the elements of ψ . Then the rank order property, Assumption 4.1, holds.

This lemma was proved in Goeree, Holt and Palfrey (2005) and is a slight generalization of a result in Manski (1975).²²

¹⁹Unmatched agents 0 could be included in matches in B_1 . For nonparametric identification, data on unmatched agents will not be needed, except for one theorem.

²⁰The necessary conditions from pairwise stability are enough to nonparametrically identify production functions. Extra inequalities from the stronger equilibrium concept the core will not be required. However, the social planning property ensures a unique assignment with probability 1. If the total payoff to a match, say $f(\bar{x}(i, C_i^u)) + \sum_{a \in C^u} \epsilon_{(a,i)}$, has a continuous distribution because of continuous elements of $\bar{x}(i, C_i^u)$ or $\epsilon_{(a,i)}$, then the event that two or more assignments solve the social planner's problem has probability zero. Here the probability is taken across markets. To eliminate the role of the equilibrium-assignment selection rule, this section considers only games where the outcome is in the core.

²¹More generally, there is no need for the density S to be the same for all markets X . See Fox (2007) for more discussion of letting X be a conditioning argument in S , for the single-agent, multinomial-choice case.

²²An exchangeable joint density satisfies $g(y_1, y_2, \dots, y_n) = g(\pi y_1, \pi y_2, \dots, \pi y_n)$ for any permutation π of any vector of arguments (y_1, \dots, y_n) .

The social-planner errors can be interpreted as errors in the deterministic model from finding the true core solution. One could then view exchangeability of the joint density as a structural assumption on the equilibrium-assignment selection process. Adding errors to a deterministic model is similar to the quantile-response-equilibrium method of perturbing behavior (Goeree et al.). The social planning problem is a structural assumption that does exactly generalize the intuition from the empirical matching literature (without error terms) that assignments with, say, more assortative matching are more likely to occur.²³

4.4 Match-specific error terms

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

The analog of the practice for perfect-information Nash games in matching games (a non-nested class of games) is match-specific error terms $\varepsilon_{(a,i)}$.²⁴ Let the total output of a set of downstream firm partners C^u for upstream firm i be

$$f(\vec{x}(i, C^u)) + \sum_{a \in C^u} \varepsilon_{(a,i)}, \quad (13)$$

where $\varepsilon_{(a,i)}$ is the match- $\varepsilon_{(a,i)}$ -specific error term, which is independent of all components of X and Q . Let the stochastic structure S represent the distribution of $\varepsilon_{(a,i)}$. Let the game's outcome be in the core, again to eliminate the need to discuss equilibrium-assignment selection rules. Then, in the perfect information world where $\varepsilon_{(a,i)}$ is observed by all agents in the model but is not in the data,

$$\Pr(A | X, Q; f, S) = \int_{\varepsilon} 1[A \text{ maximizes output} | X, Q, \varepsilon] dS(\varepsilon), \quad (14)$$

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

Unfortunately, the integrand in (14) has errors at the match and not assignment level, and the resulting $\Pr(A | X, Q; f, S)$ does not have the nice properties for the single-agent multinomial choice model that Manski

²³An exchangeable joint density for assignment-level errors is a sufficient but not necessary condition for the rank order property. Consider the comparison of an assignment $A_1 = \{\langle 1, 1 \rangle, \langle 2, 2 \rangle, \langle 3, 0 \rangle, \langle 0, 3 \rangle\}$, where downstream and upstream firms 3 are both unmatched, to another assignment $A_2 = \{\langle 1, 2 \rangle, \langle 2, 3 \rangle, \langle 3, 1 \rangle\}$ where all firms are matched. The local-production-maximization inequality in Assumption 4.1 does not allow comparing A_1 and A_2 because A_2 is not a rearrangement of downstream-firm partners from A_1 : A_2 does not reallocate the former states of being unmatched. Therefore, the comparison of A_1 and A_2 is not relevant for the rank order property. However, the model of a social planner with exchangeable errors makes predictions about the relative frequencies of A_1 and A_2 based on their relative sums of production. For example, Lemma 4.1 says $\sum_{(a,i) \in A_1} f(\vec{x}(i, C_i^u(A))) = \sum_{(a,i) \in A_2} f(\vec{x}(i, C_i^u(A)))$ if and only if $\Pr(A_1 | X; f, S) = \Pr(A_2 | X; f, S)$. This is despite the fact that in A_1 match $\langle 3, 3 \rangle$ will often form if it gives more production than being unmatched, so, in many realizations of uncertainty, A_1 will not be a stable assignment. Therefore, the rank order property is a weaker assumption than a social planner with exchangeable assignment-level errors.

²⁴Pairwise stability is equivalent to the core for one-to-one matching games and many-to-many matching games where payoffs are additively separable across multiple matches. The core is a stronger solution concept for many-to-many matching games where payoffs are not additively separable across multiple matches.

(1975) found when, say, $\varepsilon_{a,i}$ was i.i.d. across choices for the same agent. For matching, unlike single-agent discrete choice, it is not a theorem that i.i.d. errors yield the rank order property for matching, Assumption 4.1. If the rank order property for matching is a natural, stochastic generalization of the core property found by Becker (1973) and others, then i.i.d. errors is not a primitive condition for the stochastic structure \mathcal{S} that gives this natural generalization, exactly.

All models are approximations to reality. If the true production function is thought to include i.i.d. match-specific shocks as in (13) and therefore assignment probabilities are given by (14), then the rank order property may actually be a pretty close approximation. After all, the transferable utility and price-taking structure of the game does naturally imply that adding production functions is much more natural than in a noncooperative Nash game. I now present simulation results that examine how closely a perfect-information matching game with shocks as in (13) is approximated by the rank order property, a natural generalization of prior work on matching games without econometric errors.

Table 1 includes results from simulations that compute assignment probabilities for a one-to-one, two-sided matching game where match production is $f(\bar{x}(i, \{a\})) + \varepsilon_{(a,i)}$.²⁵ The matching game has three upstream firms and three downstream firms. The deterministic payoffs of the game are chosen so that $\sum_{(a,i) \in A_1} f(\bar{x}(i, \{a\})) = \sum_{(a,i) \in A_2} f(\bar{x}(i, \{a\}))$ for two assignments, A_1 and A_2 . As A_1 and A_2 differ by rearrangements of one downstream firm per upstream firm, the rank order property, Assumption 4.1, requires that $\Pr(A_1 | X; f, S) = \Pr(A_2 | X; f, S)$. Neither A_1 or A_2 is a deterministic stable assignment.²⁶ The deterministic payoffs $\sum_{(a,i) \in A_1} f(\bar{x}(i, \{a\}))$ are constructed so that deviation by agents in A_2 is more attractive in an ease metric (two matched pairs could exchange partners, leaving the third pair alone) to provide a more compelling test against the idea that the rank order property is satisfied. The details of the game are in the notes to Table 1.

Table 1 considers six distributions S for i.i.d. match-specific unobservables. The table uses a simulation of the integral in (14) to compute $\Pr(A_1 | X; f, S) - \Pr(A_2 | X; f, S)$, a measure of how far off the rank order property is. The first line considers a standard normal distribution. As the variance is small and both A_1 and A_2 are not stable assignments in the deterministic game, the assignment probabilities are individually small. However, the difference $\Pr(A_1 | X; f, S) - \Pr(A_2 | X; f, S) = -0.01514$ is large relative to the magnitudes. The second line increases the normal standard deviation to 6. Both assignment probabilities increase to around 0.079, but the difference $\Pr(A_1 | X; f, S) - \Pr(A_2 | X; f, S)$ decreases in absolute value to 0.00009. The third line increases the standard deviation to 20; now the probabilities are around 0.065, although $\Pr(A_1 | X; f, S) - \Pr(A_2 | X; f, S)$ remains small, at 0.00011.

I also investigate to what degree the previous simulations relied on normality. The final three experiments in Table 1 consider asymmetric, mixed normal distributions with two modes. Again, it appears that the absolute value of $\Pr(A_1 | X; f, S) - \Pr(A_2 | X; f, S)$ is smaller when the standard deviation of the errors is higher.

Table 1 implies that assignment probabilities that differ by exchanges of only one downstream firm for each upstream firm are nearly rank ordered by their deterministic payoffs, when true payoffs include match-specific stochastic components. In Fox (2009), I introduce a maximum-score estimator using this paper's identification results and present a Monte Carlo study that shows the estimator has good finite-sample performance when the data are generated by matching games with match-specific unobservables.

²⁵The game is one-to-one matching, so Q is observed and the same for all markets.

²⁶In other words, neither A_1 or A_2 would solve the social planner's assignment problem if all errors $\varepsilon_{(a,i)}$ were 0.

4.5 Multiple equilibrium assignments

Assumption 4.1 may hold without equilibrium-assignment selection rules if the researcher is willing to specify that the outcome to the game is in the core. As argued earlier, the equilibrium assignment will be generically unique.

Return to using pairwise stability as the only equilibrium concept. In a one-to-one matching game or a many-to-many matching game with additive separability across multiple matches (Sotomayor, 1999), generically there will be only one equilibrium assignment. In more complex, many-to-many matching games, multiple assignments may be pairwise stable. In a game with multiple equilibrium assignments, (14) becomes

$$\Pr(A | X, Q; f, S) = \int_{\varepsilon} 1[A \text{ selected assignment} | A \text{ stable}, X, Q, \varepsilon] \cdot 1[A \text{ pairwise stable} | X, Q, \varepsilon] dS(\varepsilon). \quad (15)$$

Let $\Upsilon(A | X, Q; f, S)$ be equal to $\int_{\varepsilon} 1[A \text{ pairwise stable} | X, Q, \varepsilon] dS(\varepsilon)$. Define A_1 and A_2 as in Assumption 4.1. For a model with multiple equilibrium assignments, the rank order property, Assumption 4.1, will hold under the following conditions: 1) $\Upsilon(A_1 | X, Q; f, S) > \Upsilon(A_2 | X, Q; f, S)$ if and only if inequality (11) holds, and 2) $\Pr(A_1 | X, Q; f, S) > \Pr(A_2 | X, Q; f, S)$ if and only if $\Upsilon(A_1 | X, Q; f, S) > \Upsilon(A_2 | X, Q; f, S)$. Part 1 says A_1 will more likely to be stable than A_2 if A_1 has a higher production after an exchange of one downstream firm for each upstream firm in some set $B_1 \subseteq A_1$. Part 2 says the equilibrium-assignment selection rule preserves the rank ordering of stability: assignments that are more likely to be stable are more likely to occur. These conditions together imply Assumption 4.1 and hence allow a unified framework to be used to study identification and estimation of matching games, regardless of the number of stable assignments for each ε and X combination.

