Quantitative Economics 1 (2010), 203–254 1759-7331/20100203

Identification in matching games

Jeremy T. Fox
Department of Economics, University of Michigan and NBER

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 production function is identified up to a positive monotonic transformation.

Keywords. Matching, identification, complementarities, two-sided matching, assignment games, vertical relationships.

JEL classification. C14, C78.

1. Introduction

Matching games are a new and important area of empirical interest. Consider the example of marriage. A researcher may have data on the marriages in each of several 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 types 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

Jeremy T. Fox: jtfox@umich.edu
Thanks to participants at seminars at many universities and conferences. I thank the National Science Foundation, the NET Institute, the Olin Foundation, and the Stigler Center for generous funding.

DOI: 10.3982/QE3

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 features 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) showed that complementarities result in assortative matching. I extend the informal identification result 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 production function for different types of matches up to a positive monotonic transformation. I learn whether match production is higher at one set of characteristics for the matched parties than at another set of characteristics. I extend an identification result from the single agent, multinomial choice literature of Matzkin (1993, Theorem 1). The extension is nontrivial 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, you 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 data sets 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 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, full stability. 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 fully stable 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 fully stable would be more controversial.

The identification arguments are fully 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 property 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 single agent, multinomial choice, a similar rank order property is derived as a consequence of the payoff of each discrete choice having an independent and identically distributed (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, but such an i.i.d. shock at the assignment level does. 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 rank order property is that the identification arguments lead to a computationally simple estimator. Matching markets can have hundreds or thousands of agents in them. In Fox (2010), I presented such a maximum score estimator for matching games and showed 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 (2010), we studied matching between bidders and licenses for sale in an FCC spectrum auction. In Fox (2010), I studied matching between automotive parts suppliers and automotive assemblers. In both cases, the data sets 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), Baccara, Imrohoroglu, Wilson, and Yariv (2009), Levine (2009), Mindruta (2009), and Yang, Shi, and Goldfarb (2009) conducted empirical work using the matching maximum score estimator in Fox (2010). The identification results here are directly relevant for the above empirical papers.

No paper has performed a fully nonparametric analysis of any sort of matching game. Choo and Siow (2006) provided a logit-based estimator for one-to-one matching games using infinite numbers of agents and a finite number of observed agent characteristics.[1] They identified production functions conditional on the parametric structure of the logit model for the error terms. In Fox and Bajari (2010), we presented a rank order property for many-to-one matching games using infinite numbers of agents and a continuum of observed agent characteristics. We argued that the rank order property in Fox and Bajari (2010) is implied by the logit assumptions in Choo and Siow. Once the rank order property is established, the fully nonparametric identification results in this paper extend almost trivially to the case with an infinite number of agents. Therefore, the assumption of logit errors in Choo and Siow is sufficient but not necessary to identify features of production functions in markets with an infinite number of agents. A literature has explored parametric estimation in nonnested Gale and Shapley (1962) matching games, that is, games without endogenous transfers (Boyd, Lankford, Loeb, and Wyckoff (2003), Gordon and Knight (2009), Sørensen (2007)). These papers do not study identification.

The results in this paper are relevant for empirical work. One goal of matching empirical work 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) found changes in the sorting patterns between broad

[1] Dagsvik (2000) also used the logit model.

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. Identification of production functions, or features of them, is thus important to answer these questions.

The identification results in this paper are specifically referenced in Fox and Bajari (2010). In that paper, we estimated the production function for companies matching to geographic mobile phone licenses in an FCC spectrum auction using qualitative data on which bidders win what packages of licenses. We use our estimates to measure the efficiency of the assignment in the auction. The current paper does not prove that the production function is identified up to an affine transformation. Such a theorem would be sufficient to identify the relative efficiency of multiple assignments nonparametrically. In Fox and Bajari, we first used the derivative-based identification arguments to cardinally measure the relative complementarities from higher-value bidders sorting to packages with greater scale and from packages of geographically nearby licenses being grouped together. Second, we use the results on the identification of production functions up to a positive monotonic transformation to argue that both of these types of bidder and license characteristics are "goods" that raise output. Given a parametric functional form that, informally, is specified up to the features of the production function that I prove are nonparametrically identified, this lets us use our production function 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.

I first present the intuition for the identification strategies and results for the simple example of marriage in Section 2. Once the intuition for the results has been presented, Section 3 introduces notation for many-to-many, two-sided matching games. Section 4 presents sufficient conditions for the maximum score-like rank order property. The last two sections present the main identification theorems. Section 5 discusses derivative-based identification and Section 6 discusses the identification of production levels up to positive monotonic transformations. The proofs of the theorems are provided 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, Chap. 8). The results in this section focus on intuition; there will be no formal proofs. The formal details will be shown for the many-to-many matching case in Sections 5 and 6.

2.1 *Utility functions, production functions, and transfers*

This section studies a market with two men, m_1 and m_2, and two women, w_1 and w_2. 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, $m_1 = (m_1^1, m_1^2)$, where the vector of man m_1's characteristics is equal to man m_1's schooling m_1^1 and wealth m_1^2. Superscripts refer to individual, scalar characteristics and subscripts refer to different agents. For woman w_1, $w_1 = (w_1^1, w_1^2)$. 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.

Each person can be married only once and to a person of the opposite sex. Let the utility of a man m for a woman w before transfers be $v^{\text{men}}(m, w)$, which is a function of the characteristics of m and w, rather than their indices directly. Likewise, let the utility of woman w for man m before transfers be $v^{\text{women}}(m, w)$. After a scalar equilibrium transfer $t_{\langle m,w\rangle}$ for match $\langle m, w\rangle$ is made, the utilities of the man and woman are $v^{\text{men}}(m, w) + t_{\langle m,w\rangle}$ and $v^{\text{women}}(m, w) - t_{\langle m,w\rangle}$, respectively. Transfers can be negative, so the convention that women pay men is an innocuous normalization. Being single is allowed. If man m is unmatched, write his utility as $v^{\text{men}}(m, 0) + 0$, where we say he is matched to a dummy agent 0. Women can also be single.

A key object of interest in this paper is the production from a match,

$$f(m, w) = f((m^1, m^2), (w^1, w^2)) = v^{\text{men}}(m, w) + v^{\text{women}}(m, w),$$

which is a function of the schooling of the man and the woman and the wealth of both. This paper seeks to identify features of the production function $f(m, w)$, and not the utility functions $v^{\text{men}}(m, w)$ and $v^{\text{women}}(m, w)$ separately. This is because, absent using the individual rationality decision to remain single, the production function $f(m, w)$ will govern the sorting patterns in the transferable utility matching games studied here. One can only identify *features* of the production function $f(m, w)$, in part because production is not directly observed in the data; only matches are.

Becker (1973) showed 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 the schooling of women are complements in production:

$$\frac{\partial^2 f((m^1), (w^1))}{\partial m^1 \, \partial w^1} > 0 \quad \forall m^1 \in \mathbb{R}, w^1 \in \mathbb{R}.$$

Assortative matching means highly schooled men match with highly schooled women. Likewise, antiassortative 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 $m = (m^1, m^2)$. 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.

2.2 *Data on multiple matching markets*

The econometrician has access to data on (A, X) with $X = (m_1, m_2, w_1, w_2)$ 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 X. Likewise, the researcher observes the assignment A, the set of matches, in each market. For example, if m_1 and w_1 as well as m_2 and w_2 marry, $A = \{\langle m_1, w_1\rangle, \langle m_2, w_2\rangle\}$, where $\langle m_1, w_1\rangle$ is the match between man m_1 and woman w_1. If m_2 and w_2 are instead single, $A = \{\langle m_1, w_1\rangle, \langle m_2, 0\rangle, \langle 0, w_2\rangle\}$.

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 $X = (m_1, m_2, w_1, w_2)$ 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.

Assume that all characteristics in X vary across markets and are continuous random variables with (if needed) full support. Let the support be a product space across 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. The researcher has access to the population data on i.i.d. market observations (A, X) and hence can identify the joint distribution of A and X. Let $\Pr(A \mid X)$ be the probability of observing assignment A given agent characteristics X. $\Pr(A \mid X)$ is known given the joint distribution of A and X, as is $G(X)$, the marginal distribution of X.

For the case of one-to-one matching (this property will not always generalize to many-to-many matching), Koopmans and Beckmann (1957), Shapley and Shubik (1972), and Becker (1973) proved that any equilibrium assignment A will maximize the sum of production of all matches in the economy, or A will maximize $\sum_{\langle m,w\rangle\in A} f(m, w)$, where $\langle m, w\rangle$ 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. The equilibrium concept is pairwise stability, which is formally defined in Section 3.2.

Let S index the stochastic structure of the model. If the matching market has error terms ψ, S will be the distribution of the error terms. Assumptions on how error terms enter the model will be informally discussed in Section 2.3 and formally discussed in Section 4. Given a production function f, stochastic structure S, and the assumption of independence between X and ψ, the model induces a probability distribution over assignments A given observed agent characteristics X:

$$\Pr(A \mid X; f, S) = \int_\psi 1[A \text{ is the chosen assignment } \mid X, \psi; f]\, dS(\psi). \tag{1}$$

This formulation requires that the assignment chosen (or the equilibrium assignment) is unique with probability 1, which holds in the one-to-one matching model if agent characteristics or error terms have continuous and product supports. If f^0 is the true production function and S^0 is the true stochastic structure, then the observed data are generated as $\Pr(A \mid X) = \Pr(A \mid X; f^0, S^0)$.

2.3 *The rank order property*

Identification will be based on what I call the rank order property. The rank order property defines the econometric model being studied. The property 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.

Property 2.1. Given f and S, a strong version of the *rank order property* states that $\Pr(A_1 \mid X; f, S) > \Pr(A_2 \mid X; f, S)$ if and only if $\sum_{\langle m,w\rangle \in A_1} f(m, w) > \sum_{\langle m,w\rangle \in A_2} f(m, w)$.

For example, focus on the two assignments where no agent is single. Let $A_1 = \{\langle m_1, w_1\rangle, \langle m_2, w_2\rangle\}$ and $A_2 = \{\langle m_1, w_2\rangle, \langle m_2, w_1\rangle\}$. Given the observable agent characteristics in X and the true f^0, the rank order property states that assignment A_1 is more frequently observed if and only if

$$f(m_1, w_1) + f(m_2, w_2) > f(m_1, w_2) + f(m_2, w_1). \tag{2}$$

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 deterministic production.

The rank order property holds exactly if the total production to an assignment A is $\sum_{\langle m,w\rangle \in A} f(m, w) + \psi_A$, where ψ_A is independently and identically distributed across assignments A (or is exchangeable across A). The error ψ_A occurs in the social planner's problem and can be thought of as a market equilibrating error (a friction). One might decompose $\psi_A = \sum_{\langle m,w\rangle \in A} \varepsilon_{\langle m,w\rangle,A}$, where $\varepsilon_{\langle m,w\rangle,A}$ is specific to match $\langle m, w\rangle$ and assignment A, but these errors enter the social planning problem only and should not be seen as unobserved heterogeneity in the payoff to each match.

The rank order property holds approximately in simulations if the payoff to a match is $f(m, w) + \varepsilon_{\langle m,w\rangle}$, where $\varepsilon_{\langle m,w\rangle}$ is independently and identically distributed across matches $\langle m, w\rangle$ and reflects unobserved heterogeneity in the payoff to a match observed by the agents in the matching game. This type of i.i.d. error is typically invoked on the literature on estimating static Nash games. Section 4 presents the relevant simulation evidence.

The rank order property is unlikely to hold if the production to a match is $f((m^1, m^2, \eta^{\text{man}}), (w^1, w^2, \eta^{\text{woman}}))$, where η^{man} is a scalar unobserved characteristic of a man and η^{woman} is a scalar unobserved characteristic of a woman, and both unobservables are i.i.d. None of the previous papers on parametric estimation in any sort of matching game allows these sorts of unobserved agent characteristics that enter the production of every match involving agents m or w (Boyd et al. (2003), Choo and Siow (2006), Sørensen (2007)). As the first paper on identification in any sort of matching game, I stick to the simple rank order property and do not explore identification in the presence of unobserved characteristics η^{man} and η^{woman}. Such an extension is a subject of my ongoing research, which is important because empirical papers that do not formally estimate an equilibrium matching model, such as Ackerberg and Botticini (2002), have found that such characteristics may be important.

The rank order property is convenient because it allows identification and estimation to proceed without solving integrals. Simulation arguments directly involve the integral in (1), so could involve numerically integrating out error terms equal to the number of men times the number of women if the payoff to a match is $f(m, w) + \varepsilon_{\langle m,w\rangle}$. This could involve hundreds of agents and thousands of error terms in a realistic data set. The identification arguments in this paper lead to the computationally simple maximum score estimator in Fox (2010).

2.4 *Identification of features and within classes of production functions*

As I wrote above, 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 (2).

There are two ways I proceed over the two classes of identification results that I will present: identifying features $c(f)$ of the production function f and sufficiently restricting the class of production functions $\mathcal{F}$ so that f^0 is point identified within it. I use an extension of a standard definition for point identification by Gourieroux and Monfort (1995, Section 3.4). The probability of a set of market characteristics $\mathcal{Y}$ is $\int_{\mathcal{Y}} dG(X)$.

Definition 2.1. Let $\mathcal{F}$ be a class of production functions and let $\mathcal{S}$ be a class of stochastic structures. 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.

(i) f^0 is *identified within the class of production functions* $\mathcal{F}$ if there do 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 $\Pr(A \mid X; f^1, S^1) = \Pr(A \mid X; f^0, S^0)$ for all feasible (A, X) with $X \notin \mathcal{Y}$.