The above assumptions are strong. In the literature on estimating Nash games, a non-nested class with matching games, some researchers assume a particular selection rule when such a rule is easy to implement (Jia, 2008) and others use a numerical procedure and report the first equilibrium the routine converges to (Seim, 2006). Assumptions about an equilibrium assignment selection rule may be just as arbitrary as the above approaches, but for now they are currently the only feasible alternative for matching games with large numbers of agents and multiple equilibrium assignments.²⁷

5 Derivative-based nonparametric identification

No previous paper has studied the nonparametric identification of matching games. In this section, I generalize the identification results in Becker (1973) to the case of many-to-many matching with vectors of agent characteristics, among other extensions. The intuition for some of these results has been given in Section 2 for the case of one-to-one matching. Here, I eschew intuition and focus on stating general theorems precisely.

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

²⁷The literature on estimating parametric Nash games, a non-nested class with matching games, presents some more rigorous, but computationally demanding, strategies for dealing with multiple equilibria. See Bajari, Hong and Ryan (2008) and Ciliberto and Tamer (2008). I do not believe the nonparametric identification of other model components has been studied while simultaneously employing these methods.

5.1 Derivative-based identification preliminaries

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

Definition 5.1. Let \mathcal{F} be a class of production functions. Let $f^0 \in \mathcal{F}$ be the production function and let $S^0 \in \mathcal{S}$ be the stochastic structure in the data generating process.

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

The probability of a set \mathcal{Y} is $\int_{\mathcal{Y}} dG(X)$. I maintain the following assumption for derivative-based identification.

Assumption 5.1.

1. Each $f \in \mathcal{F}$ is three-times differentiable in all of its arguments.
2. X has support equal to the product of the marginal supports of the scalar elements of the vectors that comprise X . Each scalar element has continuous support on an open rectangle of \mathbb{R} .

I make this assumption to focus on cross-partial derivatives, for example. These conditions can be relaxed.²⁸

The features of the production functions that govern sorting depend on how the characteristics that enter $\vec{x}(i, C^u)$ vary. I will present results where characteristics vary at the levels of the firm i , the individual match $\langle a, i \rangle$, and the group C^u of downstream firms matching with one upstream firm. Keep in mind that a unit of observation is a market. I use variation in market-level observables X for identification.

5.2 Derivative-based identification theorems

5.2.1 Derivative-based identification with firm-specific characteristics

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

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

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

Theorem 5.1. Let $\vec{x}(i, C^u) = \text{cat} \left(\left(x_{i,1}^u, \dots, x_{i,K^u}^u \right), \left(x_{a_1,1}^d, \dots, x_{a_1,K^d}^d \right), \dots, \left(x_{a_n,1}^d, \dots, x_{a_n,K^d}^d \right) \right)$. Let \vec{x} be a given point of evaluation of f .

²⁸While there are definitions such as increasing differences (Milgrom and Shannon, 1994) that encompass complementarities without relying on differentiable f 's and continuous support for the x 's, working with broader definitions makes the results harder to interpret and to compare to Becker's.

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

The theorem is stated for a given point of evaluation \bar{x} for clarity. As the theorem holds for all points of evaluation with nonzero cross-partial derivatives, the theorem establishes the global identification of the listed properties. The intuition behind Part 1 was presented in Section 2.4.1 and Section 2.4.2 presented the intuition for Part 2.

Part 1 shows that Becker's result for the case of scalar types for men and women applies to each pair of scalar characteristics for distinct firms. Part 1 also extends Becker to a local notion of identification: f 's inputs can be complements at some areas of support and substitutes at other areas.

Becker studies only scalar types. Part 2 of the theorem shows the econometrician can go further and identify the relative importance of the complementarities for two pairs of characteristics. This is perhaps the most important result on identification in this paper. The econometrician can run a "horse race" where he or she tries to measure the relative importance of several pairs of characteristics. This is not just an ordering of magnitudes, as in $\frac{\partial^2 f^0(\bar{x})}{\partial x_1 \partial x_2} > \frac{\partial^2 f^0(\bar{x})}{\partial x_3 \partial x_4}$. Rather, the ratio of the degree of complementarities, $\frac{\partial^2 f^0(\bar{x})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{x})}{\partial x_3 \partial x_4}$, can be identified from qualitative data on who matches with whom.

5.2.2 Derivative-based identification with match-specific characteristics

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

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

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

Theorem 5.2. *Let the two scalars x_1 and x_2 be distinct elements of $\bar{x}(i, C^u) = \text{cat} \left(\left(x_{\langle a_1, i \rangle, 1}^{u,d}, \dots, x_{\langle a_1, i \rangle, K}^{u,d} \right), \dots, \left(x_{\langle a_n, i \rangle, 1}^{u,d}, \dots, x_{\langle a_n, i \rangle, K}^{u,d} \right) \right)$, corresponding to different matches. Let \bar{x} be a given point of evaluation of f .*

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

Like with firm-specific characteristics, a researcher can measure the relative importance of sorting on various characteristics in the production function, $\frac{\partial^2 f^0(\bar{x})}{\partial^2 x_1} / \frac{\partial^2 f^0(\bar{x})}{\partial^2 x_2}$.²⁹

²⁹Some might think firm-specific characteristics are a special case of match-specific characteristics. Firm-specific characteristics increase the difficulty of showing the identification of features of f^0 because the same firm characteristics must appear on the left and the right sides of a local-production-maximization inequality. By contrast, hypothetical markets exist where the match-specific characteristics for the matches $\langle a, i \rangle$ and $\langle a, j \rangle$ may take on any pair values, under Assumption 5.1.

Note that the proofs of the identification theorems are more general than the statements. The proofs do not require strong properties on the characteristics not given special attention in the statement of the theorems. For example, x_1 and x_2 could be match-specific, allowing Theorem 5.2 to be applied, while x_3 – x_6 could be firm specific, requiring Theorem 5.1. The presence of the match-specific x_1 and x_2 does not invalidate applying Theorem 5.1 to x_3 – x_6 .

5.2.3 Derivative-based identification with group-specific characteristics

Characteristics can be specific to a group of downstream firms and an upstream firm that the group matches with. Let $\vec{x}(i, C^u) = \vec{x}_{(i, C^u)}^{\text{group}} = (x_{(i, C^u), 1}^{\text{group}}, \dots, x_{(i, C^u), K}^{\text{group}})$, where there are K scalar group- (i, C^u) -specific characteristics, with the e th such scalar characteristic being $x_{(i, C^u), e}^{\text{group}}$. An example of using the estimator in this paper for the group-characteristic case is Fox and Bajari (2009), who model bidders matching to a package of geographic licenses in a spectrum auction. A characteristic of a package of licenses is the extent of the geographic complementarities among the licenses. Fox and Bajari use a measure like the gravity equation in international trade to create a proxy for these geographic complementarities.

Theorem 5.3. *Let the two scalars x_1 and x_2 be distinct elements of $\vec{x}(i, C^u) = \vec{x}_{(i, C^u)}^{\text{group}} = (x_{(i, C^u), 1}^{\text{group}}, \dots, x_{(i, C^u), K}^{\text{group}})$, corresponding to different matches. Let \vec{x} be a given point of evaluation of f .*

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

The proof of this theorem is omitted because the mathematical argument is nearly identical to the proof of Theorem 5.2.

6 Nonparametric identification of orderings of production

Section 2.5 motivated why identifying orderings of production levels is distinct from derivative-based identification. Here I focus on precise statements of general theorems.

6.1 Preliminaries for the identification of orderings of production

Positive monotonic transformations preserve rankings, so we must rule those transformations out.

Assumption 6.1. Let \mathcal{F} be a class of production functions. For any two members of this class \mathcal{F} , f^1 and f^2 , for no positive, strictly monotonic function m is it the case that $f^1(\vec{x}) = m \circ f^2(\vec{x})$ for all \vec{x} .

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

Recall that \vec{x} is one long vector of scalar characteristics. Call the first, scalar element of this vector, x_1 . Call all other elements \vec{x}_{-1} . The collection of market characteristics is X .

Assumption 6.2.

- The conditional density of characteristic 1, $g(x_1 | \bar{x}_{-1}, X \setminus \bar{x})$, has an everywhere positive density in \mathbb{R} .³⁰
- Each $f \in \mathcal{F}$ is continuous in its argument x_1 and is either strictly increasing or strictly decreasing in x_1 .
- X has support equal to the product of the marginal supports of the scalar elements of the vectors that comprise X .
- Each scalar element of a vector in X has either strictly discrete or strictly continuous support.
- Each $f \in \mathcal{F}$ is continuous in any scalar element of \bar{x} with continuous support.

The assumption allows all but one of the characteristics in \bar{x} to have discrete or qualitative support.³¹ This assumption replaces the earlier Assumption 5.1, which is only for the derivative-based identification theorems.³²

The technical use of Assumption 6.2 involves the lack of a positive monotonic transformation and its relationship to a strict inequality. I state the argument in a separate lemma because the continuous-covariate argument is used in the same way in the proofs of the three identification theorems on the orderings of production.

Lemma 6.1. Let f^1 and f^2 be production functions in a class \mathcal{F} satisfying Assumption 6.1. If Assumption 6.2 holds, then there exists two vectors \bar{x}_a and \bar{x}_b such that either

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

or

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

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

6.2 Identifying orderings of production

Identification proofs in the single-agent maximum score tradition (Matzkin, 1993) typically amount mathematically to Lemmas 4.1 and 6.1. Consequently, the identification proof for each case focuses on an issue that is new to matching games: embedding the inequalities from Lemma 6.1 in a local-production-maximization inequality, meaning an inequality where each upstream firm switches at most one downstream firm at a time. Thus, the proofs look for market characteristics X where the comparisons in Lemma 6.1 are decisive in rank ordering the production of two larger, otherwise similar assignments, A_1 and A_2 . I consider group, match and firm-specific characteristics separately. Identification is Definition 5.1 subject to the lack of a positive monotonic transformation in Assumption 6.1. I list the theorems in increasing difficulty of the proofs.

³⁰The notation $X \setminus \bar{x}$ means all elements of X other than those in the specified vector \bar{x} . Recall \bar{x} can include firm-, match- and group-specific characteristics. This is a slight notational abuse as \bar{x} is a long vector formed from the concatenation of subvectors, rather than the set of such subvectors.

³¹The assumption that the support of x_1 is \mathbb{R} , rather than some compact subset of \mathbb{R} , is made for convenience. Manski (1988) and Horowitz (1998) show how to relax the full support assumption for the identification of single-agent binary-choice models. A continuous product quality could be a candidate for the continuous upstream product characteristic x_1 .

³²The identification arguments in this paper are not related to the identification-at-infinity arguments made in the literature on selection and the related work on the special regressor estimator of Lewbel (2000). Identification based on special regressor arguments might be possible if there are match-specific regressors with full support and independence from the error terms. Arguments exist to weaken the full support assumptions (Magnac and Maurin, 2007). Special-regressor identification arguments do not lead to tractable estimators for matching games. The Lewbel single-agent, multinomial-choice estimator requires multidimensional density estimation and therefore suffers from a data curse of dimensionality in the number of choices.

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

Theorem 6.1. *Let $\bar{x}(i, C^u) = \bar{x}_{(i, C^u)}^{\text{group}} = (x_{(i, C^u), 1}^{\text{group}}, \dots, x_{(i, C^u), K}^{\text{group}})$. Then the production function is identified in the class \mathcal{F} .*

Match-specific characteristics make the identification proof more complex than before. The reason is that the equilibrium concept of pairwise stability, Definition 3.2, involves only one unmatched pair deviating at a time. To show identification, we must start with Lemma 6.1 and be able to construct local-production-maximization inequalities where the coalition characteristics differ by the arguments corresponding to only one match between an upstream and a downstream firm. Remember the production function requires a vector of arguments for each match of an upstream firm. To apply this and the following theorems, the maximum quota of an upstream firm must be known.

Theorem 6.2. *Let $\bar{x}(i, C^u) = \text{cat} \left((x_{(a_1, i), 1}^{u, d}, \dots, x_{(a_1, i), K}^{u, d}), \dots, (x_{(a_n, i), 1}^{u, d}, \dots, x_{(a_n, i), K}^{u, d}) \right)$. Also, let there be assignments A that contain as many groups of matches as the maximum quota of an upstream firm. The production function is identified in the class \mathcal{F} .*

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

An alternative way of identifying a production function involving the characteristics of the matches involving many downstream firms may be to use a solution concept such as the core. The core solution concept would give inequalities where the researcher can have groups of downstream firms matched to the same upstream firm deviating at once. An achievement of this paper is to show identification without relying on the crutch of a stronger equilibrium concept: only pairwise stability is imposed. This is important, as the core is a very strong equilibrium concept. Core outcomes are less likely to exist and believing that a many-to-many outcome is in the core would require a lot of communication and coordination for the agents in a decentralized matching market. I show that pairwise stability, which requires only communication between one upstream and one downstream firm at a time, is enough for point identification in many-to-many matching games where production functions are not additively separable in the characteristics of the matches for multiple downstream firms.

Firm-specific characteristics require an additional normalization. As in the example in Section 2.5, the value of being unmatched will be 0. Informally, identification considers the probabilities of assignments where certain firms are unmatched.³³ Firms that are more likely to be unmatched in an assignment are likely to have lower contributions to production.

Theorem 6.3. *Let $\bar{x}(i, C^u) = \text{cat} \left((x_{i, 1}^u, \dots, x_{i, K^u}^u), (x_{a_1, 1}^d, \dots, x_{a_1, K^d}^d), \dots, (x_{a_n, 1}^d, \dots, x_{a_n, K^d}^d) \right)$. Let the value of any firm remaining unmatched be 0, or $f(\bar{x}_i^d) = f(\bar{x}_i^u) = 0$ for all $i \in U$, $a \in D$ and $f \in \mathcal{F}$. Further, let there be assignments A that contain as many matched coalitions as three times the maximum quota of an upstream firm. Then the production function is identified in the class \mathcal{F} .*

³³Choo and Siow (2006) estimate a logit-based one-to-one matching model of marriage that requires data on the fraction of each observable type of man or woman that is single.

7 Conclusions

This paper discusses identification 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 one-to-one matching games this condition is related to social efficiency, but for general many-to-many matching games it is not.

It is not obvious what types of economic parameters are identified from data on only who matches with whom. The identification theorems cover both certain derivatives of production functions and orderings of production levels. The derivative-based results generalize the work of Becker (1973) to the case of each agent having a vector of types, many-to-many matching as well as production functions where pairs of inputs are not complements over their entire supports. One can identify whether any two inputs or complements or substitutes. Importantly, one can identify the value of the ratio of complementarities for two pairs of inputs, at any point. Thus, one can run “horse race” analyses to identify the relative importance of different pairs of characteristics in match production.

The results on the ordering of production levels extend the single-agent work of Matzkin (1993) to matching games, where agents cannot unilaterally choose partners and so identification requires working with the equilibrium structure of the game. We can learn whether certain inputs are “goods” that always raise output. We can distinguish between $f^1(a_1, i_1) = -(a_1 - i_1)^2$ and $f^2(a_1, i_1) = 2a_1 \cdot i_1$.

The identification results are relevant for empirical work. In Fox and Bajari (2009), we use both of these sets of identification theorems to measure the efficiency of a FCC spectrum auction compared to some counterfactual outcomes. Without identification theorems, it would be unclear whether qualitative match data is sufficient to identify some notion of efficiency. This paper shows qualitative match data can allow such measurements in transferable utility matching games.

A Proofs

A.1 Lemma 3.1: Pairwise stability implies local production maximization

Substitute $\tilde{t}_{(b,i)}$ into (7) and cancel the transfers $\sum_{c \in C_i^u(A) \setminus \{a\}} t_{(c,i)}$ to give

$$r^u(\bar{x}(i, C_i^u(A))) + t_{(a,i)} \geq r^u(\bar{x}(i, (C_i^u(A) \setminus \{a\}) \cup \{b\})) + r^d(\bar{x}(i, \{b\})) - (r^d(\bar{x}(j, \{b\})) - t_{(b,j)}).$$

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

B_1 . Now form $\sum_{\langle a,i \rangle \in B_1} \text{nd} \left(\langle a,i \rangle, \langle \pi \langle a,i \rangle, j(\pi \langle a,i \rangle) \rangle, A, t_{\langle a,i \rangle}, t_{\langle \pi \langle a,i \rangle, j(\pi \langle a,i \rangle) \rangle} \right)$. This gives

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

By the definition of a permutation, each $t_{\langle a,i \rangle}$ for $\langle a,i \rangle \in B_1$ appears on both the left and right sides. The transfers cancel. Similarly, each equilibrium $r^d(\bar{x}(j(\pi \langle a,i \rangle), \{\pi \langle a,i \rangle\}))$ appears on the right side with a negative sign and each deviation $r^d(\bar{x}(i, \{\pi \langle a,i \rangle\}))$ appears on the right side with a positive sign. Moving $\sum_{\langle a,i \rangle \in B_1} r^d(\bar{x}(j(\pi \langle a,i \rangle), \{\pi \langle a,i \rangle\}))$ to the left side and substituting the definition of a production function, Definition 3.1, gives the local-production-maximization inequality in the lemma.

A.2 Theorem 5.1: Derivative-based identification with firm-specific covariates

A.2.1 Part 1

The vector \bar{x} is given in the statement of the theorem. To avoid confusion of the point \bar{x} with the function $\bar{x}(i, C_i^\mu)$, I relabel the vector \bar{x} as \bar{y} inside this proof. I will focus on the case where $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} > 0$. The proof for the case where $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2} < 0$ is very similar.

For an arbitrary $f^1 \in \mathcal{F}$, $f^1 \neq f^0$ where $\frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} < 0$, by Definition 5.1, the definition of identification of a feature of a function, I must show that there does not exist S^1 corresponding to f^1 where $\Pr(A | X; f^0, S^0) = \Pr(A | X; f^1, S^1)$ for all (A, X) except perhaps a set of X of probability 0. By the key Assumption 4.1, a sufficient condition involves showing that there exists a continuum of market characteristics X with positive probability and a corresponding matching situation where f^0 and f^1 give different implications for a local-production-maximization inequality, of the form in Definition 3.3. At each of these markets X , there will be a particular assignment A_1 and another assignment A_2 where, by Assumption 4.1, $\Pr(A_1 | X; f^0, S^0) > \Pr(A_2 | X; f^0, S^0)$ while $\Pr(A_1 | X; f^1, S^1) < \Pr(A_2 | X; f^1, S^1)$ for any $S_1 \in \mathcal{S}$. Therefore, the conditions of Definition 5.1 will be satisfied.

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

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

$$\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2} = \lim_{h \rightarrow 0} \frac{f(\bar{y} + he_1 + he_2) - f(\bar{y} + he_1) - f(\bar{y} + he_2) + f(\bar{y})}{h^2}. \quad (16)$$

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

$$\left| \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_2} - \frac{f(\bar{y} + h e_1 + h e_2) - f(\bar{y} + h e_1) - f(\bar{y} + h e_2) + f(\bar{y})}{h^2} \right| < v.$$

As $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} > 0$, there will be a $h^0 > 0$ such that the numerator of the middle difference quotient at $f = f^0$ is positive, or

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

or

$$f^0(\bar{y} + h^0 e_1 + h^0 e_2) + f^0(\bar{y}) > f^0(\bar{y} + h^0 e_1) + f^0(\bar{y} + h^0 e_2). \quad (17)$$

As $\frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} < 0$, there exists $h^1 > 0$ where

$$f^1(\bar{y} + h^1 e_1 + h^1 e_2) + f^1(\bar{y}) < f^1(\bar{y} + h^1 e_1) + f^1(\bar{y} + h^1 e_2). \quad (18)$$

The argument for f^1 is symmetric to the argument for f^0 and is omitted. Set $h = \min\{h^0, h^1\}$. The inequalities (17) and (18) hold for any such h .

Now let me argue that (17), and by a similar argument (18), is a local-production-maximization inequality: it satisfies Definition 3.3. To do this I need to form B_1 and B_2 , as in the definition, and show how a hypothetical swap of downstream firm partners could produce (17). Let $B_1 = \{\langle a, i \rangle, \langle b, j \rangle\}$ and $B_2 = \{\langle b, i \rangle, \langle a, j \rangle\}$, where these indices refer to arbitrary firms I am creating to show (17) satisfies Definition 3.3. Also, let there be some C_i^u and C_j^u of sufficient size to reproduce the number of non-empty (\emptyset , representing unfilled matches) elements of \bar{y} . We require $a \in C_i^u$, $a \notin C_j^u$, $b \in C_j^u$, and $b \notin C_i^u$. Then define $\bar{x}(i, C_i^u) = \bar{y} + h e_1 + h e_2$, $\bar{x}(j, C_j^u) = \bar{y}$, $\bar{x}(i, (C_i^u \setminus \{a\}) \cup \{b\}) = \bar{y} + h e_1$ and $\bar{x}(j, (C_j^u \setminus \{b\}) \cup \{a\}) = \bar{y} + h e_2$. With $\pi \langle a, i \rangle = b$ and $\pi \langle b, j \rangle = a$, inspection shows (17) satisfies Definition 3.3.