(ii) Let $c(\cdot)$ be a known function of f that produces either a scalar, vector, another function of the arguments of f, or a vector of functions of the arguments of f. A *feature* of f^0 $c(f^0)$ is *identified* for the class of production functions $\mathcal{F}$ if there do 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 $\Pr(A \mid X; f^1, S^1) = \Pr(A \mid X; f^0, S^0)$ for all feasible (A, X) with $X \notin \mathcal{Y}$.

Identification of a feature implies identification within a suitably restricted class of production functions. If the feature $c(f^0)$ is identified for the class of production functions $\mathcal{F}$, then we can define a new class $\mathcal{F}^c$ where for no two $f^1, f^2 \in \mathcal{F}^c$ it is the case that $c(f^1) = c(f^2)$ and for every $f \in \mathcal{F}$ there exists an $f^1 \in \mathcal{F}^c$ such that $c(f^1) = c(f)$. Then we can argue that the unique $f^3 \in \mathcal{F}^1$ such that $c(f^3) = c(f^0)$ is identified within the class of production functions $\mathcal{F}^c$ whenever the feature $c(f^0)$ is identified for the class of production functions $\mathcal{F}$.

Identification of f^0 within the class of production functions $\mathcal{F}$ is the key step to proving that the probability limit of a maximum score objective function has a unique global optimum at the true production function f^0, when optimized over $f \in \mathcal{F}$.[2] Thus identification in the sense of either part (by the previous paragraph's argument) of Definition 2.1 leads immediately to a constructive identification result, where f^0 is constructed as the unique solution to a maximum score optimization problem. Fox (2010) discussed the maximum score estimation of production functions in matching games in more detail.

Let me now provide a template for how we can combine the definition of identification and the rank order property to show identification of a feature $c(f^0)$ of f^0, the true production function. Therefore, let f^1 be some other production function such that $c(f^1) \neq c(f^0)$. Identification will require us to find a set of market characteristics X where (2) holds for $f = f^0$ and

$$f(m_1, w_1) + f(m_2, w_2) \leq f(m_1, w_2) + f(m_2, w_1) \tag{3}$$

holds for $f = f^1$. Let $A_1 = \{\langle m_1, w_1\rangle, \langle m_2, w_2\rangle\}$ and $A_2 = \{\langle m_1, w_2\rangle, \langle m_2, w_1\rangle\}$. The inequality (2) for $f = f^0$ implies that $\Pr(A_1 \mid X) > \Pr(A_2 \mid X)$ in the population data, while (3) for $f = f^1$ implies $\Pr(A_1 \mid X) \leq \Pr(A_2 \mid X)$ if f^1 happened to generate the data, which it does not. $\Pr(A_1 \mid X) > \Pr(A_2 \mid X)$ and $\Pr(A_1 \mid X) \leq \Pr(A_2 \mid X)$ are exclusive possibilities. If the production function is continuous, an open set of market characteristics around this particular X will be decisive because f^0 and f^1 give different implications for the population data $\Pr(A \mid X)$ for all X in the open set.

2.5 *Derivative-based identification*

I will use the notion of a feature $c(f)$ to identify the signs of cross-partial derivatives and the ratios of cross-partial derivatives.

2.5.1. *Are two inputs complements or substitutes at a point?*

The first feature of f^0 that will be identifiable is the sign of $\frac{\partial^2 f^0(m,w)}{\partial m^1 \partial w^1}$, or whether the schooling levels of men and women are complements or substitutes in production. Here m and w are vectors of male and female characteristics. This extends the informal identification result of Becker (1973) in two ways: each m is a vector of two characteristics, schooling and wealth (not just schooling), and the signs of $\frac{\partial^2 f^0(m,w)}{\partial m^1 \partial w^1}$ can be positive for some couple characteristics (m, w) and negative for other (m, w). The sign of $\frac{\partial^2 f^0(m,w)}{\partial m^1 \partial w^1}$ will be learned for each (m, w) separately and so the signs will be known for all points in the support of f.

Let $\frac{\partial^2 f^0(m,w)}{\partial m^1 \partial w^1} > 0$ and so, to show identification, let f^1 be some other production function where $\frac{\partial^2 f^1(m,w)}{\partial m^1 \partial w^1} < 0$. We will need to find some characteristics $X = (m_1, m_2, w_1, w_2)$, which comprises eight characteristics because each element is a vector, where (2) holds for $f = f^0$ and (3) holds for $f = f^1$.

[2]The rank order property ensures that f^0 is one global optimum of the probability limit of the maximum score objective function. Identification of f^0 within the class of production functions $\mathcal{F}$ immediately shows that the global optimum is unique.

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,

$$\begin{aligned}\frac{\partial^2 f(m,w)}{\partial m^1\,\partial w^1} = \lim_{h\to 0}\big(\big(&f((m^1+h,m^2),(w^1+h,w^2)) - f((m^1+h,m^2),(w^1,w^2))\\ &- f((m^1,m^2),(w^1+h,w^2)) + f((m^1,m^2),(w^1,w^2))\big)/h^2\big),\end{aligned} \tag{4}$$

where h is the limit argument. The value of (m,w) where we wish to identify the sign of $\frac{\partial^2 f^0(m,w)}{\partial m^1\,\partial w^1}$ is given. We will work with markets with observables X of the form $X = ((m^1,m^2),(m^1+h,m^2),(w^1,w^2),(w^1+h,w^2))$. Men (m^1,m^2) and (m^1+h,m^2) have identical observable characteristics except that (m^1+h,m^2) has h more units of schooling than (m^1,m^2). Likewise, women (w^1,w^2) and (w^1+h,w^2) are identical except that (w^1+h,w^2) has h more units of schooling.

The numerator of the middle-difference quotient in (4) for $f=f^0$ will be positive for sufficiently small h, because $\frac{\partial^2 f^0(m,w)}{\partial m^1\,\partial w^1} > 0$. Likewise for sufficiently small $h>0$, the numerator of the middle-difference quotient in (4) 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

$$\begin{aligned}&f((m^1+h,m^2),(w^1+h,w^2)) + f((m^1,m^2),(w^1,w^2))\\ &\quad > f((m^1+h,m^2),(w^1,w^2)) + f((m^1,m^2),(w^1+h,w^2)).\end{aligned}$$

Likewise, the opposite inequality will hold for $f=f^1$. Now let there be two hypothetical assignments, $A_1 = \{\langle(m^1+h,m^2),(w^1+h,w^2)\rangle, \langle(m^1,m^2),(w^1,w^2)\rangle\}$ and $A_2 = \{\langle(m^1+h,m^2),(w^1,w^2)\rangle, \langle(m^1,m^2),(w^1+h,w^2)\rangle\}$. By the rank order property, f^0 implies $\Pr(A_1 \mid X) > \Pr(A_2 \mid X)$ and f^1 implies the reverse. We thus have identification of f^0 and hence we learn the sign of $\frac{\partial^2 f^0(m,w)}{\partial m^1\,\partial w^1}$. We can use data on the equilibrium outcomes to matching markets to learn whether two inputs are 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 (A_1) in more markets than they would antiassortatively match (A_2), while f^1 predicted antiassortative matching would occur in more markets. We can look at the population data on $\Pr(A \mid X)$ to see which is more common. The continuity of f will extend identification to an open set of market characteristics with positive probability.

Note that the identification result does not allow a researcher to tell whether a pair of inputs is "more" complementary at (m_1,w_1) than some other point (m_2,w_2). For example, $\frac{\partial^2 f^1(m_1,w_1)}{\partial m^1\,\partial w^1} = 5$ and $\frac{\partial^2 f^1(m_2,w_2)}{\partial m^1\,\partial w^1} = 7$ cannot be distinguished from any other pair of positive values.

2.5.2. *How important are complementarities for one pair of inputs compared to another pair?*

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? The multivariate model has multiple characteristics and we wish to identify the relative importance of the different pairs of characteristics in match production.

Say both schooling and wealth are complements: $\frac{\partial^2 f^0(m,w)}{\partial m^1\,\partial w^1} > 0$ and $\frac{\partial^2 f^0(m,w)}{\partial m^2\,\partial w^2} > 0$. We will identify the ratio of the complementarities of schooling to the complementarities of wealth,

$$\frac{\partial^2 f^0(m,w)}{\partial m^1\,\partial w^1} \Big/ \frac{\partial^2 f^0(m,w)}{\partial m^2\,\partial w^2}.$$

As before, the analysis is local: for a given value of the vectors of male and female characteristics (m, w), we can establish these ratios globally by varying (m, w). Also note that we are identifying the ratio of complementarities, which is an actual numerical value. This will be harder than identifying whether $\frac{\partial^2 f^0(m,w)}{\partial m^1\,\partial w^1} > \frac{\partial^2 f^0(m,w)}{\partial m^2\,\partial w^2}$, which is a qualitative comparison instead of a quantitative value.[3]

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

$$\frac{\partial^2 f^0(m,w)}{\partial m^1\,\partial w^1} \Big/ \frac{\partial^2 f^0(m,w)}{\partial m^2\,\partial w^2} > \frac{\partial^2 f^1(m,w)}{\partial m^1\,\partial w^1} \Big/ \frac{\partial^2 f^1(m,w)}{\partial m^2\,\partial w^2}.$$

We will need to embellish the running example and allow each matching market to have three men and three women. Let all men start at the baseline characteristics (m^1, m^2). One man $(m^1 + h_1, m^2)$ has h_1 extra units of schooling and another man $(m^1, m^2 + h_2)$ has h_2 extra units of wealth. Likewise, there are three women, (w^1, w^2), $(w^1 + h_1, w^2)$, and $(w^1, w^2 + h_2)$. Now $X = ((m^1, m^2), (m^1 + h_1, m^2), (m^1, m^2 + h_2), (w^1, w^2), (w^1 + h_1, w^2), (w^1, w^2 + h_2))$.

The formal argument using the limits of middle-difference quotients is somewhat technical and will appear in the proof of Theorem 5.3 for the many-to-many matching case. 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,

$$\begin{aligned} &f((m^1 + h_1, m^2), (w^1 + h_1, w^2)) + f((m^1, m^2 + h_2), (w^1, w^2)) \\ &\quad + f((m^1, m^2), (w^1, w^2 + h_2)) \end{aligned}$$

[3]The identification of $\frac{\partial^2 f^0(m,w)}{\partial m^1\,\partial w^1} / \frac{\partial^2 f^0(m,w)}{\partial m^2\,\partial w^2}$ may seem 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.

$$> f((m^1, m^2 + h_2), (w^1, w^2 + h_2)) + f((m^1 + h_1, m^2), (w^1, w^2)) + f((m^1, m^2), (w^1 + h_1, w^2)).$$

The reverse inequality will hold under $f = f^1$. On the left side, there is the total production from an assignment $A_1 = \{\langle(m^1 + h_1, m^2), (w^1 + h_1, w^2)\rangle, \langle(m^1, m^2 + h_2), (w^1, w^2)\rangle, \langle(m^1, m^2), (w^1, w^2 + h_2)\rangle\}$, where the man and woman with h_1 extra units of schooling marry. Also, the man and woman with h_2 extra units of wealth each marry baseline individuals. On the right side, there is the total production from an assignment

$$A_2 = \{\langle(m^1, m^2 + h_2), (w^1, w^2 + h_2)\rangle, \langle(m^1 + h_1, m^2), (w^1, w^2)\rangle, \langle(m^1, m^2), (w^1 + h_1, w^2)\rangle\},$$

where the couple who both have h_2 of extra wealth marry, and the man and the woman with h_1 of extra schooling each marry a baseline person.

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

2.6 *Identification of production functions up to positive monotonic transformations*

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 $f^1(m^1, w^1) = -(m^1 - w^1)^2$ and $f^2(m^1, w^1) = 2m^1 \cdot w^1$ both have $\frac{\partial^2 f(m^1, w^1)}{\partial m^1 \partial w^1} = 2$. 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 who has the same level of schooling. Schooling is a horizontal attribute. Under f^2, any man's marriage's production will be maximized by matching with the most educated woman. Schooling is a vertical attribute.

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 (m, w) than another argument. For any two sets of characteristics for a pair of matches $\langle m_1, w_1\rangle$ and $\langle m_2, w_2\rangle$, we need to identify whether $f^0(m_1, w_1) > f^0(m_2, w_2)$ or the reverse. Ordering production function levels is helpful in distinguishing whether an individual match characteristic such as schooling is actually a vertical attribute that raises output.

Matzkin (1993, Theorem 1) proved the identification of utility functions for the single-agent, multinomial choice model. Matzkin showed that production functions are identified within a class $\mathcal{U}$ of functions where no two utility functions are related by a positive monotonic transformation. Formally, Matzkin found identification of the utility function in part by sufficiently restricting the class of utility functions to rule out positive monotonic transformations. Following Matzkin, restrict attention to a class of production functions $\mathcal{F}$ where no two members are related by a positive monotonic transformation. Let $f^0 \in \mathcal{F}$ be the true production function and let $f^1 \in \mathcal{F}$ be some alternative not related to f^0 by a positive monotonic transformation. Matzkin proved there exists two points (m_1, w_1) and (m_2, w_2) where, without loss of generality, $f^0(m_1, w_1) > f^0(m_2, w_2)$ and $f^1(m_1, w_1) < f^1(m_2, w_2)$. 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 (m_1, w_1) more frequently than the product with characteristics (m_2, w_2). 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 the variation in the exogenous market-level characteristics of matches to show identification.