Further, it is important to show that this exchange of partners can be accomplished with firm-specific characteristics, as that is a maintained hypothesis in the theorem being proved. The theorem requires that x_1 and x_2 be from different firms. As only one upstream firm's characteristics enter each production function, it is without loss of generality to say that x_2 is a characteristic of a downstream firm. To complete the argument that (17) satisfies Definition 3.3 for the case of firm-specific characteristics, let downstream firms a and b have the same baseline characteristics, except that firm a has $h e_2$ more of characteristic x_2 than firm b . Using the notation $\bar{x}(i, C^u) = \text{cat}\left(\left(x_{i,1}^u, \dots, x_{i,K^u}^u\right), \left(x_{d_1,1}^d, \dots, x_{d_1,K^d}^d\right), \dots, \left(x_{d_n,1}^d, \dots, x_{d_n,K^d}^d\right)\right)$ for $C^u = \{d_1, \dots, d_n\}$, then $x_{a,2}^d - h e_2 = x_{b,2}^d$, where now a is one of the firms in C_i^u . On the left of (17), the match $\langle a, i \rangle$ puts the downstream firm a with $h e_2$ more x_2 in either a direct partnership with an upstream firm i with $h_1 e_1$ more x_1 than upstream firm j or an indirect partnership with another downstream firm, say $c_i \in C_i^u$, with $h_1 e_1$ more x_1 than the corresponding downstream firm c_j in C_j^u . In notation, either $x_{i,1}^u - h e_1 = x_{j,1}^u$ or $x_{c_i,1}^d - h e_1 = x_{c_j,1}^d$.³⁴

The matches $\langle b, i \rangle$ and $\langle a, j \rangle$ form on the right side of (17). The firm a with $h_2 e_2$ more of x_2 is transferred from the set of matches involving i with $h_1 e_1$ more x_1 to the set of matches involving j without any more x_1 . Likewise, the downstream firm b without any more x_2 matches to i and its downstream firm partners, which together have $h_1 e_1$ more x_1 than the matches involving j . The important requirement that is satisfied is that each

³⁴Note some abuses of notation: if x_2 is the second characteristic of $\bar{x}(i, C_i^u)$, I then say it is also the second characteristic of $\bar{x}_a^d, \bar{x}_{a,2}^d$. This is done for clarity: to keep "2" referring to the same variable whether I am referring to it as an element of the entire vector of production function arguments $\bar{x}(i, C_i^u)$ or as an element of the vector of characteristics of firm a , \bar{x}_a^d .

move switches the characteristics of only the firm that is actually switching. Therefore, (17) satisfies Definition 3.3 for the case of firm-specific characteristics.

The second step of the proof is that I will argue that I can embed B_1 and B_2 in an entire matching market. Let B_3 be a larger set of matches that includes the matches corresponding to the downstream firms in $C_i^u \setminus \{a\}$ and $C_j^u \setminus \{b\}$. The exact choice of B_3 plays no role in the proof, other than to ensure $C_i^u \setminus \{a\}$ and $C_j^u \setminus \{b\}$ are large enough given the number of non-empty elements (representing filled quota slots) in \bar{y} . Then set $A_1 = B_1 \cup B_3$ and $A_2 = B_2 \cup B_3$.

Consider the set of the unique firm-specific characteristics in an inequality, here

$$\zeta = \left\{ \left(x_{i,1}^u, \dots, x_{i,K^u}^u \right) \right\} \cup \left(\bigcup_{a \in C_i^u} \left\{ \left(x_{da,1}^d, \dots, x_{da,K^d}^d \right) \right\} \right) \cup \left\{ \left(x_{j,1}^u, \dots, x_{j,K^u}^u \right) \right\} \cup \left(\bigcup_{a \in C_j^u} \left\{ \left(x_{da,1}^d, \dots, x_{da,K^d}^d \right) \right\} \right).$$

Let there be some collection X of characteristics for all firms in the market. The choice of X plays no role in the proof, except that ζ must be a subset of X .

Assumption 4.1 and (17) imply $\Pr(A_1 | X; f^0, S^0) > \Pr(A_2 | X; f^0, S^0)$ while Assumption 4.1 and (18) imply $\Pr(A_1 | X; f^1, S^1) < \Pr(A_2 | X; f^1, S^1)$ for any $S_1 \in \mathcal{S}$.

I now move to the third stage of the proof. Definition 5.1 requires a continuum of markets with characteristics X with this property. Let v be the change of notation that makes ζ into one long vector, where all elements enter as scalars. Likewise, let \bar{z} be an arbitrary vector of length $|v|$. For \bar{z} , let $\bar{z}_1, \dots, \bar{z}_4$ be the equivalent of the four objects $\bar{x}(i, C_i^u)$, $\bar{x}(j, C_j^u)$, $\bar{x}(i, (C_i^u \setminus \{a\}) \cup \{b\})$ and $\bar{x}(j, (C_j^u \setminus \{b\}) \cup \{a\})$. Switching from \bar{z} to $\bar{z}_1, \dots, \bar{z}_4$ and back is just a change of notation as the characteristics of all firms enter the long vector \bar{z} .

\mathcal{F} contains only continuous functions, so

$$g(\bar{z} | f) \equiv f(\bar{z}_1) + f(\bar{z}_2) - f(\bar{z}_3) - f(\bar{z}_4)$$

is also a continuous function in \bar{z} , for any $f \in \mathcal{F}$. The definition of a continuous function $g(z)$ states that $g^{-1}(V)$ is an open neighborhood of v whenever V is an open neighborhood of $g(v)$. Because the inequalities are strict, there is an open neighborhood V^0 around $g(v; f^0)$ where $p > 0$ for $p \in V^0$ and there is an open neighborhood V^1 around $g(v; f^1) < 0$ where $p < 0$ for $p \in V^1$. Because the object v is the same in both the f^0 and f^1 constructions, $W = g^{-1}(V^0; f^0) \cap g^{-1}(V^1; f^1) \neq \emptyset$ and is itself an open neighborhood of v , as both $g^{-1}(V^0; f^0)$ and $g^{-1}(V^1; f^1)$ are open sets by the definition of continuity and because topologies are closed under finite intersections.

Divide any market characteristics X into X_1 and X_2 , where X_1 is the set ζ for that X , where here the firm-specific characteristics in ζ represent the characteristics of the corresponding firms given an arbitrary X (like the argument \bar{z} above) rather than terms involving the \bar{y} from the statement of the theorem. Also, $X_2 = X \setminus X_1$, so X is a one-to-one change of variables (or just notation) of (X_1, X_2) . By Assumption 5.1, X has support equal to the product of the marginal supports of the scalar elements of the vectors that comprise X . Therefore, the probability that (X_1, X_2) lies in $W \times \mathbb{R}^s$ is strictly positive, where s is the number of elements of X_2 . By construction, all market characteristics (X_1, X_2) in $W \times \mathbb{R}^s$ satisfy $\Pr(A_1 | X; f^0, S^0) > \Pr(A_2 | X; f^0, S^0)$ while $\Pr(A_1 | X; f^1, S^1) < \Pr(A_2 | X; f^1, S^1)$ for any $S_1 \in \mathcal{S}$.

A.2.2 Part 2

We are given a point \bar{y} (reabeled from \bar{x} in the statement of the theorem) and there is an arbitrary $f^1 \in \mathcal{F}$ where $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_3 \partial x_4} \neq \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\bar{y})}{\partial x_3 \partial x_4}$. The goal in broad generality is the same as Part 1: show there exists a continuum of X and two assignments A_1 and A_2 where $\Pr(A_1 | X; f^0, S^0) > \Pr(A_2 | X; f^0, S^0)$ while $\Pr(A_1 | X; f^1, S^1) < \Pr(A_2 | X; f^1, S^1)$ for any $S_1 \in \mathcal{S}$. The proof is more challenging than the proof of Part 1 because now we are trying to identify the value of some feature of f^0 , here $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_3 \partial x_4}$, rather than just the sign of some feature, as before.

I will show that the term $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3}$ is identified, where x_1 is the same characteristic in the numerator and the denominator. Then, by Young's / Clairaut's / Schwarz's theorem, arbitrary ratios $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_3 \partial x_4}$ can be identified by comparing, say, $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3}$ to $\frac{\partial^2 f^0(\bar{y})}{\partial x_3 \partial x_4} / \frac{\partial^2 f^0(\bar{y})}{\partial x_3 \partial x_1}$. Cross-partial derivatives are symmetric if the second-partial derivatives are continuous, which they are because Assumption 6.1 states f is three-times differentiable.

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

The outline of the proof follows. The most novel step comes first: I find a key inequality that arises from the numerator of the middle difference quotient, (16), and that has a different direction for f^0 and f^1 . For example, this can be seen as a situation where f^0 would predict sorting on characteristics x_1 and x_2 while f^1 would predict sorting on characteristics x_1 and x_3 when sorting on both pairs simultaneously is physically impossible. The second step is that I show that this inequality is a local-production-maximization inequality. Some final steps follow arguments in Part 1 and are omitted for brevity.

Let $h_{1,2}$ be the limit argument from the middle difference quotient, (16), for $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2}$. Likewise, let $h_{1,3}$ be the limit argument for $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_3}$. Consider the case $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3} > \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_3}$ and let $\{h_{1,2,n}\}_{n \in \mathbb{N}}$ be a sequence that converges to 0. Let $\{h_{1,3,n}\}_{n \in \mathbb{N}}$ be a sequence

$$h_{1,3,n} = h_{1,2,n} \sqrt{\frac{1}{2} \left(\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3} + \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_3} \right)}. \quad (19)$$

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

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

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

$$\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3} > \frac{h_{1,3,n}^2}{h_{1,2,n}^2} > \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_3} \quad (20)$$

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

Let $\tau = \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3} - \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_3}$ and

$$\Upsilon(h_{1,2}, h_{1,3}; f) = \frac{f(\bar{y} + h_{1,2}e_1 + h_{1,2}e_2) - f(\bar{y} + h_{1,2}e_1) - f(\bar{y} + h_{1,2}e_2) + f(\bar{y})}{h_{1,2}^2} \cdot \left(\frac{f(\bar{y} + h_{1,3}e_1 + h_{1,3}e_3) - f(\bar{y} + h_{1,3}e_1) - f(\bar{y} + h_{1,3}e_3) + f(\bar{y})}{h_{1,3}^2} \right)^{-1} \quad (21)$$

for $f \in \mathcal{F}$. By the definition of a cross-partial derivative, (16), the ratio $\Upsilon(h_{1,2,n}, h_{1,3,n}; f)$ converges to $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_3}$ for $f \in \{f^0, f^1\}$ as $n \rightarrow \infty$ and $(h_{1,2,n}, h_{1,3,n}) \rightarrow (0, 0)$. Then there exists some $n_1 \in \mathbb{N}$ where, for all $n \geq n_1$, $n \in \mathbb{N}$, $|\Upsilon(h_{1,2,n}, h_{1,3,n}; f^0) - \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3}| < \frac{\tau}{3}$ and $|\Upsilon(h_{1,2,n}, h_{1,3,n}; f^1) - \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_3}| < \frac{\tau}{3}$. The choice of distance $\frac{\tau}{3}$ ensures that

$$\Upsilon(h_{1,2,n}, h_{1,3,n}; f^0) > \frac{h_{1,3,n}^2}{h_{1,2,n}^2} > \Upsilon(h_{1,2,n}, h_{1,3,n}; f^1) \quad (22)$$

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

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

for $f \in \mathcal{F}$. Choose $(h_{1,2}, h_{1,3}) = (h_{1,2,n}, h_{1,3,n})$. Substituting the definition of $\Upsilon(h_{1,2,n}, h_{1,3,n}; f)$ into (22) and resulting algebra shows that, at $(h_{1,2}, h_{1,3})$, the ratios $h_{1,3,n}^2/h_{1,2,n}^2$ cancel in all terms and

$$\frac{f^0(\bar{y} + h_{1,2}e_1 + h_{1,2}e_2) - f^0(\bar{y} + h_{1,2}e_1) - f^0(\bar{y} + h_{1,2}e_2) + f^0(\bar{y})}{f^0(\bar{y} + h_{1,3}e_1 + h_{1,3}e_3) - f^0(\bar{y} + h_{1,3}e_1) - f^0(\bar{y} + h_{1,3}e_3) + f^0(\bar{y})} > 1 > \frac{f^1(\bar{y} + h_{1,2}e_1 + h_{1,2}e_2) - f^1(\bar{y} + h_{1,2}e_1) - f^1(\bar{y} + h_{1,2}e_2) + f^1(\bar{y})}{f^1(\bar{y} + h_{1,3}e_1 + h_{1,3}e_3) - f^1(\bar{y} + h_{1,3}e_1) - f^1(\bar{y} + h_{1,3}e_3) + f^1(\bar{y})} \quad (23)$$

and so

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

At this value $(h_{1,2}, h_{1,3})$, f^0 and f^1 have different signs for a key term $\Delta(h_{1,2}, h_{1,3}; f)$. The same style of arguments and the same choice of $h_{1,3,n}$, (19), will apply to the case $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3} < \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_3}$. Only a few inequalities are reversed.