We need to transform the inequality $f^0(m_1, w_1) > f^0(m_2, w_2)$ 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 $f^0(m_1, w_1) > f^0(m_2, w_2)$. Focusing on single people will resolve this dilemma. Adding the payoff for each agent being single to each side of the inequality gives

$$\begin{aligned} &f^0(m_1, w_1) + f^0(m_1, 0) + f^0(m_2, 0) + f^0(0, w_1) + f^0(0, w_2) \\ &\quad > f^0(m_2, w_2) + f^0(m_1, 0) + f^0(m_2, 0) + f^0(0, w_1) + f^0(0, w_2). \end{aligned} \tag{5}$$

This inequality almost involves two assignments to the same market, X. An issue is that a man with characteristics m_1 appears on the left side twice, as does a woman w_1. Similarly, man m_2 and woman w_2 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

$$f^0(m_1, w_1) + f^0(m_2, 0) + f^0(0, w_2) > f^0(m_2, w_2) + f^0(m_1, 0) + f^0(0, w_1).$$

We could set all single matches to 0, but doing so would return us to $f^0(m_1, w_1) > f^0(m_2, w_2)$. On the left side, we have an assignment $A_1 = \{\langle m_1, w_1\rangle, \langle m_2, 0\rangle, \langle 0, w_2\rangle\}$ and on the right side we have an assignment $A_2 = \{\langle m_2, w_2\rangle, \langle m_1, 0\rangle, \langle 0, w_1\rangle\}$. The production function f^0 implies $\Pr(A_1 \mid X) > \Pr(A_2 \mid X)$ and f^1 implies the reverse, for

$X = (m_1, m_2, w_1, w_2)$. By the continuity of f, there will be a set of X with positive probability where the assignment probabilities for A_1 and A_2 are differently ordered by the true f^0 and the alternative f^1. Thus, f^0 is identified in the class $\mathcal{F}$, where $\mathcal{F}$ contains no two production functions that are related by a positive monotonic transformation.

The economic intuition here is simple. If we wish to know whether or not $f^0(m_1, w_1) > f^0(m_2, w_2)$, we merely need to see the relative frequency of m_1 and w_1 being married and m_2 and w_2 being single (A_1) compared to m_2 and w_2 being married and m_1 and w_1 being single (A_2). The couple with the higher marital production will be single less often.

Observations on single people are often found in marriage data. The need to use single people arises because the example used agent-specific characteristics. Identification of production functions up to a positive monotonic transformation will be proved below for match-specific and group-of-matches-specific characteristics. Those results will not rely on single people.

The results on identification of cross-partial derivatives and the results on identification of production functions up to a positive monotonic transformation are complements. Both sets of results show different features of production functions that are identified. Neither identification result nests the other.

It is instructive to compare both results to the identification of utility functions up to a positive monotonic transformation in the single-agent model (Matzkin (1993, Theorem 1)). Taking a positive monotonic transformation of a production function may change the assignment that is observed, while such a transformation does not change outcomes in the single-agent model. Thus, the results about identification up to a positive monotonic transformation are less sharp than for the single-agent model. On the other hand, in single-agent choice, one cannot usually identify cardinal features of utility functions such as the ratio of the complementarities on one pair of inputs to the complementarities for another pair. Identification of cardinal features arises from the transferable utility structure of the matching games studied here.

3. Many-to-many matching games

The rest of the paper studies the general case of many-to-many matching games 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 (2010), Fox (2010)). This section outlines a two-sided, many-to-many matching game where all match characteristics are observable to other firms. In the next section, I reintroduce the rank order property and discuss sufficient conditions on error terms. The running example will be downstream firms matching with upstream firms.

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, 1999), Camiña (2006), and Jaume, Massó, and Neme (2009) 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, multisided matching, and Garicano and Rossi-Hansberg (2006) for the one-sided matching of workers into coalitions known as firms with hierarchical production. This paper uses the term "matching game" to encompass a broad class of transferable utility models, including some games where the original theoretical analyses used different names.

3.1 *Many-to-many matching markets*

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

Several exogenous objects define a matching market. Let U be a finite set of upstream firms indexed by u. Let D be a finite set of downstream firms indexed by d. Let $Q : U \cup D \to \mathbb{N}_+$ be the set of quotas, where $q_d^{\text{down}} \in Q$ is the quota of the downstream firm d and $q_u^{\text{up}} \in Q$ is the quota of the upstream firm u. 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 $\langle u, d\rangle$ be a match between downstream firm u and upstream firm d. As before, $\langle 0, d\rangle$ refers to an unfilled quota slot for a downstream firm and $\langle u, 0\rangle$ 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, $A = \{\langle u_1, d_1\rangle, \langle u_2, d_2\rangle, \dots, \langle u_N, d_N\rangle\}$, $T = \{t_{\langle u_1,d_1\rangle}, t_{\langle u_2,d_2\rangle}, \dots, t_{\langle u_N,d_N\rangle}\}$ is a set of transfers for all matches in A. Each $t_{\langle u,d\rangle} \in \mathbb{R}$ 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; the notation allows transfers to be negative. In a market with 100 upstream–downstream relationships, A is a set of 100 relationships and T is a set of 100 transfers between each of the matched partners. 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 $u \in U$ is

$$v^{\text{up}}\big(\vec{x}(u, D_u(A))\big) + \sum_{d \in D_u(A)} t_{\langle u,d\rangle}. \tag{6}$$

Here, the function $D_u(A)$ gives the collection of downstream firms u is matched to in assignment A,

$$D_u(A) = \begin{cases} \{d \in D \mid \langle u, d\rangle \in A\}, & \text{if } \{d \in D \mid \langle u, d\rangle \in A\} \neq \emptyset, \\ \{0\}, & \text{if } \{d \in D \mid \langle u, d\rangle \in A\} = \emptyset, \end{cases}$$

$v^{\text{up}}(\cdot)$ is the structural revenue function for upstream firms, and $\vec{x}(u, D_u)$ is the vector of characteristics entering the production function for the matches involving upstream firm u and an arbitrary collection of downstream partners $D_u \subseteq D \cup \{0\}$.[4] The payoff at (A, T) for $d \in D$ for the match $\langle u, d\rangle \in A$ is $v^{\text{down}}(\vec{x}(u, \{d\})) - t_{\langle u,d\rangle}$. I assume that the structural revenues of downstream firms are additively separable across multiple upstream firm partners: $v^{\text{down}}(\vec{x}(U_d, \{d\})) = \sum_{u\in U_d} v^{\text{down}}(\vec{x}(u, \{d\}))$, where $U_d \subseteq U \cup \{0\}$ is a collection of upstream firms. The reason for assuming additive separability in downstream firms' payoffs across multiple upstream firms will be clear when I reintroduce production functions below.

I will consider three types of characteristics. First consider the case studied in Section 2, where each agent has a fixed type, a vector $\vec{x}_d^{\text{down}}$ for downstream firm d, and a vector $\vec{x}_u^{\text{up}}$ for upstream firm u. For example, $\vec{x}_d^{\text{down}}$ could be the geographic location of d's assembly plant, information on the products manufactured by d, the markets d sells to, and so forth. Likewise, $\vec{x}_u^{\text{up}}$ could be the geographic location of u, the past experience of u, and so forth. If all characteristics are firm-specific, $\vec{x}(u, D_u) = \text{cat}(\vec{x}_u^{\text{up}}, \vec{x}_{d_1}^{\text{down}}, \dots, \vec{x}_{d_n}^{\text{down}})$, where $D_u = \{d_1, \dots, d_n\}$.[5]

I also consider cases where covariates vary directly at the match $\langle u, d\rangle$ or group-of-matches D_u levels. If all characteristic are match-specific, the long vector $\vec{x}(u, D_u) = \text{cat}(\vec{x}_{\langle u,d_1\rangle}^{\text{match}}, \dots, \vec{x}_{\langle u,d_n\rangle}^{\text{match}})$ for $D_u = \{d_1, \dots, d_n\}$, where each $\vec{x}_{\langle u,d\rangle}^{\text{match}}$ is the vector of characteristics of the match $\langle u, d\rangle$. An example of an element of $\vec{x}_{\langle u,d\rangle}^{\text{match}}$ is a measure of the degree of compatibility of two firms' inventory information systems. For group-specific characteristics, the vector $\vec{x}(u, D_u) = \vec{x}_{u,D_u}^{\text{group}}$ is not a concatenation of shorter vectors for covariates that operate at the match $\langle u, D_u\rangle$ level. An example of a group-specific characteristic is that $\vec{x}_{u,D_u}^{\text{group}}$ might include the percentage of upstream firm u's downstream firm partners that are located in countries with rigorous environmental regulations.[6]

X is the tuple of vectors 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,

$$\begin{aligned} X = \text{cat}\big(&(\vec{x}_u^{\text{up}})_{u\in U\cup\{0\}}, (\vec{x}_d^{\text{down}})_{d\in D\cup\{0\}}, \\ &\big(\vec{x}_{\langle u,d\rangle}^{\text{match}}\big)_{u\in U\cup\{0\}, d\in D\cup\{0\}}, \big(\vec{x}_{u,D_u}^{\text{group}}\big)_{u\in U\cup\{0\}, D_u\in\mathcal{P}(D\cup\{0\})}\big), \end{aligned} \tag{7}$$

where $\mathcal{P}(D \cup \{0\})$ is the power set of downstream firms. All elements of X may not be present in many applications. For example, an application may lack match-specific

[4] I use the vector notation only for characteristics, in part to reserve x for an element of $\vec{x}$.

[5] The 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.

[6] Some group or match characteristics may be built up from underlying firm characteristics, such as environmental standards in a firm's home country, that do not directly enter the production function. Other data sets may list characteristics that directly vary at the level of the match or group of matches. For example, we could measure whether each country sends an ambassador to each other country.

characteristics. In that case, just treat the corresponding terms as not being present in the definition of X.[7]

The fact that transfers enter total profits additively separably for both upstream and downstream firms allows us to focus on the following production function.

Definition 3.1. The *production function* for (u, D_u) for $u \in U$ and $D_u \subseteq D \cup \{0\}$ is

$$f(\vec{x}(u, D_u)) \equiv v^{\text{up}}(\vec{x}(u, D_u)) + \sum_{d \in D_u} v^{\text{down}}\big(\vec{x}(u, \{d\})\big).$$

For the Section 2 example of one-to-one matching with fixed types, $f(\vec{x}(u, \{d\})) \equiv v^{\text{up}}(\text{cat}(\vec{x}_u^{\text{up}}, \vec{x}_d^{\text{down}})) + v^{\text{down}}(\text{cat}(\vec{x}_u^{\text{up}}, \vec{x}_d^{\text{down}}))$.

I assume that the maximum quota for all upstream firms, $\max_{u \in U} q_u^{\text{up}}$, is known and finite. This means that $\vec{x}(u, D_u)$ has a known maximum number of elements. For many-to-many matching, a maximum quota and the additive separability of $v^{\text{down}}(\vec{x}(u, \{d\}))$ across multiple upstream firm partners makes the set of arguments of f finite.[8] 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 $f(\vec{x})$, where the argument $\vec{x}$ is an arbitrary vector of characteristics. When an upstream firm does not use all of its quota, null arguments can be included in the argument vector $\vec{x}$ of $f(\vec{x})$ to refer to the unfilled match slots.

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* under the following conditions:

[7]This 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 $\vec{x}_{u,D_u}^{\text{group}}$ for a $|D_u| > q_u^{\text{up}}$. The definition of X requires dummy arguments for characteristics involving the partner of being unmatched, 0, for notational conciseness later.

[8]Consider an example with matches $\langle u_1, d_1\rangle$, $\langle u_1, d_2\rangle$, and $\langle u_2, d_2\rangle$. If the model allowed arbitrary nonlinearities in both upstream and downstream firms' structural revenue functions, there would be a set of firms $\{u_1, u_2, d_1, d_2\}$ with production $f(\vec{x}(\{u_1, u_2\}, \{d_1, d_2\}))$, even though u_1 and u_2, u_2 and d_1, and d_1 and d_2 have no direct links.

(i) For all $\langle u_1, d_1\rangle \in A$, $\langle u_2, d_2\rangle \in A$, $\langle u_1, d_2\rangle \notin A$, and $\langle u_2, d_1\rangle \notin A$,

$$\begin{aligned} & v^{\mathrm{up}}\big(\vec{x}\big(u_1, D_{u_1}(A)\big)\big) + \sum_{d \in D_{u_1}(A)\setminus\{d_1\}} t_{\langle u_1, d\rangle} + t_{\langle u_1, d_1\rangle} \\ & \quad \geq v^{\mathrm{up}}\big(\vec{x}\big(u_1, \big(D_{u_1}(A)\setminus\{d_1\}\big) \cup \{d_2\}\big)\big) + \sum_{d \in D_{u_1}(A)\setminus\{d_1\}} t_{\langle u_1, d\rangle} + \tilde{t}_{\langle u_1, d_2\rangle}, \end{aligned} \tag{8}$$

where $\tilde{t}_{\langle u_1, d_2\rangle} \equiv v^{\mathrm{down}}(\vec{x}(u_1, \{d_2\})) - (v^{\mathrm{down}}(\vec{x}(u_2, \{d_2\})) - t_{\langle u_2, d_2\rangle})$.