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

$$f(\bar{y} + h_{1,2}e_1 + h_{1,2}e_2) + f(\bar{y} + h_{1,3}e_1) + f(\bar{y} + h_{1,3}e_3) + f(\bar{y}) > f(\bar{y} + h_{1,3}e_1 + h_{1,3}e_3) + f(\bar{y} + h_{1,2}e_1) + f(\bar{y} + h_{1,2}e_2) + f(\bar{y}). \quad (24)$$

Clearly this inequality is satisfied when $\Delta(h_{1,2}, h_{1,3}; f) > 0$. The inequality (24) satisfies Definition 3.3 for some choice of B_1 and B_2 . Let $B_1 = \{\langle a, i \rangle, \langle b, j \rangle, \langle c, k \rangle, \langle g, l \rangle\}$ and $B_2 = \{\langle g, i \rangle, \langle c, j \rangle, \langle b, k \rangle, \langle a, l \rangle\}$, where the permutation π is implied by the definitions of B_1 and B_2 . Also, let $a \in C_i^u$, $b \in C_j^u$, $c \in C_k^u$ and $g \in C_l^u$. Let $\bar{x}(i, C_i^u) = \bar{y} + h_{1,2}e_1 + h_{1,2}e_2$, $\bar{x}(j, C_j^u) = \bar{y} + h_{1,3}e_1$, $\bar{x}(k, C_k^u) = \bar{y} + h_{1,3}e_3$, $\bar{x}(l, C_l^u) = \bar{y}$, $\bar{x}(i, (C_i^u \setminus \{a\}) \cup \{g\}) = \bar{y} +$

$h_{1,2}e_1$, $\bar{x}(j, (C_j^u \setminus \{b\}) \cup \{c\}) = \bar{y} + h_{1,3}e_1 + h_{1,3}e_3$, $\bar{x}(k, \{C_k^u \setminus \{c\}\} \cup \{b\}) = \bar{y}$, and $\bar{x}(l, (C_l^u \setminus \{g\}) \cup \{a\}) = \bar{y} + h_{1,2}e_2$. Using the notation $\bar{x}(i, C_i^u) = \text{cat}((x_{i,1}^u, \dots, x_{i,K^u}^u), (x_{d_1,1}^d, \dots, x_{d_1,K^d}^d), \dots, (x_{d_n,1}^d, \dots, x_{d_n,K^d}^d))$: $x_{a,2}^d - h_{1,2}e_2 = x_{b,2}^d$; either $x_{i,1}^u - h_{1,2}e_1 = x_{l,1}^u$ or $x_{m_i,1}^d - h_{1,2}e_1 = x_{m_l,1}^d$ for two firms $m_i \in C_i^u$, $m_i \neq a$ and $m_l \in C_l^u$, $m_l \neq g$; $x_{c,2}^d - h_{1,3}e_3 = x_{b,2}^d$; and either $x_{j,1}^u - h_{1,3}e_1 = x_{k,1}^u$ or $x_{m_j,1}^d - h_{1,3}e_1 = x_{m_k,1}^d$ for two firms $m_j \in C_j^u$, $m_j \neq b$ and $m_k \in C_k^u$, $m_k \neq c$. By inspection, it can be seen that each match in B_1 exchanges a downstream firm partner for a match in B_2 . Meanwhile, each set of arguments $\bar{x}(i, C_i^u)$ on the right can be formed by an exchange of single downstream firm's characteristics from a set of arguments on the left. Therefore, this construction satisfies Definition 3.3 for the case of firm-specific characteristics.

As in the proof of Part 1, I can embed B_1 and B_2 into a larger matching market. Next, Assumption 4.1 states that if (24) holds for $f \in \mathcal{F}$, then $\Pr(A_1 | X; f, S) > \Pr(A_2 | X; f, S)$ for any $S \in \mathcal{S}$. Above, we found $(h_{1,2}, h_{1,3})$ where $\Delta(h_{1,2}, h_{1,3}; f^0) > 0$ and hence where (24) holds for the true f^0 . Likewise, as this point $\Delta(h_{1,2}, h_{1,3}; f^1) < 0$ and hence (24) does not hold for the alternative f^1 . Now we need to show that there exists a continuum of markets X where $\Pr(A_1 | X; f^0, S^0) > \Pr(A_2 | X; f^0, S^0)$ while $\Pr(A_1 | X; f^1, S^1) < \Pr(A_2 | X; f^1, S^1)$ for any $S^1 \in \mathcal{S}$. This part of the proof uses similar arguments as the corresponding proof of Part 1 and so is omitted for brevity.

Recall that by Part 1 of the theorem we can focus on cases where the signs of the cross partials are the same for f^0 and f^1 . Before we restricted attention to the case $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2} > 0$, $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_3} > 0$ and $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2} > \frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_3}$ for $f \in \{f^0, f^1\}$. $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2} > \frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_3}$ is without loss of generality,³⁵ but $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2} > 0$ and $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_3} > 0$ for $f \in \{f^0, f^1\}$ are conditions with some loss of generality. Now we need to argue that the above arguments go through for the other three cases: $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2} < 0$ and $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_3} > 0$; $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2} > 0$ and $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_3} < 0$; as well as $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2} < 0$ and $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_3} < 0$. This is simple: in some of these new cases key inequalities may reverse direction, but as $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3} \neq \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_3}$ for all cases, the same arguments as above will show $(h_{1,2}, h_{1,3})$ can be chosen to lie in between $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3}$ and $\frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_3}$. One this is done, the above can all be repeated without much change.³⁶

A.3 Theorem 5.2: Derivative-based identification, match-specific covariates

For conciseness, certain steps of the proof will be replaced with references to similar arguments in Theorem 5.1. Some notation, such as \bar{y} , is also explained in Theorem 5.1.

A.3.1 Part 1

One definition of a second derivative is

$$\frac{\partial^2 f(\bar{y})}{\partial^2 x_1} = \lim_{h \rightarrow 0} \frac{f(\bar{y} + 2he_1) - 2f(\bar{y} + he_1) + f(\bar{y})}{h^2}, \quad (25)$$

where as in the proof of Theorem 5.1, $e_1 = (1, 0, 0, \dots, 0)$ is a vector of 0's except in the first element. Because both $\frac{\partial^2 f^0(\bar{y})}{\partial^2 x_1}$ and $\frac{\partial^2 f^1(\bar{y})}{\partial^2 x_1}$ are limits, there will be some $h = \min\{h^0, h^1\}$, where h^0 and h^1 are the respective limit

³⁵In part, there is no loss in generality because identifying $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3}$ is equivalent to identifying its inverse, $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2}$.

³⁶If $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_3} < 0$ for $f \in \{f^0, f^1\}$, then (19) will involve the square root of a negative number. To fix this, let (19) involve the absolute values of $\frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f(\bar{y})}{\partial x_1 \partial x_3}$ for $f \in \{f^0, f^1\}$. For the case $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3} > \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_3}$, (29) will become $\frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\bar{y})}{\partial x_1 \partial x_3} > \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\bar{y})}{\partial x_1 \partial x_3}$. Following the steps of the algebra in the earlier argument, the 1 in (23) will be a -1 and the pair of inequalities in (23) will reverse directions once both sides are multiplied by the -1. A different local-production-maximization inequality will arise, but otherwise the argument is similar to the earlier argument.

arguments for f^0 and f^1 , where both

$$f^0(\bar{y} + 2he_1) + f^0(\bar{y}) > 2f^0(\bar{y} + he_1) \quad (26)$$

and

$$f^1(\bar{y} + 2he_1) + f^1(\bar{y}) < 2f^1(\bar{y} + he_1) \quad (27)$$

hold.

Now I will argue that (26) and by the same argument (27) are local-production-maximization equations, Definition 3.3. Let $B_1 = \{\langle a, i \rangle, \langle b, j \rangle\}$ and $B_2 = \{\langle b, i \rangle, \langle a, j \rangle\}$. Also let there be sufficiently large (to handle \bar{y}) sets C_i^u and C_j^u where $a \in C_i^u$, $a \notin C_j^u$, $b \in C_j^u$, and $b \notin C_i^u$. Also define $\bar{x}(i, C_i^u) = \bar{y} + 2he_1$, $\bar{x}(j, C_j^u) = \bar{y}$, $\bar{x}(i, (C_i^u \setminus \{a\}) \cup \{b\}) = \bar{y} + he_1$ and $\bar{x}(j, (C_j^u \setminus \{b\}) \cup \{a\}) = \bar{y} + he_1$. Using the match-specific characteristics notation $\bar{x}(i, C^u) = \text{cat}\left(\left(x_{\langle d_1, i \rangle, 1}^{u,d}, \dots, x_{\langle d_1, i \rangle, K}^{u,d}\right), \dots, \left(x_{\langle d_n, i \rangle, 1}^{u,d}, \dots, x_{\langle d_n, i \rangle, K}^{u,d}\right)\right)$, let $x_{\langle a, i \rangle, 1}^{u,d} - he_1 = x_{\langle a, j \rangle, 1}^{u,d} = x_{\langle b, i \rangle, 1}^{u,d}$ and $x_{\langle a, i \rangle, 1}^{u,d} - 2he_1 = x_{\langle b, j \rangle, 1}^{u,d}$. With $\pi\langle a, i \rangle = b$ and $\pi\langle b, j \rangle = a$, inspection shows (17) satisfies Definition 3.3 for the case of match-specific characteristics.

The embedding of B_1 and B_2 in a matching market and the finding of a continuum of markets X with the key property both follow similar arguments in the proof of Part 1 of Theorem 5.1.