(ii) For all $\langle u, d_1\rangle \in A$,

$$\begin{aligned} & v^{\mathrm{up}}\big(\vec{x}(u, D_u(A))\big) + \sum_{d \in D_u(A)\setminus\{d_1\}} t_{\langle u, d\rangle} + t_{\langle u, d_1\rangle} \\ & \quad \geq v^{\mathrm{up}}\big(\vec{x}(u, D_u(A)\setminus\{d_1\})\big) + \sum_{d \in D_u(A)\setminus\{d_1\}} t_{\langle u, d\rangle}. \end{aligned}$$

(iii) For all $\langle u, d\rangle \in A$,

$$v^{\mathrm{down}}\big(\vec{x}(u, \{d\})\big) - t_{\langle u, d\rangle} \geq 0.$$

(iv) The inequality (8) holds if either or both of the existing matches represent a free quota slot, namely $\langle u_1, d_1\rangle = \langle u_1, 0\rangle$ or $\langle u_2, d_2\rangle = \langle 0, d_2\rangle$. In this case, the transfers corresponding to the free quota slots in (8), $t_{\langle u_1, d_1\rangle}$ or $t_{\langle u_2, d_2\rangle}$, are set equal to 0.

Part (i) of the definition of pairwise stability says that upstream firm u_1 prefers its current downstream firm d_1 instead of some alternative downstream firm d_2 at the transfer $\tilde{t}_{\langle u_1, d_2\rangle}$ that makes d_2 switch to sourcing its supplies from u_1 instead of its equilibrium upstream firm, u_2. Because of transferable utility, u_1 can always cut its price and attract d_2's business; at an equilibrium, u_1 would lower its profit from doing so if the new business supplanted the relationship with d_1. Part (i) is the component of the definition of pairwise stability that I will focus on in this paper. Parts (ii) and (iii) deal with matched agents that do not profit by unilaterally dropping a relationship and becoming unmatched. These are individual rationality conditions: all matches must give an incremental positive surplus. Finally, part (iv) states that two firms, where one or both have free quota, do not want to form a match.

3.3 *Properties of pairwise stable outcomes in many-to-many matching games*

The literature has established existence and uniqueness results for pairwise stable equilibria (A, T) under a restriction on the upstream firm preferences $v^{\mathrm{up}}(\vec{x}(u, D_u))$ known as *substitutes*. Substitutes restricts how one upstream firm ranks sets of downstream firms.[9] In the interest of brevity, I refer the interested reader to Definitions 5 and 6 in

[9]Note that in this context "substitutes" refers to how multiple downstream firms enter preferences, not how different individual firm characteristics enter production functions. The latter use is the more common use of "substitutes" and "complements" in this article.

Hatfield and Kominers (2010) for this restriction expressed in a more general matching game in contract space. Hatfield and Kominers showed several results (some are implied by more general results). Under substitutable preferences, a pairwise stable equilibrium is guaranteed to exist. Furthermore, any pairwise stable equilibrium is automatically also *fully stable*. A fully stable equilibrium is robust to deviations by any coalition of downstream and upstream firms. One such coalition is the coalition of all firms. Because of this paper's focus on transferable utility, an assignment that is part of a fully stable equilibrium outcome will maximize the sum of production in the entire matching economy. Indeed, the equilibrium assignment will be computable by a sufficiently general social planner's linear programming problem. Furthermore, if the characteristics entering the production function have continuous support and the production function is continuous in its arguments, then a unique assignment A will solve the social planner's linear programming problem and hence be part of any pairwise stable equilibrium (A, T) with probability 1. To summarize, the substitutes condition ensures existence of a pairwise stable assignment and the uniqueness of that assignment with probability 1, in a many-to-many, transferable utility matching game. The substitutes condition also leads to a useful algorithm for computing equilibrium assignments.

This paper will not, for most sections, impose the substitutes condition, as many empirical applications of many-to-many matching games are to situations where an upstream firm could view two or more downstream firms as complements. For example, Fox (2010) empirically examined specialization by automotive suppliers, where a supplier could have higher production (or lower cost) by supplying multiple car parts to the same automotive assembler (here view a car part as a downstream firm), compared to one car part each to several assemblers.

If preferences allow for complementarities between multiple downstream firms matched to the same upstream firm, Hatfield and Milgrom (2005), Pycia (2008), Hatfield and Kojima (2008), and Hatfield and Kominers (2010) presented examples of preferences where no pairwise stable equilibrium exists. The counterexamples mean that general existence theorems do not exist. However, recent empirical and theoretical work on nontransferable utility matching games has presented situations under which the probability that a pairwise stable equilibrium exists converges to 1 as the number of agents in the matching market grows large (Kojima, Pathak, and Roth (2010)). In this sense, nonexistence of pairwise stable equilibria may be a minor problem in some markets of empirical interest. This paper will maintain the assumption that the data reflect a pairwise stable equilibrium.[10]

Under complementarities, there is no equivalence of pairwise and fully stable equilibria, so pairwise stable equilibria will not solve a social planner's problem. Pairwise stable equilibria can be inefficient. The latter property means multiple equilibrium assignments can exist with positive probability. I discuss multiple equilibrium assignments in Section 4.5.

[10]Other researchers have examined games with potential existence issues. In the nonnested-with-matching literature on estimating normal-form Nash games, Ciliberto and Tamer (2009) threw out a particular realization of the error term's contribution to the likelihood if no pure strategy equilibrium exists. In the current paper, probabilities will not sum to 1 if equilibria sometimes do not exist. In computational experiments, I rarely stumble across parameters where no pairwise stable outcome exists.

3.4 *Using matches only: Local production maximization*

A matching-game outcome (A, T) has two components. 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 u_1 and u_2, consider an example where $D_{u_1}(A) = \{d_1\}$ and $D_{u_2}(A) = \{d_2\}$. The inequality (8) becomes

$$\begin{aligned} v^{\text{up}}\big(\vec{x}(u_1, \{d_1\})\big) + t_{\langle u_1, d_1\rangle} \geq {} & v^{\text{up}}\big(\vec{x}(u_1, \{d_2\})\big) + v^{\text{down}}\big(\vec{x}(u_1, \{d_2\})\big) \\ & - \big(v^{\text{down}}\big(\vec{x}(u_2, \{d_2\})\big) - t_{\langle u_2, d_2\rangle}\big) \end{aligned} \tag{9}$$

after substituting in the definition of $\tilde{t}_{\langle u_1, d_2\rangle}$. Likewise, there is another inequality for u_2's deviation to match with d_1 instead of d_2:

$$\begin{aligned} v^{\text{up}}\big(\vec{x}(u_2, \{d_2\})\big) + t_{\langle u_2, d_2\rangle} \geq {} & v^{\text{up}}\big(\vec{x}(u_2, \{d_1\})\big) + v^{\text{down}}\big(\vec{x}(u_2, \{d_1\})\big) \\ & - \big(v^{\text{down}}\big(\vec{x}(u_1, \{d_1\})\big) - t_{\langle u_1, d_1\rangle}\big). \end{aligned} \tag{10}$$

Adding (9) and (10), canceling the transfers $t_{\langle u_1, d_1\rangle}$ and $t_{\langle u_2, d_2\rangle}$ 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

$$f\big(\vec{x}(u_1, \{d_1\})\big) + f\big(\vec{x}(u_2, \{d_2\})\big) \geq f\big(\vec{x}(u_1, \{d_2\})\big) + f\big(\vec{x}(u_2, \{d_1\})\big).$$

This is 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 $f(\vec{x}(u, \{d\})) = \vec{x}_u^{\text{up}} \beta^{\text{up}} + \vec{x}_d^{\text{down}} \beta^{\text{down}}$ for parameter vectors β^{up} and β^{down}, then a local production maximization inequality is

$$\begin{aligned} & \vec{x}_{u_1}^{\text{up}} \beta^{\text{up}} + \vec{x}_{d_1}^{\text{down}} \beta^{\text{down}} + \vec{x}_{u_2}^{\text{up}} \beta^{\text{up}} + \vec{x}_{d_2}^{\text{down}} \beta^{\text{down}} \\ & \quad \geq \vec{x}_{u_1}^{\text{up}} \beta^{\text{up}} + \vec{x}_{d_2}^{\text{down}} \beta^{\text{down}} + \vec{x}_{u_2}^{\text{up}} \beta^{\text{up}} + \vec{x}_{d_1}^{\text{down}} \beta^{\text{down}} \end{aligned} \tag{11}$$

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]

[11] For some policy questions, the cancellation of characteristics that are not interactions between the characteristics of multiple firms is an empirical advantage. Many data sets lack data on all important char-

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

LEMMA 3.3. *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*

$$B_2 = \{\langle \pi\langle u, d\rangle, u\rangle \mid \langle u, d\rangle \in B_1\}.$$

Then the inequality

$$\sum_{\langle u,d\rangle \in B_1} f\big(\vec{x}(u, D_u(A))\big) \geq \sum_{\langle u,d\rangle \in B_2} f\big(\vec{x}\big(u, D_u((A \backslash B_1) \cup B_2)\big)\big) \tag{12}$$

holds.

All proofs are found in the Appendix.[12] The definition of a local production maximization inequality is similar to (12), except that no particular outcome (A, T) needs to be stated. This definition will be used formally in the identification proofs.

DEFINITION 3.4. Let there be a set of matches B_1 and let B_2 be a permutation π of B_1, $B_2 = \{\langle \pi\langle u, d\rangle, u\rangle \mid \langle u, d\rangle \in B_1\}$. For each u where $\langle u, d\rangle \in B_1$, let there be a set of downstream firms D_u such that $\langle u, d\rangle \in B_1$ implies $d \in D_u$. Call

$$\sum_{\langle u,d\rangle \in B_1} f(\vec{x}(u, D_u)) \geq \sum_{\langle u,d\rangle \in B_1} f\big(\vec{x}\big(u, (D_u \backslash \{d\}) \cup \{\pi\langle u, d\rangle\}\big)\big)$$

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]

acteristics 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.

[12]A permutation π of the downstream firm partners applied to a set of matches $\{\langle u_1, d_1\rangle, \langle u_2, d_2\rangle, \langle u_3, d_3\rangle\}$ gives each upstream firm a new downstream firm partner. An example of a permutation is $\{\langle u_1, d_3\rangle, \langle u_2, d_1\rangle, \langle u_3, d_2\rangle\}$. Let $\pi\langle u_1, d_1\rangle = d_3$ give the index of the new downstream firm partner d_3 of the upstream firm u_1.

[13]The notation B_2 is not, strictly speaking, needed for Definition 3.4. Later I use B_1 and B_2 when showing that an inequality satisfies Definition 3.4.

[14]I have no proof that satisfying (12) 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.

4. The rank order property

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 data sets 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 so as 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. Then

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

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

[15]For 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 the elements of X corresponding to some firms, matches, or groups of matches are missing. See Fox (2007) for a related discussion on estimating the single-agent multinomial choice model without data on all available choices.

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

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

[18]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)

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 can be seen as a stochastic version of local production maximization. The benefits of the rank order property were given in the Introduction.

Property 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 $B_2 = \{\langle \pi\langle u,d\rangle, u\rangle \mid \langle u,d\rangle \in B_1\}$. Let $A_2 = (A_1 \backslash B_1) \cup B_2$. Let $S \in \mathcal{S}$ be any distribution of the error terms and let $f \in \mathcal{F}$ be any production function. The *rank order property* states that

$$\sum_{\langle u,d\rangle \in B_1} f\big(\vec{x}(u, D_u(A_1))\big) > \sum_{\langle u,d\rangle \in B_2} f\big(\vec{x}(u, D_u(A_2))\big) \tag{13}$$

if and only if

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

Keep in mind that X, f, and S are held fixed: the rank order property is an assumption about the stochastic structure of the model. To understand the rank order property, consider a situation where A_1 contains thousands of matches and $B_1 = \{\langle u_1, d_1\rangle, \langle u_2, d_2\rangle\}$ contains only two matches. Then $A_2 = (A_1 \backslash B_1) \cup B_2$ is equal to A_1 except that the matches $B_2 = \{\langle u_1, d_2\rangle, \langle u_2, d_1\rangle\}$ form. Given X and Q, neither A_1 nor 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 (13) 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 property says that A_1 will be more likely to be a pairwise stable assignment to some realized model than A_2.[19]

As the quotas in Q are not observed in many empirical applications, a slightly more primitive version of Property 4.1 is that (13) holds if and only if $\Pr(A_1 \mid X, Q; f, S) > \Pr(A_2 \mid X, Q; f, S)$ for any valid Q. Then taking expectations with respect to $Q \mid X$ gives Assumption 4.1. Even if Q is unobserved, for the most part I have only considered inequalities where the total number of matches of each agent in A_1 and A_2 is kept the same.[20] If unmatched agents are not considered in B_1 and B_2, and if A_1 is a feasible assignment for Q, then A_2 is also a feasible assignment for that Q.

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

[19]The rank order property can be rejected by the data if the following set is empty: the set of $f \in \mathcal{F}$ such that $\sum_{\langle u,d\rangle \in B_1} f(\vec{x}(u, D_u(A_1))) > \sum_{\langle u,d\rangle \in B_2} f(\vec{x}(u, D_u(A_2)))$ if and only if $\Pr(A_1 \mid X) > \Pr(A_2 \mid X)$ holds for all (A_1, A_2, B_1, B_2, X), where A_1, A_2, B_1, and B_2 are defined in Property 4.1 and $\Pr(A \mid X)$ is identified from the data. Itemizing over all $f \in \mathcal{F}$ is computationally difficult; more work needs to be done to operationalize a test based on this idea.

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

4.3 *A Sufficient condition for the rank order property*

This subsection explores a sufficient condition for the rank order property, Property 4.1, in the context of models where assignments have unobserved components in production. In this subsection only, I impose the condition that the equilibrium (A, T) satisfies full stability, as discussed in Section 3.3. This ensures that a pairwise stable equilibrium (A, T) always exists, that the equilibrium assignment A is unique with probability 1, and that the equilibrium assignment A is computable as a social planner's problem. Substitutable preferences is a sufficient condition for these properties.

There is a finite, although potentially large, number of assignments. The social planning problem is the single-agent, unordered, discrete choice problem of Manski (1975), where the single agent is the social planner choosing from the finite (but often large) number of assignments. From Manski's work, we know the sufficient condition that will arise. For an assignment A, let its total production be $\sum_{\langle u,d\rangle\in A} f(\vec{x}(u, D_u(A))) + \psi_A$, 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 density S and let ψ be independent of Q and X.[21] Let $\Pr(A \mid Q, X; f, S)$ be the probability A is picked by the social planner.

Lemma 4.1. *Let the payoff to the social planner for assignment A be $\sum_{\langle u,d\rangle\in A} f(\vec{x}(u, D_u(A))) + \psi_A$ and let the social planner choose an assignment to maximize its payoff. Let the density S be exchangeable. Then the rank order property* (*Property* 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]

The social planner errors can be interpreted as errors in the deterministic model from finding the true stable assignment. 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, Holt, and Palfrey (2005)). 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]

[21] 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 in single-agent, multinomial choice. While the density S can depend on X, it must still be exchangeable conditional on X. Thus, the rank order property is unlikely to hold if, across matches within a market, observed and unobserved components of match production are correlated: there is omitted variable bias.

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

[23] An exchangeable joint density for assignment-level errors is a sufficient but not necessary condition for the rank order property. Consider the comparison of an assignment $A_1 = \{\langle 1, 1\rangle, \langle 2, 2\rangle, \langle 3, 0\rangle, \langle 0, 3\rangle\}$, where downstream and upstream firms 3 are both unmatched, to another assignment $A_2 = \{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 3, 1\rangle\}$, where all firms are matched. The local production maximization inequality in Property 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 er-

4.4 *Match-specific error terms*

The closest analog to the practice of adding action-specific error terms to perfect information Nash games in matching games (a nonnested class of games) is adding match-specific error terms $\varepsilon_{\langle u,d\rangle}$ (Andrews, Berry, and Jia (2004), Bajari, Hong, and Ryan (2010), Beresteanu, Molchanov, and Molinari (2008), Berry (1992), Bresnahan and Reiss (1991), Ciliberto and Tamer (2009), Galichon and Henry (2008), Jia (2008), Mazzeo (2002), Tamer (2003)). Let the total output of a set of downstream firm partners D_u for upstream firm u be

$$f(\vec{x}(u, D_u)) + \sum_{u\in D_u} \varepsilon_{\langle u,d\rangle}, \tag{14}$$

where $\varepsilon_{\langle u,d\rangle}$ is the match-$\langle u, d\rangle$-specific error term, which is independent of all components of X and Q. Let the stochastic structure S represent the distribution of $\varepsilon_{\langle u,d\rangle}$. Let the game's outcome be fully stable to ensure uniqueness of equilibrium assignments with probability 1. Then, in the perfect information world where $\varepsilon_{\langle u,d\rangle}$ is observed by all agents in the model but is not in the data,

$$\Pr(A \mid X, Q; f, S) = \int_{\varepsilon} 1[A \text{ maximizes output} \mid X, Q, \varepsilon]\, dS(\varepsilon), \tag{15}$$

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.

For matching, unlike single-agent discrete choice, it is not a theorem that i.i.d. errors yield the rank order property for matching, Property 4.1. All models are approximations to reality. If the true production function is thought to include i.i.d. match-specific shocks as in (14) and, therefore, assignment probabilities are given by (15), 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 (14) is approximated by the rank order property, a natural generalization of prior work on matching games without econometric errors.

Table 1 includes results from simulations that compute assignment probabilities for a one-to-one, two-sided matching game where match production is $f(\vec{x}(u, \{d\})) + \varepsilon_{\langle u,d\rangle}$. There are three upstream firms and three downstream firms. The details of the game are given in the footnote to Table 1. The parameters of the game are chosen so that two assignments A_1 and A_2 give equal deterministic production, neither is the deterministic

rors 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 $\sum_{\langle u,d\rangle\in A_1} f(\vec{x}(u, D_u(A_1))) = \sum_{\langle u,d\rangle\in A_2} f(\vec{x}(u, D_u(A_2)))$ if and only if $\Pr(A_1 \mid X; f, S) = \Pr(A_2 \mid X; f, S)$. 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 weaker than a social planner with exchangeable assignment-level errors.

TABLE 1. Assignment probabilities for two assignments with equal deterministic production, with i.i.d. match-specific unobservables, by distribution.[a]

No. Firms $U = D$	Distribution for i.i.d. Errors, S	Error Standard Deviation	$\Pr(A_1 \mid X; f, S)$	$\Pr(A_2 \mid X; f, S)$	$\Pr(A_1 \mid X; f, S) - \Pr(A_2 \mid X; f, S)$
3	$N(0, 1)$	1	0.02126	0.03640	−0.01514
3	$N(0, 36)$	6	0.07897	0.07888	0.00009
3	$N(0, 400)$	20	0.06554	0.06543	0.00011
3	$0.33 \cdot N(0.5, 0.04) + 0.67 \cdot N(-0.5, 0.123)$	0.53	0.000098	0.001791	−0.00163
3	$0.33 \cdot N(2.5, 0.04) + 0.67 \cdot N(-2.5, 0.123)$	2.44	0.047894	0.045908	0.001986
3	$0.33 \cdot N(8.0, 4.0) + 0.67 \cdot N(-6.0, 6.25)$	6.98	0.033811	0.033967	−0.000156

[a]The rank order property says $\Pr(A_1 \mid X; f^0, S) - \Pr(A_2 \mid X; f^0, S) = 0$ for any S. Total match production is $f(X(u, \{d\})) + \varepsilon_{\langle u,d\rangle}$, 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.

stable assignment, and deviation by agents in A_2 is more attractive in an ease metric (two matched pairs could exchange partners, leaving the third pair alone), which provides a more compelling test against the idea that the rank order property holds approximately.

Table 1 considers six distributions S for i.i.d. match-specific unobservables. The table uses a simulation of the integral in (15) to compute $\Pr(A_1 \mid X; f, S) - \Pr(A_2 \mid X; f, S)$, a measure of how far off the rank order property is. The first line considers a standard normal distribution. As the variance is small, and both A_1 and A_2 are not pairwise stable assignments in the deterministic game, the assignment probabilities are individually small. However, the difference $\Pr(A_1 \mid X; f, S) - \Pr(A_2 \mid X; f, S) = -0.01514$ is large relative to the magnitudes. The second line increases the normal standard deviation to 6. Both assignment probabilities increase to around 0.079, but the difference $\Pr(A_1 \mid X; f, S) - \Pr(A_2 \mid X; f, S)$ decreases in absolute value to 0.00009. The third line increases the standard deviation to 20; now the probabilities are around 0.065, although $\Pr(A_1 \mid X; f, S) - \Pr(A_2 \mid X; f, S)$ remains small, at 0.00011.

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

Table 1 implies that assignment probabilities that differ by exchanges of only one downstream firm for each upstream firm are nearly rank ordered by their deterministic payoffs when true payoffs include match-specific stochastic components. In Fox

(2010), I introduced a maximum score estimator and presented 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.

4.5 *Multiple equilibrium assignments*

As argued in Section 3.3, games with complementarities across multiple downstream partners for the same upstream firm can have multiple equilibrium assignments with positive probability. In a game with multiple equilibrium assignments, (1) becomes

$$\begin{aligned}\Pr(A \mid X, Q; f, S) = \int_\psi & 1[A \text{ selected assignment} \mid A \text{ stable}, X, Q, \psi] \\ & \times 1[A \text{ pairwise stable} \mid X, Q, \psi]\, dS(\psi).\end{aligned} \tag{16}$$

Let $Y(A \mid X, Q; f, S)$ be equal to $\int_\psi 1[A \text{ pairwise stable} \mid X, Q, \psi]\, dS(\psi)$. Define A_1 and A_2 as in Property 4.1. For a model with multiple equilibrium assignments, the rank order property (Property 4.1) will hold under the following conditions: (i) $Y(A_1 \mid X, Q; f, S) > Y(A_2 \mid X, Q; f, S)$ if and only if inequality (13) holds, and (ii) $\Pr(A_1 \mid X, Q; f, S) > \Pr(A_2 \mid X, Q; f, S)$ if and only if $Y(A_1 \mid X, Q; f, S) > Y(A_2 \mid X, Q; f, S)$. Part (i) says A_1 will be 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 (ii) says the equilibrium-assignment selection rule preserves the rank ordering of pairwise stability: assignments that are more likely to be pairwise stable are more likely to occur. These conditions together imply Property 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 ψ and X combination.

In the literature on estimating Nash games, a nonnested class with matching games, some researchers assume a particular selection rule when such a rule is easy to implement (Jia (2008)); other researchers 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.[24]

5. Derivative-based nonparametric identification

In this section, I prove theorems about the identification of features of production functions in many-to-many matching games, following Definition 2.1. The intuition for some of these results was given in Section 2.5 for the case of one-to-one matching. Here, I focus on stating general theorems precisely.

[24]The literature on estimating parametric Nash games, a nonnested class with matching games, presents strategies with perhaps fewer assumptions but higher computational demands in estimation for dealing with multiple equilibria. See Bajari, Hong, and Ryan (2010) and Ciliberto and Tamer (2009). The nonparametric identification of other model components has not been studied while simultaneously employing these methods.

As mentioned before, I will explore identification with market-level data on (A, X). Let (A, X) be i.i.d. across matching markets. With these data, I can identify both $\Pr(A \mid X)$ and $G(X)$, the distribution of X across markets. I maintain the following assumption for derivative-based identification.

ASSUMPTION 5.1.

(i) *Each $f \in \mathcal{F}$ is three-times differentiable in all of its arguments.*

(ii) *X has support equal to the product of the marginal supports of the scalar elements of the vectors that comprise X. Each scalar element that derivatives are taken with respect to 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.[25]

The features of the production functions that govern sorting depend on how the characteristics that enter $\vec{x}(u, D_u)$ vary. I will present results where characteristics vary at the levels of the firm u or d, the individual match $\langle u, d\rangle$, and the group (u, D_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.

5.1 *Derivative-based identification with firm-specific characteristics*

First I consider firm-specific characteristics.

THEOREM 5.2. *Let $\vec{x}$ be a given point of evaluation of f. Let x_1 and x_2 be scalar characteristics in $\vec{x}$ from two different firms, either one upstream firm and one downstream firm or two downstream firms. Assume $\frac{\partial^2 f^0(\vec{x})}{\partial x_1\,\partial x_2} \neq 0$. Then the sign of $\frac{\partial^2 f^0(\vec{x})}{\partial x_1\,\partial x_2}$ is identified.*

The theorem is stated for a given point of evaluation $\vec{x}$ for clarity.[26] 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 the theorem was presented in Section 2.5.1. The next result follows.

THEOREM 5.3. *Let $\vec{x}$ be a given point of evaluation of f. Let x_1 and x_2 be scalar characteristics in $\vec{x}$ from two different firms, and let x_3 and x_4 be two scalar characteristics from two different firms as well. The identities of the firms in the two pairs (x_1, x_2) and (x_3, x_4)*

[25] While there are definitions such as increasing differences (Milgrom and Shannon (1994)) that encompass complementarities without relying on differentiable f's and continuous support for the x's, working with broader definitions makes the results harder to interpret and to compare to Becker's (1973).

[26] The assumption $\frac{\partial^2 f^0(\vec{x})}{\partial x_1\,\partial x_2} \neq 0$ greatly shortens the proof, which otherwise would require more complex limit arguments. This unfortunately rules out the interesting possibility that f^0 is a constant function, as then $\frac{\partial^2 f^0(\vec{x})}{\partial x_1\,\partial x_2} = 0$. However, the assumption that $\frac{\partial^2 f^0(\vec{x})}{\partial x_1\,\partial x_2} \neq 0$ was not made to rule out constant functions. Indeed, the rank order property ensures (if f^0 is a constant) that all assignments with the same set of nonsingle agents occur with equal probabilities. Allowing for f^0 to be a constant function is equivalent to allowing for 0 parameter values in the single-agent, multinomial logit: all choices have the same probabilities.

can be the same or not. Assume $\frac{\partial^2 f^0(\vec{x})}{\partial x_1 \partial x_2} \neq 0$ *and* $\frac{\partial^2 f^0(\vec{x})}{\partial x_3 \partial x_4} \neq 0$. *Then the ratio* $\frac{\partial^2 f^0(\vec{x})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\vec{x})}{\partial x_3 \partial x_4}$ *is identified.*

Section 2.5.2 presented the intuition for this theorem.[27] This is perhaps the most important result on identification in this paper. The ratio of the degree of complementarities, $\frac{\partial^2 f^0(\vec{x})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\vec{x})}{\partial x_3 \partial x_4}$, can be identified from qualitative data on who matches with whom.

5.2 *Derivative-based identification with match-specific characteristics*

The characteristics in $\vec{x}(u, D_u)$ can be specific to the individual matches $\langle u, d\rangle$.[28] In this case, the feature of f that governs sorting is f's second derivatives.