A.3.2 Part 2

We are given a point \bar{y} (relabelled from \bar{x} in the statement of the theorem) and there is an arbitrary $f^1 \in \mathcal{F}$ where $\frac{\partial^2 f^0(\bar{y})}{\partial^2 x_1} / \frac{\partial^2 f^0(\bar{y})}{\partial^2 x_2} \neq \frac{\partial^2 f^1(\bar{y})}{\partial^2 x_1} / \frac{\partial^2 f^1(\bar{y})}{\partial^2 x_2}$. I first consider the case where $\frac{\partial^2 f(\bar{y})}{\partial^2 x_1} > 0$ and $\frac{\partial^2 f(\bar{y})}{\partial^2 x_2} > 0$ for $f \in \{f^0, f^1\}$. The other cases are discussed at the end of the proof.

Let h_1 be the index for the approximation term on the right side of (25) for $\frac{\partial^2 f(\bar{y})}{\partial^2 x_1}$ and let h_2 be the index for $\frac{\partial^2 f(\bar{y})}{\partial^2 x_2}$. Consider the case $\frac{\partial^2 f^0(\bar{y})}{\partial^2 x_1} / \frac{\partial^2 f^0(\bar{y})}{\partial^2 x_2} > \frac{\partial^2 f^1(\bar{y})}{\partial^2 x_1} / \frac{\partial^2 f^1(\bar{y})}{\partial^2 x_2}$ and let $\{h_{1,n}\}_{n \in \mathbb{N}}$ be a sequence that converges to 0. Let $\{h_{2,n}\}_{n \in \mathbb{N}}$ be a sequence

$$h_{2,n} = h_{1,n} \sqrt{\frac{1}{2} \left(\frac{\partial^2 f^0(\bar{y})}{\partial^2 x_1} / \frac{\partial^2 f^0(\bar{y})}{\partial^2 x_2} + \frac{\partial^2 f^1(\bar{y})}{\partial^2 x_1} / \frac{\partial^2 f^1(\bar{y})}{\partial^2 x_2} \right)}. \quad (28)$$

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

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

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

$$\frac{\partial^2 f^0(\bar{y})}{\partial^2 x_1} / \frac{\partial^2 f^0(\bar{y})}{\partial^2 x_2} > \frac{h_{2,n}^2}{h_{1,n}^2} > \frac{\partial^2 f^1(\bar{y})}{\partial^2 x_1} / \frac{\partial^2 f^1(\bar{y})}{\partial^2 x_2} \quad (29)$$

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

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

$$\Upsilon(h_1, h_2; f) = \frac{f(\bar{y} + 2h_1 e_1) - 2f(\bar{y} + h_1 e_1) + f(\bar{y})}{h_1^2} \cdot \left(\frac{f(\bar{y} + 2h_2 e_2) - 2f(\bar{y} + h_2 e_2) + f(\bar{y})}{h_2^2} \right)^{-1} \quad (30)$$

for $f \in \mathcal{F}$. By the definition of a second partial derivative, (25), the ratio $\Upsilon(h_1, h_2; f)$ converges to $\frac{\partial^2 f(\bar{y})}{\partial^2 x_1} / \frac{\partial^2 f(\bar{y})}{\partial^2 x_2}$ for $f \in \{f^0, f^1\}$ as $n \rightarrow \infty$ and $(h_{1,n}, h_{2,n}) \rightarrow (0, 0)$. Then there exists some $n_1 \in \mathbb{N}$ where, for all $n \geq n_1$, $n \in \mathbb{N}$, $|\Upsilon(h_1, h_2; f^0) - \frac{\partial^2 f^0(\bar{y})}{\partial^2 x_1} / \frac{\partial^2 f^0(\bar{y})}{\partial^2 x_2}| < \frac{\tau}{3}$ and $|\Upsilon(h_1, h_2; f^1) - \frac{\partial^2 f^1(\bar{y})}{\partial^2 x_1} / \frac{\partial^2 f^1(\bar{y})}{\partial^2 x_2}| < \frac{\tau}{3}$. The choice of distance $\frac{\tau}{3}$ ensures that

$$\Upsilon(h_1, h_2; f^0) > \frac{h_{2,n}^2}{h_{1,n}^2} > \Upsilon(h_1, h_2; f^1) \quad (31)$$

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

$$\Delta(h_1, h_2; f) = f(\bar{y} + 2h_1 e_1) - 2f(\bar{y} + h_1 e_1) + f(\bar{y}) - (f(\bar{y} + 2h_2 e_2) - 2f(\bar{y} + h_2 e_2) + f(\bar{y}))$$

for $f \in \mathcal{F}$. Choose $(h_1, h_2) = (h_{1,n}, h_{2,n})$ for $n \geq n_1$. Substituting the definition of $\Upsilon(h_1, h_2; f)$ into (31) and resulting algebra shows that at (h_1, h_2) , the ratios $h_{2,n}^2/h_{1,n}^2$ cancel in all terms and

$$\frac{f^0(\bar{y} + 2h_1 e_1) - 2f^0(\bar{y} + h_1 e_1) + f^0(\bar{y})}{f^0(\bar{y} + 2h_2 e_2) - 2f^0(\bar{y} + h_2 e_2) + f^0(\bar{y})} > 1 > \frac{f^1(\bar{y} + 2h_1 e_1) - 2f^1(\bar{y} + h_1 e_1) + f^1(\bar{y})}{f^1(\bar{y} + 2h_2 e_2) - 2f^1(\bar{y} + h_2 e_2) + f^1(\bar{y})}$$

and so

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

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

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

$$f(\bar{y} + 2h_1 e_1) + 2f(\bar{y} + h_2 e_2) + f(\bar{y}) > f(\bar{y} + 2h_2 e_2) + 2f(\bar{y} + h_1 e_1) + f(\bar{y}). \quad (32)$$

We can show that this is a local-production-maximization inequality, Definition 3.3, for some choice of B_1 and B_2 . Let $B_1 = \{\langle a, i \rangle, \langle b, j \rangle, \langle c, k \rangle, \langle g, l \rangle\}$ and $B_2 = \{\langle g, i \rangle, \langle a, j \rangle, \langle b, k \rangle, \langle c, l \rangle\}$, where the permutation π is implied by the definitions of B_1 and B_2 . Also, let $a \in C_i^u$, $b \in C_j^u$, $c \in C_k^u$ and $g \in C_l^u$. Let $\bar{x}(i, C_i^u) = \bar{y} + 2h_1 e_1$, $\bar{x}(j, C_j^u) = \bar{y} + h_2 e_2$, $\bar{x}(k, C_k^u) = \bar{y} + h_2 e_2$, $\bar{x}(l, C_l^u) = \bar{y}$, $\bar{x}(i, (C_i^u \setminus \{a\}) \cup \{g\}) = \bar{y} + h_1 e_1$, $\bar{x}(j, (C_j^u \setminus \{b\}) \cup \{a\}) = \bar{y} + h_1 e_1$, $\bar{x}(k, (C_k^u \setminus \{c\}) \cup \{b\}) = \bar{y} + 2h_2 e_2$, and $\bar{x}(l, (C_l^u \setminus \{g\}) \cup \{c\}) = \bar{y}$. Using the match-specific characteristics notation $\bar{x}(i, C^u) = \text{cat}\left(\left(x_{(d_1, i), 1}^{u, d}, \dots, x_{(d_1, i), K}^{u, d}\right), \dots, \left(x_{(d_n, i), 1}^{u, d}, \dots, x_{(d_n, i), K}^{u, d}\right)\right)$, let $x_{(a, i), 1}^{u, d} - 2h_1 e_1 = x_{(g, l), 1}^{u, d} = x_{(b, k), 1}^{u, d} = x_{(b, j), 1}^{u, d} = x_{(c, k), 1}^{u, d} = x_{(c, l), 1}^{u, d}$; $x_{(a, i), 1}^{u, d} - h_1 e_1 = x_{(a, j), 1}^{u, d} = x_{(g, i), 1}^{u, d}$; $x_{(b, k), 2}^{u, d} - 2h_2 e_2 = x_{(a, i), 2}^{u, d} = x_{(a, j), 2}^{u, d} = x_{(c, l), 2}^{u, d} = x_{(g, l), 2}^{u, d} = x_{(g, i), 2}^{u, d}$; and $x_{(b, k), 2}^{u, d} - h_2 e_2 = x_{(b, j), 2}^{u, d} = x_{(c, k), 2}^{u, d}$. By inspection, it can be seen that each match in B_1 exchanges a downstream firm partner for a match in B_2 . Meanwhile, each set of arguments $\bar{x}(i, C_i^u)$ on the right can be formed by replacing the characteristics associated with a single match in a set of arguments $\bar{x}(i, C_i^u)$ on the left. Therefore, Definition 3.3 is satisfied.

The remainder of the proof follows arguments in the proofs of Parts 1 and 2 of Theorem 5.1, and so is omitted.

A.4 Lemma 6.1: Continuous characteristics for the identification of production orderings

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

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

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

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

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

is non-empty. Assume not. Then f^1 and f^2 induce the same ordering, or preference relation in utility theory. The “only if” direction of Theorem 1.2 in Jehle and Reny (2000) shows that there must exist some positive, strictly monotonic function m such that $f^1(\bar{x}_a) = m \circ f^2(\bar{x}_a)$ over the range of values taken on by f^2 . As this contradicts Assumption 6.1, W^2 must be non-empty.

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

A.5 Theorem 6.1: Identification of production orderings, group characteristics

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

Lemma 6.1 produces \bar{x}_1 and \bar{x}_2 such that $f^0(\bar{x}_1) > f^0(\bar{x}_2)$ and $f^1(\bar{x}_1) < f^1(\bar{x}_2)$ or $f^0(\bar{x}_1) < f^0(\bar{x}_2)$ and $f^1(\bar{x}_1) > f^1(\bar{x}_2)$. Focus on the first case. An inequality such as $f^0(\bar{x}_1) > f^0(\bar{x}_2)$ considers a group of matches centered around an upstream firm on the left and another group of matches centered around an upstream firm on the right. This is not a local-production-maximization inequality (Definition 3.3), which would require at least two groups, each centered on an upstream firm, on both the left and the right.

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

$$f(\bar{x}_1) + f(\bar{x}_3) > f(\bar{x}_2) + f(\bar{x}_3). \quad (33)$$

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

I will now argue that this is a local-production-maximization inequality, Definition 3.3. Let $B_1 = \{\langle a, i \rangle, \langle b, j \rangle\}$ and $B_2 = \{\langle b, i \rangle, \langle a, j \rangle\}$. Also let there be sufficiently large (to handle \bar{y}) sets C_i^u and C_j^u where $a \in C_i^u$, $a \notin C_j^u$, $b \in C_j^u$, and $b \notin C_i^u$. Also define $\bar{x}(i, C_i^u) = \bar{x}_1$, $\bar{x}(j, C_j^u) = \bar{x}_3$, $\bar{x}(i, (C_i^u \setminus \{a\}) \cup \{b\}) = \bar{x}_2$ and $\bar{x}(j, (C_j^u \setminus \{b\}) \cup \{a\}) = \bar{x}_3$. Using the group-specific characteristics notation $\bar{x}(i, C^u) = (x_{(i, C^u), 1}^{\text{group}}, \dots, x_{(i, C^u), K}^{\text{group}})$, these four distinct groups can have the four different covariate vectors listed. With $\pi \langle a, i \rangle = b$ and $\pi \langle b, j \rangle = a$, inspection shows (33) satisfies

Definition 3.3 for the case of match-specific characteristics.