Theorem 5.4. *Let* $\vec{x}$ *be a given point of evaluation of* f. *Let the scalar* x *be a match-specific element of* $\vec{x}$. *Assume* $\frac{\partial^2 f^0(\vec{x})}{\partial x^2} \neq 0$. *The sign of* $\frac{\partial^2 f^0(\vec{x})}{\partial x^2}$ *is identified.*

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

Theorem 5.5. *Let* $\vec{x}$ *be a given point of evaluation of* f. *Let the two scalars* x_1 *and* x_2 *be distinct match-specific elements of* $\vec{x}$, *corresponding to different matches. Assume* $\frac{\partial^2 f^0(\vec{x})}{\partial x_1^2} \neq 0$ *and* $\frac{\partial^2 f^0(\vec{x})}{\partial x_2^2} \neq 0$. *The ratio* $\frac{\partial^2 f^0(\vec{x})}{\partial x_1^2} / \frac{\partial^2 f^0(\vec{x})}{\partial x_2^2}$ *is identified.*

The theorems 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 5.5 to be applied, while x_3–x_6 could be firm-specific, requiring Theorem 5.3. The presence of the match-specific x_1 and x_2 does not invalidate applying Theorem 5.3 to x_3–x_6.

5.3 *Derivative-based identification with group-specific characteristics*

Characteristics can be specific to a group (u, D_u) of downstream firms and the upstream firm that the downstream firms match with. An example of using the estimator in this paper for the group-characteristic case is Fox and Bajari (2010), who modeled 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 used a measure like the gravity equation in international trade to create a proxy for these geographic complementarities.

[27] The assumptions $\frac{\partial^2 f^0(\vec{x})}{\partial x_1 \partial x_2} \neq 0$ and $\frac{\partial^2 f^0(\vec{x})}{\partial x_3 \partial x_4} \neq 0$ allow Theorem 5.2 to be used in the proof and avoid division by zero in the ratio $\frac{\partial^2 f^0(\vec{x})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\vec{x})}{\partial x_3 \partial x_4}$.

[28] Some might think firm-specific characteristics are a special case of match-specific characteristics. Firm-specific characteristics increase the difficulty of showing the identification of features of f^0 because the same firm characteristics must appear on the left and the right sides of a local production maximization inequality. By contrast, hypothetical markets exist where the match-specific characteristics for the matches $\langle u_1, d_1\rangle$ and $\langle u_2, d_2\rangle$ may take on any pair of values, under Assumption 5.1.

Theorem 5.6. *Let $\vec{x}$ be a given point of evaluation of f and let x be a group-specific element of $\vec{x}$. Assume $\frac{\partial^2 f^0(\vec{x})}{\partial x^2} \neq 0$. The sign of $\frac{\partial^2 f^0(\vec{x})}{\partial x^2}$ is identified.*

The proof of this theorem is omitted because the mathematical argument is nearly identical to the proof of Theorem 5.4.

Theorem 5.7. *Let $\vec{x}$ be a given point of evaluation of f, and let the two scalars x_1 and x_2 be distinct group-specific elements of $\vec{x}$. Assume $\frac{\partial^2 f^0(\vec{x})}{\partial x_1^2} \neq 0$ and $\frac{\partial^2 f^0(\vec{x})}{\partial x_2^2} \neq 0$. The ratio $\frac{\partial^2 f^0(\vec{x})}{\partial x_1^2} / \frac{\partial^2 f^0(\vec{x})}{\partial x_2^2}$ is identified.*

Likewise, the proof is omitted because it is nearly identical to that of Theorem 5.5.

6. Nonparametric identification up to a positive monotonic transformation

Section 2.6 motivated why identifying production functions up to a positive monotonic transformation is distinct from derivative-based identification. Here I focus on precise statements of general theorems.

6.1 *Preliminaries for identification up to a positive monotonic transformation*

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 f^1 and f^2 of this class $\mathcal{F}$, 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) presented 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, $g(x_1 \mid \vec{x}_{-1}, X\backslash\vec{x})$, has an everywhere positive density in $\mathbb{R}$.*[29]
- *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.*

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

- *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 $\vec{x}$ with continuous support.*

The assumption allows all but one of the characteristics in $\vec{x}$ to have discrete or qualitative support.[30] This assumption replaces the earlier Assumption 5.1, which is only for the derivative-based identification theorems.[31]

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 where f is learned up to a positive monotonic transformation.

LEMMA 6.3. *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 $\vec{x}_a$ and $\vec{x}_b$ such that either*

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

or

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

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

6.2 *Theorems for identification up to a positive monotonic transformation*

Identification proofs in the single-agent maximum score tradition (Matzkin (1993, Theorem 1)) typically amount mathematically to Lemmas 4.1 and 6.3. Consequently, the identification proof for each case focuses on an issue that is new to matching games: embedding the inequalities from Lemma 6.3 in a local production maximization inequality, meaning an inequality where each upstream firm switches at most one downstream firm at a time. Thus, the proofs look for market characteristics X where the comparisons in Lemma 6.3 are decisive in rank ordering the production of two larger,

[30] The assumption that the support of x_1 is $\mathbb{R}$, rather than some compact subset of $\mathbb{R}$, is made for convenience. Manski (1988) and Horowitz (1998) showed 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.

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

otherwise similar assignments, A_1 and A_2. I consider group-, match-, and firm-specific characteristics separately. Identification is Definition 2.1 subject to the lack of a positive monotonic transformation in Assumption 6.1. I list the theorems in increasing difficulty of the proofs.

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

Theorem 6.4. *Let all elements of each $\vec{x}$ be group-specific. 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 6.3 and be able to construct local production maximization inequalities, where the coalition characteristics differ by the arguments corresponding to only one match between an upstream and a downstream firm. Remember, the production function allows a vector of arguments for each match of an upstream firm. To apply this and the following theorem, the maximum quota of an upstream firm must be known.

Theorem 6.5. *Let all elements of each $\vec{x}$ be match-specific. Also, let there be assignments A that contain as many groups of matches as the maximum quota of an upstream firm. Then 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 to identify a production function that involves the characteristics of the matches involving many downstream firms may be to use a solution concept such as full stability. The full stability 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 full stability is a very strong equilibrium concept. Fully stable outcomes are less likely to exist if preferences exhibit complementarities across multiple matches, and believing that a many-to-many outcome is fully stable 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 2.6, the value of being unmatched will be 0. Informally, identification considers the probabilities of assignments where certain firms are unmatched.[32] Firms that are more likely to be unmatched in an assignment are likely to have lower contributions to production.

THEOREM 6.6. *Let all the elements of each $\vec{x}$ be firm-specific. Let the value of any firm remaining unmatched be* 0, *or $f(\vec{x}_u^{\text{up}}) = f(\vec{x}_d^{\text{down}}) = 0$ for all $u \in U$, $d \in D$, and $f \in \mathcal{F}$. Furthermore, 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 $\mathcal{F}$.*

Data on single or unmatched agents are often available in marriage applications. Likewise, in an analysis of mergers or business alliances, single firms could be those not participating in a merger or business alliance. In an auction, bidders who do not win any items or items that are unsold are often observed.

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 and many-to-many games with substitutable preferences across multiple downstream firm partners, this condition is related to social efficiency. 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 derivative-based features of production functions and the identification of production functions up to a positive monotonic transformation. The derivative-based theorems generalize the informal identification results 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 or substitutes over their entire supports. One can identify whether any two inputs are complements or substitutes. Importantly, one can identify the value of the ratio of complementarities for two pairs of inputs at any point. Therefore, a researcher can identify the relative importance of different pairs of characteristics in match production.

The results on the identification of production functions up to a positive monotonic transformation extend the single-agent work of Matzkin (1993, Theorem 1) to matching games, where agents cannot unilaterally choose partners and so identification requires working with the equilibrium structure of the game. Researchers can distinguish between $f^1(m^1, w^1) = -(m^1 - w^1)^2$, the case of horizontal attributes, and $f^2(m^1, w^1) = 2m^1 \cdot w^1$, the case of vertical attributes.

[32] Choo and Siow (2006) estimated 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.

Appendix: Proofs

A.1 *Lemma 3.3: Pairwise stability implies local production maximization*

Substitute $\tilde{t}_{\langle u_1,d_2\rangle}$ into (8) and cancel the transfers $\sum_{d\in D_{u_1}(A)\setminus\{d_1\}} t_{\langle u_1,d\rangle}$ to give

$$\begin{aligned}
&v^{\text{up}}\big(\vec{x}(u_1, D_{u_1}(A))\big) + t_{\langle u_1,d_1\rangle} \\
&\quad\geq v^{\text{up}}\big(\vec{x}\big(u_1, (D_{u_1}(A)\setminus\{d_1\}) \cup \{d_2\}\big)\big) + v^{\text{down}}\big(\vec{x}(u_1, \{d_2\})\big) \\
&\qquad - \big(v^{\text{down}}\big(\vec{x}(u_2, \{d_2\})\big) - t_{\langle u_2,d_2\rangle}\big).
\end{aligned}$$

Call this no-deviation inequality $\text{nd}(\langle u_1, d_1\rangle, \langle u_2, d_2\rangle, A, t_{\langle u_1,d_1\rangle}, t_{\langle u_2,d_2\rangle})$. If $\pi\langle u, d\rangle$ is the new downstream partner of u in the permutation, let $\tilde{u}(\pi\langle u, d\rangle)$ be a function that gives the original partner of $\pi\langle u, d\rangle$, in B_1. So $\langle \tilde{u}(\pi\langle u, d\rangle), \pi\langle u, d\rangle\rangle \in B_1$. Now form $\sum_{\langle u,d\rangle\in B_1} \text{nd}(\langle u, d\rangle, \langle \tilde{u}(\pi\langle u, d\rangle), \pi\langle u, d\rangle\rangle, A, t_{\langle u,d\rangle}, t_{\langle \tilde{u}(\pi\langle u,d\rangle),\pi\langle u,d\rangle\rangle})$. This gives

$$\begin{aligned}
&\sum_{\langle u,d\rangle\in B_1} v^{\text{up}}\big(\vec{x}(u, D_u(A))\big) + \sum_{\langle u,d\rangle\in B_1} t_{\langle u,d\rangle} \\
&\quad\geq \sum_{\langle u,d\rangle\in B_1} v^{\text{up}}\big(\vec{x}\big(u, (D_u(A)\setminus\{d\}) \cup \{\pi\langle u, d\rangle\}\big)\big) \\
&\qquad + \sum_{\langle u,d\rangle\in B_1} \big\{v^{\text{down}}\big(\vec{x}\big(u, \{\pi\langle u, d\rangle\}\big)\big) \\
&\qquad - \big(v^{\text{down}}\big(\vec{x}\big(\tilde{u}(\pi\langle u, d\rangle), \{\pi\langle u, d\rangle\}\big)\big) - t_{\langle \tilde{u}(\pi\langle u,d\rangle),\pi\langle u,d\rangle\rangle}\big)\big\}.
\end{aligned}$$

By the definition of a permutation, each $t_{\langle u,d\rangle}$ for $\langle u, d\rangle \in B_1$ appears on both the left and right sides. The transfers cancel. Similarly, each equilibrium $v^{\text{down}}(\vec{x}(\tilde{u}(\pi\langle u, d\rangle), \{\pi\langle u, d\rangle\}))$ appears on the right side with a negative sign and each deviation $v^{\text{down}}(\vec{x}(u, \{\pi\langle u, d\rangle\}))$ appears on the right side with a positive sign. Moving $\sum_{\langle u,d\rangle\in B_1} v^{\text{down}}(\vec{x}(\tilde{u}(\pi\langle u, d\rangle), \{\pi\langle u, d\rangle\}))$ 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 5.2: Identification of the sign of a cross-partial derivative with firm-specific covariates*

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

For an arbitrary $f^1 \in \mathcal{F}$, $f^1 \neq f^0$, where $\frac{\partial^2 f^1(\vec{y})}{\partial x_1\,\partial x_2} < 0$, by Definition 2.1, 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 $\Pr(A \mid X; f^0, S^0) = \Pr(A \mid X; f^1, S^1)$ 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 3.4. At each of these markets X, there will be a particular assignment A_1 and another assignment A_2 where, by Property 4.1, $\Pr(A_1 \mid X; f^0, S^0) > \Pr(A_2 \mid X; f^0, S^0)$ while $\Pr(A_1 \mid X; f^1, S^1) < \Pr(A_2 \mid X; f^1, S^1)$ for any $S_1 \in \mathcal{S}$. Therefore, the conditions of Definition 2.1 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 $\Pr(A_1 \mid X; f^0, S^0) > \Pr(A_2 \mid X; f^0, S^0)$ while $\Pr(A_1 \mid X; f^1, S^1) < \Pr(A_2 \mid X; f^1, S^1)$ for any $S_1 \in \mathcal{S}$.

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 kth position. Without loss of generality, let x_1 be the first position of $\vec{x}$ and let x_2 be the second position. One definition of a cross-partial derivative is the limit of the middle difference quotient:

$$\frac{\partial^2 f(\vec{y})}{\partial x_1\,\partial x_2} = \lim_{h\to 0} \frac{f(\vec{y}+he_1+he_2) - f(\vec{y}+he_1) - f(\vec{y}+he_2) + f(\vec{y})}{h^2}. \tag{17}$$

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

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

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

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

or

$$f^0(\vec{y}+h^0e_1+h^0e_2) + f^0(\vec{y}) > f^0(\vec{y}+h^0e_1) + f^0(\vec{y}+h^0e_2). \tag{18}$$

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

$$f^1(\vec{y}+h^1e_1+h^1e_2) + f^1(\vec{y}) < f^1(\vec{y}+h^1e_1) + f^1(\vec{y}+h^1e_2). \tag{19}$$

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 (18) and (19) hold for any such h.

Now let me argue that (18), and by a similar argument (19), is a local production maximization inequality: it satisfies Definition 3.4. 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 (18). Let $B_1 = \{\langle u_1, d_1\rangle, \langle u_2, d_2\rangle\}$ and $B_2 = \{\langle u_1, d_2\rangle, \langle u_2, d_1\rangle\}$, where these indices refer to arbitrary firms I am creating to show (18) satisfies Definition 3.4. Also, let there be some D_{u_1} and D_{u_2} of sufficient size to reproduce the number of nonempty elements (representing unfilled matches) of $\vec{y}$. We require $d_1 \in D_{u_1}$, $d_2 \notin D_{u_2}$, $d_2 \in D_{u_2}$, and $d_2 \notin D_{u_1}$. Then define $\vec{x}(u_1, D_{u_1}) = \vec{y} + he_1 + he_2$, $\vec{x}(u_2, D_{u_2}) = \vec{y}$, $\vec{x}(u_1, (D_{u_1}\backslash\{d_1\}) \cup \{d_2\}) =$

$\vec{y} + he_1$, and $\vec{x}(u_2, (D_{u_2}\backslash\{d_2\}) \cup \{d_1\}) = \vec{y} + he_2$. With $\pi\langle u_1, d_1\rangle = d_2$ and $\pi\langle u_2, d_2\rangle = d_1$, inspection shows (18) satisfies Definition 3.4.

To complete the argument that (18) satisfies Definition 3.4 for the case of firm-specific characteristics, it is necessary to show that this exchange of partners can be accomplished with firm-specific characteristics, as 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. Let downstream firms d_1 and d_2 have the same baseline characteristics, except that firm d_1 has h more of characteristic x_2 than firm d_2: $x^{\text{down}}_{d_1,2} - h = x^{\text{down}}_{d_2,2}$, where $x^{\text{down}}_{d,2}$ is characteristic x_2 for firm $d \in D$.[33] On the left of (18), the match $\langle u_1, d_1\rangle$ puts d_1 in either a direct partnership with an upstream firm u_1 with h more x_1 than u_2 or an indirect partnership with another downstream firm $d_3 \in D_{u_1}$ with h more x_1 than the corresponding $d_4 \in D_{u_2}$. In notation, either $x^{\text{up}}_{u_1,1} - h = x^{\text{up}}_{u_2,1}$ or $x^{\text{down}}_{d_3,1} - h = x^{\text{down}}_{d_4,1}$.

The matches $\langle u_1, d_2\rangle$ and $\langle u_2, d_1\rangle$ form on the right side of (18). Firm d_1, with h more of x_2, is transferred from the set of matches (u_1, D_{u_1}), with h more x_1, to the set of matches (u_2, D_{u_2}). Likewise, d_2, without any more x_2, matches to u_1 and its downstream firm partners, which together have h more x_1 than the matches involving u_2. The important requirement that is satisfied is that each move switches the characteristics of only the firm that is actually switching. Therefore, (18) satisfies Definition 3.4 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 $D_{u_1}\backslash\{d_1\}$ and $D_{u_2}\backslash\{d_2\}$. The exact choice of B_3 plays no role in the proof, other than to ensure $D_{u_1}\backslash\{d_1\}$ and $D_{u_2}\backslash\{d_2\}$ are large enough given the number of nonempty elements (representing filled quota slots) in $\vec{y}$. Then set $A_1 = B_1 \cup B_3$ and $A_2 = B_2 \cup B_3$. Let there be some collection X of characteristics as in (7). The choice of X plays no role in the proof, except that characteristics itemized above must be a subset of X.

Property 4.1 and (18) imply $\Pr(A_1 \mid X; f^0, S^0) > \Pr(A_2 \mid X; f^0, S^0)$, while Property 4.1 and (19) imply $\Pr(A_1 \mid X; f^1, S^1) < \Pr(A_2 \mid X; f^1, S^1)$ for any $S_1 \in \mathcal{S}$.

The third step of the proof is to vary all the arguments in X to show that a continuum of market characteristics give different local production maximization inequalities for f^0 and f^1. 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. Each f is continuous by Assumption 5.1, so X can be varied locally and identification can be achieved over a set of markets with positive probability.

[33] If x_2 is the second characteristic of $\vec{x}$, I then say it is also the second characteristic of $\vec{x}^{\text{down}}_d$, $x^{\text{down}}_{d,2}$. This is done for clarity: it keeps "2" referring to the same variable whether I am referring to it as an element of the entire vector of production function arguments $\vec{x}$ or as an element of the vector of characteristics of firm d, $\vec{x}^{\text{down}}_d$.

A.3 *Theorem 5.3: Identification of the ratio of two cross-partial derivatives with firm-specific covariates*

We are given a point $\vec{y}$ (relabeled from $\vec{x}$ in the statement of the theorem) and there is an arbitrary $f^1 \in \mathcal{F}$, where $\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\vec{y})}{\partial x_3 \partial x_4} \neq \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\vec{y})}{\partial x_3 \partial x_4}$. The goal in broad generality is the same as the proof of Theorem 5.2: 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 $S^1 \in \mathcal{S}$. The proof is more challenging than the proof of Theorem 5.2 because now we are trying to identify the value of some feature of f^0, here $\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\vec{y})}{\partial x_3 \partial x_4}$, rather than just the sign of a cross-partial derivative, as before.

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

By Theorem 5.2, we know the signs of $\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2}$ and $\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_3}$ if they are nonzero, as the current theorem requires. If f^1 implies different signs for $\frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_2}$ and $\frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_3}$, then by Theorem 5.2, we can distinguish f^0 and f^1. So we can restrict attention to the case where the signs of $\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2}$ and $\frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_2}$ as well as $\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_3}$ and $\frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_3}$ are the same. I will first consider the case where $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_2} > 0$ and $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_3} > 0$ 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, (17), 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 the proof of Theorem 5.2 and are omitted for brevity.

Let $h_{1,2}$ be the limit argument from the middle-difference quotient, (17), for $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_2}$. Likewise, let $h_{1,3}$ be the limit argument for $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_3}$. Consider the case $\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_3} > \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_3}$ and let $\{h_{1,2,n}\}_{n \in \mathbb{N}}$ be a sequence that converges to 0. Let $\{h_{1,3,n}\}_{n \in \mathbb{N}}$ be a sequence

$$h_{1,3,n} = h_{1,2,n} \sqrt{\frac{1}{2}\left(\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2} \bigg/ \frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_3} + \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_2} \bigg/ \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_3}\right)}. \tag{20}$$

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

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

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

$$\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2} \Big/ \frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_3} > \frac{h^2_{1,3,n}}{h^2_{1,2,n}} > \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_2} \Big/ \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_3} \tag{21}$$

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

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

$$\begin{aligned} &Y(h_{1,2}, h_{1,3}; f) \\ &\quad = \frac{f(\vec{y} + h_{1,2}e_1 + h_{1,2}e_2) - f(\vec{y} + h_{1,2}e_1) - f(\vec{y} + h_{1,2}e_2) + f(\vec{y})}{h^2_{1,2}} \\ &\qquad \times \left(\frac{f(\vec{y} + h_{1,3}e_1 + h_{1,3}e_3) - f(\vec{y} + h_{1,3}e_1) - f(\vec{y} + h_{1,3}e_3) + f(\vec{y})}{h^2_{1,3}} \right)^{-1} \end{aligned} \tag{22}$$

for $f \in \mathcal{F}$. By a definition of a cross-partial derivative, (17), the ratio $Y(h_{1,2,n}, h_{1,3,n}; f)$ converges to $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_3}$ for $f \in \{f^0, f^1\}$ as $n \to \infty$ and $(h_{1,2,n}, h_{1,3,n}) \to (0, 0)$. Then there exists some $n_1 \in \mathbb{N}$, where, for all $n \geq n_1$, $n \in \mathbb{N}$, $|Y(h_{1,2,n}, h_{1,3,n}; f^0) - \frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_3}| < \frac{\tau}{3}$ and $|Y(h_{1,2,n}, h_{1,3,n}; f^1) - \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_3}| < \frac{\tau}{3}$. The choice of distance $\frac{\tau}{3}$ ensures that

$$Y(h_{1,2,n}, h_{1,3,n}; f^0) > \frac{h^2_{1,3,n}}{h^2_{1,2,n}} > Y(h_{1,2,n}, h_{1,3,n}; f^1) \tag{23}$$

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

$$\begin{aligned} &\Delta(h_{1,2}, h_{1,3}; f) \\ &\quad = f(\vec{y} + h_{1,2}e_1 + h_{1,2}e_2) - f(\vec{y} + h_{1,2}e_1) - f(\vec{y} + h_{1,2}e_2) + f(\vec{y}) \\ &\qquad - (f(\vec{y} + h_{1,3}e_1 + h_{1,3}e_3) - f(\vec{y} + h_{1,3}e_1) - f(\vec{y} + h_{1,3}e_3) + f(\vec{y})) \end{aligned}$$

for $f \in \mathcal{F}$. Choose $(h_{1,2}, h_{1,3}) = (h_{1,2,n}, h_{1,3,n})$. Substituting the definition of $Y(h_{1,2,n}, h_{1,3,n}; f)$ into (23) and resulting algebra shows that, at $(h_{1,2}, h_{1,3})$, the ratios $h^2_{1,3,n}/h^2_{1,2,n}$ cancel in all terms and

$$\begin{aligned} &\frac{f^0(\vec{y} + h_{1,2}e_1 + h_{1,2}e_2) - f^0(\vec{y} + h_{1,2}e_1) - f^0(\vec{y} + h_{1,2}e_2) + f^0(\vec{y})}{f^0(\vec{y} + h_{1,3}e_1 + h_{1,3}e_3) - f^0(\vec{y} + h_{1,3}e_1) - f^0(\vec{y} + h_{1,3}e_3) + f^0(\vec{y})} \\ &\quad > 1 > \frac{f^1(\vec{y} + h_{1,2}e_1 + h_{1,2}e_2) - f^1(\vec{y} + h_{1,2}e_1) - f^1(\vec{y} + h_{1,2}e_2) + f^1(\vec{y})}{f^1(\vec{y} + h_{1,3}e_1 + h_{1,3}e_3) - f^1(\vec{y} + h_{1,3}e_1) - f^1(\vec{y} + h_{1,3}e_3) + f^1(\vec{y})}, \end{aligned} \tag{24}$$

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 $h_{1,3,n}$, (20), will apply to the case $\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_3} < \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_3}$. Only a few inequalities are reversed.

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

$$\begin{aligned} &f(\vec{y} + h_{1,2}e_1 + h_{1,2}e_2) + f(\vec{y} + h_{1,3}e_1) + f(\vec{y} + h_{1,3}e_3) + f(\vec{y}) \\ &\quad > f(\vec{y} + h_{1,3}e_1 + h_{1,3}e_3) + f(\vec{y} + h_{1,2}e_1) + f(\vec{y} + h_{1,2}e_2) + f(\vec{y}). \end{aligned} \tag{25}$$

The inequality (25) satisfies Definition 3.4 for some choice of B_1 and B_2. Let $B_1 = \{\langle u_1, d_1\rangle, \langle u_2, d_2\rangle, \langle u_3, d_3\rangle, \langle u_4, d_4\rangle\}$ and $B_2 = \{\langle u_1, d_4\rangle, \langle u_2, d_3\rangle, \langle u_3, d_2\rangle, \langle u_4, d_1\rangle\}$, where the permutation π is implied by the definitions of B_1 and B_2. Also, let $d_1 \in D_{u_1}$, $d_2 \in D_{u_2}$, $d_3 \in D_{u_3}$, and $d_4 \in D_{u_4}$. Let $\vec{x}(u_1, D_{u_1}) = \vec{y} + h_{1,2}e_1 + h_{1,2}e_2$, $\vec{x}(u_2, D_2) = \vec{y} + h_{1,3}e_1$, $\vec{x}(u_3, D_3) = \vec{y} + h_{1,3}e_3$, $\vec{x}(u_4, D_4) = \vec{y}$, $\vec{x}(u_1, (D_{u_1}\backslash\{d_1\}) \cup \{d_4\}) = \vec{y} + h_{1,2}e_1$, $\vec{x}(u_2, (D_{u_2}\backslash\{d_2\}) \cup \{d_3\}) = \vec{y} + h_{1,3}e_1 + h_{1,3}e_3$, $\vec{x}(u_3, \{D_{u_3}\backslash\{d_3\}\} \cup \{d_2\}) = \vec{y}$, and $\vec{x}(u_4, (D_{u_4}\backslash\{d_4\}) \cup \{d_1\}) = \vec{y} + h_{1,2}e_2$. Either $x^{\text{up}}_{u_1,1} - h_{1,2} = x^{\text{up}}_{u_4,1}$ or $x^{\text{down}}_{m_1,1} - h_{1,2} = x^{\text{down}}_{m_4,1}$ for two firms $m_1 \in D_{u_1}$, $m_1 \neq d_1$, and $m_4 \in D_{u_4}$, $m_4 \neq d_4$; $x^{\text{down}}_{d_3,2} - h_{1,3} = x^{\text{down}}_{d_2,2}$; and either $x^{\text{up}}_{u_2,1} - h_{1,3} = x^{\text{up}}_{u_3,1}$ or $x^{\text{down}}_{m_2,1} - h_{1,3} = x^{\text{down}}_{m_3,1}$ for two firms $m_2 \in D_{u_2}$, $m_2 \neq d_2$, and $m_3 \in D_{u_3}$, $m_3 \neq d_3$. By inspection, it can be seen that each match in B_1 exchanges a downstream firm partner for a match in B_2. Meanwhile, each set of arguments $\vec{x}(u, D_u)$ on the right can be formed by an exchange of a single downstream firm's characteristics from a set of arguments on the left. Therefore, this construction satisfies Definition 3.4 for the case of firm-specific characteristics.