The remainder of the proof uses quite similar arguments to those in Part 1 of Theorem 5.1. The four groups are embedded into a larger matching market. Then I show that there is a continuum of markets X with similar properties. The main change to the argument is to allow for discrete covariates. I condition on the discrete elements of X at all steps and vary only the continuous elements to show that the set of markets X has positive probability. The case with $f^0(\bar{x}_1) < f^0(\bar{x}_2)$ and $f^1(\bar{x}_1) > f^1(\bar{x}_2)$ is similar: just reverse the local production maximization inequalities.

A.6 Theorem 6.2: Identification of production orderings, match characteristics

Let $f^0, f^1 \in \mathcal{F}$, where f^0 is the production function to be identified and f^1 is an alternative where $f^1(\bar{x}) \neq m \circ f^0(\bar{x})$ for all \bar{x} and for any positive monotonic function m . Lemma 6.1 produces \bar{x}_1 and \bar{x}_2 such that $f^0(\bar{x}_1) > f^0(\bar{x}_2)$ and $f^1(\bar{x}_1) < f^1(\bar{x}_2)$ or $f^0(\bar{x}_1) < f^0(\bar{x}_2)$ and $f^1(\bar{x}_1) > f^1(\bar{x}_2)$. Focus on the first case. We will now construct a local-production-maximization inequality.

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

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

$$f(\bar{x}_1) + \sum_{c=1}^{h-1} f(\bar{w}_c) > f(\bar{x}_2) + \sum_{c=1}^{h-1} f(\bar{w}_c) \quad (34)$$

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

A local-production-maximization inequality must satisfy Definition 3.3. The main challenge is that each group characteristic on the right side of the inequality must differ in only one vector of match-specific characteristics from a characteristics vector on the left side. This is because the equilibrium concept of pairwise stability does not allow more than one downstream firm to switch for each upstream firm. To show that (34) is indeed a local-production-maximization inequality, we need to show that we can pick $\{\bar{w}_c\}_{c=1}^{h-1}$ so that each term on the right side is only one match-specific characteristic vector separate from a term on the left side. The general construction of an \bar{w}_c for $c \leq h-1$ is

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

The construction is motivated as follows. From Definition 3.3, let $B_1 = \{\langle d_0, u_0 \rangle, \langle d_1, u_1 \rangle, \dots, \langle d_{h-1}, u_{h-1} \rangle\}$ and $B_2 = \{\langle d_1, u_0 \rangle, \langle d_2, u_1 \rangle, \dots, \langle d_0, u_{h-1} \rangle\}$. The group centered around upstream firm u_0 , with characteristics $\bar{x}(u_0, C_{u_0}) = \bar{x}_1$, replaces one downstream firm, $d_0 \in C_{u_0}$, with a new firm, $d_1 \in C_{u_1}$. A valid new match-specific

value for u_0 's new partner d_1 is, by intentional choice, $\bar{x}(u_0, \{d_1\}) = \bar{x}_{(1,2)}^{u,d}$, the first vector in \bar{x}_2 .³⁷ This results in a group of matches with characteristics $\bar{w}_1 = \bar{x}(u_0, (C_{u_0} \setminus \{d_0\}) \cup \{d_1\}) = (\bar{x}_{(1,1)}^{u,d}, \dots, \bar{x}_{(h-1,1)}^{u,d}, \bar{x}_{(1,2)}^{u,d})$ appearing on the right side of (34). Recall that we need to add the same terms on the left and right sides to move from $f(\bar{x}_1) > f(\bar{x}_2)$ to (34). So we add $f(\bar{w}_1) = f(\bar{x}(u_1, C_{u_1}))$ on the left side. The group centered around upstream firm u_1 replaces one downstream firm, $d_1 \in C_{u_1}$, with $d_2 \in C_{u_2}$. On the right side, $\bar{w}_2 = \bar{x}(u_1, (C_{u_1} \setminus \{d_1\}) \cup \{d_2\}) = (\bar{x}_{(1,1)}^{u,d}, \dots, \bar{x}_{(h-2,1)}^{u,d}, \bar{x}_{(1,2)}^{u,d}, \bar{x}_{(2,2)}^{u,d})$. As before, $f(\bar{w}_2) = f(\bar{x}(u_2, C_{u_2}))$ appears on the left side as well.

This iterative process truncates. A hypothetical \bar{w}_h equals \bar{x}_2 , one of the original two vectors from the beginning of the proof. Also, \bar{x}_1 equals a hypothetical \bar{w}_0 , the beginning of the iterative process. The above construction shows that each $\bar{x}(u_c, C_{u_c}) = \bar{w}_c$ on the left side exchanges one downstream firm d_c to yield $\bar{x}(u_c, (C_{u_c} \setminus \{d_c\}) \cup \{d_{c+1}\}) = \bar{w}_{c+1}$ on the right side. By inspection, each collection of characteristics $\bar{x}(u_c, (C_{u_c} \setminus \{d_c\}) \cup \{d_{c+1}\})$ is different from $\bar{x}(u_c, C_{u_c})$ by the characteristics of one match: $\bar{x}(u_c, \{d_{c+1}\}) = \bar{x}_{(c+1,2)}^{u,d}$ instead of $\bar{x}(u_c, \{d_c\}) = \bar{x}_{(h-c,1)}^{u,d}$. Therefore, (34) is a valid local-production-maximization inequality according to Definition 3.3.

The remainder of the proof follows arguments similar to those in the conclusion of Part 1 of Theorem 5.1 and the extension to discrete covariates in Theorem 6.1. The case with $f^0(\bar{x}_1) < f^0(\bar{x}_2)$ and $f^1(\bar{x}_1) > f^1(\bar{x}_2)$ is similar: just reverse the local production maximization inequalities.

A.7 Theorem 6.3: Identification of production orderings, firm characteristics

Let $f^0, f^1 \in \mathcal{F}$, where f^0 is the production function to be identified and f^1 is an alternative where $f^1(\bar{x}) \neq m \circ f^0(\bar{x})$ for all \bar{x} and for any positive monotonic function m . Lemma 6.1 produces \bar{x}_1 and \bar{x}_2 such that $f^0(\bar{x}_1) > f^0(\bar{x}_2)$ and $f^1(\bar{x}_1) < f^1(\bar{x}_2)$ or $f^0(\bar{x}_1) < f^0(\bar{x}_2)$ and $f^1(\bar{x}_1) > f^1(\bar{x}_2)$. Focus on the first case. We will now construct a local-production-maximization inequality.

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

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

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

³⁷Keep in mind that the characteristics are match-specific, so there is no requirement that the characteristics of a firm be the same on the left and right sides.

$$\begin{aligned}
f(\bar{x}_1) + \sum_{a=1}^{l-1} f(\bar{x}_1^u, \bar{x}_1^{d,1}, \dots, \bar{x}_a^{d,1}) + f(\bar{x}_1^u) + \sum_{a=1}^l f(\bar{x}_a^{d,1}) + \sum_{a=1}^{l-1} f(\bar{x}_2^u, \bar{x}_1^{d,2}, \dots, \bar{x}_a^{d,2}) + f(\bar{x}_2^u) + \sum_{a=1}^l f(\bar{x}_a^{d,2}) > \\
\sum_{a=1}^{l-1} f(\bar{x}_1^u, \bar{x}_1^{d,1}, \dots, \bar{x}_a^{d,1}) + f(\bar{x}_1^u) + \sum_{a=1}^l f(\bar{x}_a^{d,1}) + \sum_{a=1}^{l-1} f(\bar{x}_2^u, \bar{x}_1^{d,2}, \dots, \bar{x}_a^{d,2}) + f(\bar{x}_2^u) + \sum_{a=1}^l f(\bar{x}_a^{d,2}) + f(\bar{x}_2). \quad (35)
\end{aligned}$$

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

By inspection, one can loosely verify that (35) is almost, but not quite, a local-production-maximization inequality, Definition 3.3, with firm-specific characteristics. The term \bar{x}_1 on the left exchanges $\bar{x}_1^{d,1}$ for the option of being unmatched, 0, to add $f(\bar{x}_1^u, \bar{x}_1^{d,1}, \dots, \bar{x}_{l-1}^{d,1}) + f(\bar{x}_1^{d,1})$ on the right side. Following a pattern, each term $f(\bar{x}_1^u, \bar{x}_1^{d,1}, \dots, \bar{x}_a^{d,1})$ on the left side splits away the term $\bar{x}_a^{d,1}$ to leave a $f(\bar{x}_1^u, \bar{x}_1^{d,1}, \dots, \bar{x}_{a-1}^{d,1}) + f(\bar{x}_a^{d,1})$ on the right. Each term on the left involving the characteristics originally from \bar{x}_2 , for example $f(\bar{x}_2^u, \bar{x}_1^{d,2}, \dots, \bar{x}_a^{d,2})$, combines with an unmatched $\bar{x}_{a+1}^{d,1}$ to form $f(\bar{x}_2^u, \bar{x}_1^{d,2}, \dots, \bar{x}_{a+1}^{d,2})$ on the right side.

The inequality (35) is not a local-production-maximization inequality. For example, look at the terms $f(\bar{x}_1^u) + \sum_{a=1}^l f(\bar{x}_a^{d,1})$ on the left side. These unmatched firms do not combine with other firms to make pairings on the right side of (35). Therefore, as written (35) is not a local-production-maximization inequality according to Definition 3.3. However, the statement of the theorem imposes a non-innocuous localization normalization, which gives $f(\bar{x}_1^u) + \sum_{a=1}^l f(\bar{x}_a^{d,1}) = 0$ on the left and $f(\bar{x}_2^u) + \sum_{a=1}^l f(\bar{x}_a^{d,2}) = 0$ on the right. With this change, (35) becomes

$$\begin{aligned}
f(\bar{x}_1) + \sum_{a=1}^{l-1} f(\bar{x}_1^u, \bar{x}_1^{d,1}, \dots, \bar{x}_a^{d,1}) + \sum_{a=1}^{l-1} f(\bar{x}_2^u, \bar{x}_1^{d,2}, \dots, \bar{x}_a^{d,2}) + f(\bar{x}_2^u) + \sum_{a=1}^l f(\bar{x}_a^{d,2}) > \\
\sum_{a=1}^{l-1} f(\bar{x}_1^u, \bar{x}_1^{d,1}, \dots, \bar{x}_a^{d,1}) + f(\bar{x}_1^u) + \sum_{a=1}^l f(\bar{x}_a^{d,1}) + \sum_{a=1}^{l-1} f(\bar{x}_2^u, \bar{x}_1^{d,2}, \dots, \bar{x}_a^{d,2}) + f(\bar{x}_2), \quad (36)
\end{aligned}$$

which by the above informal arguments is a local-production-maximization inequality. I intentionally do not remove from (35) all production functions with zero production. Even though the production $f(\bar{x}_1^u)$ of singleton matches is zero, these production functions are needed to show (36) satisfies the definition of a local-production-maximization inequality, Definition 3.3.

The above arguments were informal. I will now formally show that (36) satisfies Definition 3.3. There are $3l$ terms on the left side of (36). The number $3l$ explains the statement in the theorem, ‘‘Further, let there be assignments A that contain as many matched coalitions as three times the maximum quota of an upstream firm.’’ Let $B_1 = \{\langle d_1, u_1 \rangle, \dots, \langle d_{3l}, u_{3l} \rangle\}$, where the indexing $\langle d_c, u_c \rangle$ follows the order on the left side of (36), from left to right. As I will show, many of these match partners will be 0, representing being unmatched. Now let

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

The match $\langle d_{l+1}, u_1 \rangle \in B_2$ means that the upstream firm u_1 , which on the left side has characteristics $\bar{x}(u_1, C_{u_1}) = \bar{x}_1$, exchanges a downstream firm d_1 for the downstream firm $d_{l+1} \in C_{u_{l+1}}$. In this case, downstream firm d_1 has the characteristics $\bar{x}_1^{d,1}$ while d_{l+1} is actually a dummy partner, 0, representing being unmatched. For each index $c = 1, \dots, 3l$, Table 2 lists the upstream firm characteristics, the characteristics for the group of all firms u_c and C_{u_c} , downstream firm d_c 's characteristics for the match in B_1 , the downstream-firm partner in the permutation

π creating B_2 , the characteristics of that downstream-firm partner, and the characteristics of the entire group of all firms u_c and its downstream-firm partners after the switch. One can verify that the characteristics of the firm $\pi \langle d_c, u_c \rangle$ in the fifth column are always the same as the characteristics of that downstream firm in the third column. This is the key idea behind showing that (36) is a local-production-maximization inequality with firm-specific characteristics: the characteristics of downstream firms remain the same after the permutation of partners between B_1 and B_2 .

The remainder of the proof follows arguments similar to those in the conclusion of Part 1 of Theorem 5.1 and the extension to discrete covariates in Theorem 6.1. The case with $f^0(\bar{x}_1) < f^0(\bar{x}_2)$ and $f^1(\bar{x}_1) > f^1(\bar{x}_2)$ is similar: just reverse the local production maximization inequalities.

References

- Akkus, Oktay and Ali Hortacsu**, “The Determinants of Bank Mergers: A Revealed Preference Analysis,” March 2007. University of Chicago working paper.
- Andrews, Donald, Steven T. Berry, and Panle Jia**, “Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Store Locations,” 2004. working paper.
- Bajari, Patrick, Han Hong, and Stephen Ryan**, “Identification and Estimation of a Discrete Game of Complete Information,” 2008. working paper.
- , **Victor Chernozhukov, Han Hong, and Denis Nekipelov**, “Nonparametric and Semiparametric Analysis of a Dynamic Discrete Game,” April 2009. working paper.
- Becker, Gary S.**, “A Theory of Marriage: Part I,” *Journal of Political Economy*, July-August 1973, 81 (4), 813–846.
- Beresteanu, Arie, Ilya Molchanov, and Francesca Molinari**, “Sharp Identification Regions in Games,” June 2008. working paper.
- Berry, S. and E. Tamer**, “Identification in models of oligopoly entry,” in Whitney Newey Richard K. Blundell and Torsten Persson, eds., *Advances in Economics and Econometrics: Theory and Applications: Ninth World Congress*, Cambridge University Press, 2006, chapter 2, p. 46.
- Berry, Steven T.**, “Estimation of a Model of Entry in the Airline Industry,” *Econometrica*, 1992, 60 (4), 889–917.
- Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff**, “Analyzing the Determinants of the Matching Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets,” 2003. working paper.
- Bresnahan, Timothy F. and Peter C. Reiss**, “Entry and Competition in Concentrated Markets,” *The Journal of Political Economy*, October 1991, 99 (5), 977–1009.
- Camiña, Ester**, “A Generalized Assignment Game,” *Mathematical Social Sciences*, 2006, 52 (2), 152–161.
- Choo, Eugene and Aloysius Siow**, “Who Marries Whom and Why,” *The Journal of Political Economy*, November 2006, 114 (1), 175–201.

- Ciliberto, Federico and Elie Tamer**, “Market Structure and Multiple Equilibria in Airline Markets,” October 2008. working paper.
- Dagsvik, John K.**, “Aggregation in Matching Markets,” *International Economic Review*, February 2000, 41 (1), 27–57.
- Demange, Gabrielle, David Gale, and Marilda A. Oliveira Sotomayor**, “Multi-Item Auctions,” *The Journal of Political Economy*, August 1986, 94 (4), 863–872.
- Elyakime, Bernard, Jean-Jaques Laffont, Patrice Loisel, and Quang Vuong**, “First-Price Sealed-Bid Auctions with Secret Reservation Prices,” *Annales d’Economie et de Statistique*, 1994, 34 (0), 115–141.
- 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.
- , “Estimating Matching Games with Transfers,” March 2009. working paper.
- **and Patrick Bajari**, “Measuring the Efficiency of an FCC Spectrum Auction,” 2009. working paper.
- Gale, David and Lloyd Shapley**, “College Admissions and the Stability of Marriage,” *American Mathematical Monthly*, 1962, 69.
- Galichon, Alfred and Marc Henry**, “Inference in Models with Multiple Equilibria,” 2008. working paper.
- Garicano, Luis and Esteban Rossi-Hansberg**, “Organization and Inequality in a Knowledge Economy,” *The Quarterly Journal of Economics*, November 2006, 121 (4), 1383–1435.
- Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey**, “Regular Quantal Response Equilibrium,” *Experimental Economics*, December 2005, 8, 347–367.
- Gordon, Nora and Brian Knight**, “The Causes of Political Integration: An Application to School Districts,” *Journal of Public Economics*, 2009.
- Gourieroux, Christian and Alain Monfort**, *Statistics and Econometric Models*, Vol. 1 of *Themes in Modern Econometrics*, Cambridge, 1995.
- 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.
- Horowitz, Joel L.**, *Semiparametric Methods in Econometrics*, Vol. 131 of *Lecture Notes in Statistics*, Springer, 1998.
- 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.
- Jehle, Geoffrey A. and Philip J. Reny**, *Advanced Microeconomic Theory*, Addison Wesley Longman, 2000.

- Jia, Panle**, “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry,” *Econometrica*, November 2008, 76 (6), 1263–1316.
- Kelso, Alexander S. and Vincent P. Crawford**, “Job Matching, Coalition Formation, and Gross Substitutes,” *Econometrica*, November 1982, 50 (6), 1483–1504.
- Koopmans, Tjalling C. and Martin Beckmann**, “Assignment Problems and the Location of Economic Activities,” *Econometrica*, January 1957, 25 (1), 53–76.
- Kovalenkoy, 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: Cheltenham, 2003.
- Leonard, Herman B.**, “Elicitation of Honest Preferences for the Assignment of Individuals to Positions,” *The Journal of Political Economy*, June 1983, 91 (3), 461–479.
- Levine, Anna A.**, “Licensing and Scale Economies in the Biotechnology Pharmaceutical Industry,” June 2008. working paper.
- Lewbel, Arthur**, “Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables,” *Journal of Econometrics*, 2000, 97, 145–177.
- Lucas, Robert E.**, “On the Size Distribution of Business Firms,” *Bell Journal of Economics*, 1978, 9 (2), 508–523.
- Magnac, Thierry and Eric Maurin**, “Identification and information in monotone binary models,” *Journal of Econometrics*, 2007.
- Manski, Charles F.**, “Maximum Score Estimation of the Stochastic Utility Model of Choice,” *Journal of Econometrics*, 1975, 3, 205–228.
- , “Identification of Binary Response Models,” *Journal of the American Statistical Association*, September 1988, 83 (403), 729–738.
- Matzkin, Rosa L.**, “Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models,” *Econometrica*, September 1991, 59 (5), 1315–1327.
- , “Nonparametric identification and estimation of polychotomous choice models,” *Journal of Econometrics*, 1993, 58, 137–168.
- Mazzeo, Michael J.**, “Product Choice and Oligopoly Market Structure,” *The RAND Journal of Economics*, 2002, 33 (2), 221–242.
- Milgrom, Paul and Chris Shannon**, “Monotone Comparative Statics,” *Econometrica*, 1994, 62 (1), 157–180.
- Mindruta, Denisa**, “Value Creation in University-Firm Research Collaborations: A Matching Approach,” February 2009. working paper.
- Ostrovsky, Michael**, “Stability in Supply Chain Networks,” *American Economic Review*, December 2008, 98 (3), 897–923.

- Pakes, Ariel, Jack Porter, Katherine Ho, and Joy Ishii**, “Moment Inequalities and Their Application,” 2006. working paper.
- Pycia, Marek**, “Many-to-One Matching with Complementarities and Peer Effects,” April 2007. working paper.
- , “Non-Existence Result for Matching Games with Transfers,” March 2008. working paper.
- Rosen, Sherwin**, “Authority, Control, and the Distribution of Earnings,” *Bell Journal of Economics*, 1982, 13 (2), 311–323.
- Roth, Alvin E. and Marilda A. Oliveira Sotomayor**, *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis* Econometric Society Monographs, Cambridge University Press, 1990.
- Seim, Katja**, “An empirical model of firm entry with endogenous product-type choices,” *The RAND Journal of Economics*, 2006, 37 (3), 619–640.
- Shapley, Lloyd S. and Martin Shubik**, “The assignment game I: the core,” *International Journal of Game Theory*, 1972, 1, 111–130.
- Sørensen, Morten**, “How Smart is Smart Money? A Two-Sided Matching Model of Venture Capital,” *Journal of Finance*, December 2007, LXII (6), 2725–2762.
- Sotomayor, Marilda**, “The multiple partners game,” in Mukul Majumdar, ed., *Equilibrium and dynamics: essays in honour of David Gale*, Macmillan, 1992, chapter 17, pp. 322–336.
- , “The lattice structure of the set of the stable outcomes of the multiple partners assignment game,” *International Journal of Game Theory*, November 1999, 28 (4), 567–583.
- Tamer, Elie**, “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria,” *Review of Economic Studies*, 2003, 70 (1), 147–165.
- Yang, Yupin, Mengze Shi, and Avi Goldfarb**, “Estimating the Value of Brand Alliances in Professional Team Sports,” *Marketing Science*, 2009.

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

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

The rank order property says $\Pr(A_1 | X; f^0, S) - \Pr(A_2 | X; f^0, S) = 0$ for any S . Total match production is $f(X(i, \{a\})) + \varepsilon_{(a,i)}$, with the error's distribution given in the table. The assignment is calculated using linear programming (Roth and Sotomayor, 1990, Chapter 8). Each integral is simulated by using 1 million draws of the realizations for the collection of error terms for all matches and being single. Given the number of replications, the differences in the table probably do not reflect simulation error.

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

Table 2: Proof of Theorem 6.3: Demonstrating That (36) Is a Local-Production-Maximization Inequality

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