As in the proof of Theorem 5.2, I can embed B_1 and B_2 into a larger matching market, construct A_1, A_2, and X, apply the rank order property, and then vary X in a continuum to find a positive probability of markets where f^0 and f^1 give different predictions by the rank order property.

Recall that by Theorem 5.2, 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 $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_2} > 0$, $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_3} > 0$ and $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_2} > \frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_3}$ for $f \in \{f^0, f^1\}$. $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_2} > \frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_3}$ is without loss of generality,[34] but $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_2} > 0$ and $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_3} > 0$ for $f \in \{f^0, f^1\}$ are conditions with some loss of generality. Now we need to argue that the above arguments go through for the other three cases: $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_2} < 0$ and $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_3} > 0$, $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_2} > 0$ and $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_3} < 0$, as well as $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_2} < 0$ and $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_3} < 0$. This is simple: in some of these new cases, key inequalities may reverse direction, but as $\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_3} \neq \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_3}$ for all cases, the same arguments as above

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

will show $(h_{1,2}, h_{1,3})$ can be chosen to lie in between $\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_3}$ and $\frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_3}$. Once this is done, the above can all be repeated without much change.[35]

A.4 *Theorem 5.4: Identification of the sign of a second own partial derivative with match-specific covariates*

For conciseness, certain steps of the proof will be replaced with references to similar arguments in Theorem 5.2. Some notation, such as $\vec{y}$, is also explained in Theorem 5.2. Let x in the statement of the theorem be x_1 inside the proof.

One definition of a second derivative is

$$\frac{\partial^2 f(\vec{y})}{\partial x_1^2} = \lim_{h \to 0} \frac{f(\vec{y} + 2he_1) - 2f(\vec{y} + he_1) + f(\vec{y})}{h^2}, \tag{26}$$

where, as in the proof of Theorem 5.2, $e_1 = (1, 0, 0, \ldots, 0)$ is a vector of 0's except in the first element. Because both $\frac{\partial^2 f^0(\vec{y})}{\partial x_1^2}$ and $\frac{\partial^2 f^1(\vec{y})}{\partial x_1^2}$ are limits, there will be some $h = \min\{h^0, h^1\}$, where h^0 and h^1 are the respective limit arguments for f^0 and f^1, where both

$$f^0(\vec{y} + 2he_1) + f^0(\vec{y}) > 2f^0(\vec{y} + he_1) \tag{27}$$

and

$$f^1(\vec{y} + 2he_1) + f^1(\vec{y}) < 2f^1(\vec{y} + he_1) \tag{28}$$

hold.

Now I will argue that (27) and, by the same argument, (28) are local production maximization equations, Definition 3.4. Let $B_1 = \{\langle u_1, d_1\rangle, \langle u_2, d_2\rangle\}$ and $B_2 = \{\langle u_1, d_2\rangle, \langle u_2, d_1\rangle\}$. Also let there be sufficiently large (to handle $\vec{y}$) sets D_{u_1} and D_{u_2}, where $d_1 \in D_{u_1}$, $d_1 \notin D_{u_2}$, $d_2 \in D_{u_2}$, and $d_2 \notin D_{u_1}$. Also define $\vec{x}(u_1, D_{u_1}) = \vec{y} + 2he_1$, $\vec{x}(u_2, D_{u_2}) = \vec{y}$, $\vec{x}(u_1, (D_{u_1} \backslash \{d_1\}) \cup \{d_2\}) = \vec{y} + he_1$, and $\vec{x}(u_2, (D_{u_2} \backslash \{d_2\}) \cup \{d_1\}) = \vec{y} + he_1$. Let the first characteristics for matches satisfy $x^{\text{match}}_{\langle u_1, d_1\rangle, 1} - h = x^{\text{match}}_{\langle u_2, d_1\rangle, 1} = x^{\text{match}}_{\langle u_1, d_2\rangle, 1}$ and $x^{\text{match}}_{\langle u_1, d_1\rangle, 1} - 2h = x^{\text{match}}_{\langle u_2, d_2\rangle, 1}$. With $\pi\langle u_1, d_1\rangle = d_2$ and $\pi\langle u_2, d_2\rangle = d_1$, inspection shows (18) satisfies Definition 3.4 for the case of match-specific characteristics.

The embedding of B_1 and B_2 in a matching market and finding a continuum of markets X with the key property both follow similar arguments in the proof of Theorem 5.2.

[35] If $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_3} < 0$ for $f \in \{f^0, f^1\}$, then (20) will involve the square root of a negative number. To fix this, let (20) involve the absolute values of $\frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f(\vec{y})}{\partial x_1 \partial x_3}$ for $f \in \{f^0, f^1\}$. For the case $\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_3} > \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_3}$, (30) will become $\frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^0(\vec{y})}{\partial x_1 \partial x_3} > -\frac{h^2_{1,3,n}}{h^2_{1,2,n}} > \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_2} / \frac{\partial^2 f^1(\vec{y})}{\partial x_1 \partial x_3}$. Following the steps of the algebra in the earlier argument, the 1 in (24) will be a -1 and the pair of inequalities in (24) will reverse directions once both sides are multiplied by the -1. A different local production maximization inequality will arise; otherwise the argument is similar to the earlier argument.

A.5 *Theorem 5.5: Identification of the ratio of two second own partial derivatives with match-specific covariates*

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

Let h_1 be the index for the approximation term on the right side of (26) for $\frac{\partial^2 f(\vec{y})}{\partial x_1^2}$ and let h_2 be the index for $\frac{\partial^2 f(\vec{y})}{\partial x_2^2}$. Consider the case $\frac{\partial^2 f^0(\vec{y})}{\partial x_1^2} / \frac{\partial^2 f^0(\vec{y})}{\partial x_2^2} > \frac{\partial^2 f^1(\vec{y})}{\partial x_1^2} / \frac{\partial^2 f^1(\vec{y})}{\partial x_2^2}$ 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

$$h_{2,n} = h_{1,n}\sqrt{\frac{1}{2}\left(\frac{\partial^2 f^0(\vec{y})}{\partial x_1^2} \Big/ \frac{\partial^2 f^0(\vec{y})}{\partial x_2^2} + \frac{\partial^2 f^1(\vec{y})}{\partial x_1^2} \Big/ \frac{\partial^2 f^1(\vec{y})}{\partial x_2^2}\right)}. \tag{29}$$

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

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

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

$$\frac{\partial^2 f^0(\vec{y})}{\partial x_1^2} \Big/ \frac{\partial^2 f^0(\vec{y})}{\partial x_2^2} > \frac{h_{2,n}^2}{h_{1,n}^2} > \frac{\partial^2 f^1(\vec{y})}{\partial x_1^2} \Big/ \frac{\partial^2 f^1(\vec{y})}{\partial x_2^2} \tag{30}$$

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

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

$$\begin{aligned} Y(h_1, h_2; f) = {} & \frac{f(\vec{y} + 2h_1 e_1) - 2f(\vec{y} + h_1 e_1) + f(\vec{y})}{h_1^2} \\ & \times \left(\frac{f(\vec{y} + 2h_2 e_2) - 2f(\vec{y} + h_2 e_2) + f(\vec{y})}{h_2^2}\right)^{-1} \end{aligned} \tag{31}$$

for $f \in \mathcal{F}$. By the definition of a second partial derivative, (26), the ratio $Y(h_1, h_2; f)$ converges to $\frac{\partial^2 f(\vec{y})}{\partial x_1^2} / \frac{\partial^2 f(\vec{y})}{\partial x_2^2}$ for $f \in \{f^0, f^1\}$ as $n \to \infty$ and $(h_{1,n}, h_{2,n}) \to (0, 0)$. Then there exists some $n_1 \in \mathbb{N}$, where, for all $n \geq n_1$, $n \in \mathbb{N}$, $|Y(h_1, h_2; f^0) - \frac{\partial^2 f^0(\vec{y})}{\partial x_1^2} / \frac{\partial^2 f^0(\vec{y})}{\partial x_2^2}| < \frac{\tau}{3}$ and $|Y(h_1, h_2; f^1) - \frac{\partial^2 f^1(\vec{y})}{\partial x_1^2} / \frac{\partial^2 f^1(\vec{y})}{\partial x_2^2}| < \frac{\tau}{3}$. The choice of distance $\frac{\tau}{3}$ ensures that

$$Y(h_1, h_2; f^0) > \frac{h_{2,n}^2}{h_{1,n}^2} > Y(h_1, h_2; f^1) \tag{32}$$

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

$$\begin{aligned} \Delta(h_1, h_2; f) = {} & f(\vec{y} + 2h_1 e_1) - 2f(\vec{y} + h_1 e_1) + f(\vec{y}) \\ & - (f(\vec{y} + 2h_2 e_2) - 2f(\vec{y} + h_2 e_2) + f(\vec{y})) \end{aligned}$$

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 $Y(h_1, h_2; f)$ into (32) and resulting algebra shows that at (h_1, h_2), the ratios $h_{2,n}^2 / h_{1,n}^2$ cancel in all terms and

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

and so

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

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 $\frac{\partial^2 f^0(\vec{y})}{\partial x_1^2} / \frac{\partial^2 f^0(\vec{y})}{\partial x_2^2} < \frac{\partial^2 f^1(\vec{y})}{\partial x_1^2} / \frac{\partial^2 f^1(\vec{y})}{\partial x_2^2}$. Only a few inequalities are reversed.

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

$$f(\vec{y}+2h_1e_1) + 2f(\vec{y}+h_2e_2) + f(\vec{y}) > f(\vec{y}+2h_2e_2) + 2f(\vec{y}+h_1e_1) + f(\vec{y}). \quad (33)$$

We can show that this is a local production maximization inequality, Definition 3.4, for some choice of B_1 and B_2. Let $B_1 = \{\langle u_1, d_1\rangle, \langle u_2, d_2\rangle, \langle u_3, d_3\rangle, \langle u_4, d_4\rangle\}$ and $B_2 = \{\langle u_1, d_4\rangle, \langle u_2, d_1\rangle, \langle u_3, d_2\rangle, \langle u_4, d_3\rangle\}$, where the permutation π is implied by the definitions of B_1 and B_2. Also, let $d_1 \in D_{u_1}$, $d_2 \in D_{u_2}$, $d_3 \in D_{u_3}$, and $d_4 \in D_{u_4}$. Let $\vec{x}(u_1, D_{u_1}) = \vec{y} + 2h_1e_1$, $\vec{x}(u_2, D_{u_2}) = \vec{y} + h_2e_2$, $\vec{x}(u_3, D_{u_3}) = \vec{y} + h_2e_2$, $\vec{x}(u_4, D_{u_4}) = \vec{y}$, $\vec{x}(u_1, (D_{u_1} \backslash \{d_1\}) \cup \{d_4\}) = \vec{y} + h_1e_1$, $\vec{x}(u_2, (D_{u_2} \backslash \{d_2\}) \cup \{d_1\}) = \vec{y} + h_1e_1$, $\vec{x}(u_3, \{D_{u_3} \backslash \{d_3\}\} \cup \{d_2\}) = \vec{y} + 2h_2e_2$, and $\vec{x}(u_4, (D_{u_4} \backslash \{d_4\}) \cup \{d_3\}) = \vec{y}$. Let $x^{\text{match}}_{\langle u_1, d_1\rangle,1} - 2h_1 = x^{\text{match}}_{\langle u_4, d_4\rangle,1} = x^{\text{match}}_{\langle u_3, d_2\rangle,1} = x^{\text{match}}_{\langle u_2, d_2\rangle,1} = x^{\text{match}}_{\langle u_3, d_3\rangle,1} = x^{\text{match}}_{\langle u_4, d_3\rangle,1}$; $x^{\text{match}}_{\langle u_1, d_1\rangle,1} - h_1 = x^{\text{match}}_{\langle u_2, d_1\rangle,1} = x^{\text{match}}_{\langle u_1, d_4\rangle,1}$; $x^{\text{match}}_{\langle u_3, d_2\rangle,2} - 2h_2 = x^{\text{match}}_{\langle u_1, d_1\rangle,2} = x^{\text{match}}_{\langle u_2, d_1\rangle,2} = x^{\text{match}}_{\langle u_4, d_3\rangle,2} = x^{\text{match}}_{\langle u_4, d_4\rangle,2} = x^{\text{match}}_{\langle u_1, d_4\rangle,2}$; and $x^{\text{match}}_{\langle u_3, d_2\rangle,2} - h_2 = x^{\text{match}}_{\langle u_2, d_2\rangle,2} = x^{\text{match}}_{\langle u_3, d_3\rangle,2}$. By inspection, it can be seen that each match in B_1 exchanges a downstream firm partner for a match in B_2. Meanwhile, each set of arguments on the right side can be formed by replacing the characteristics associated with a single match in a set of arguments on the left side. Therefore, Definition 3.4 is satisfied.

The remainder of the proof follows arguments in previous proofs and so is omitted.

A.6 *Lemma 6.3: Continuous characteristics for the identification of production functions up to positive monotonic transformations*

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

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

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

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

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

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

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

A.7 *Theorem 6.4: Identification up to a positive monotonic transformation, 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 $f^1(\vec{x}) \neq m \circ f^0(\vec{x})$ for all $\vec{x}$ for any positive monotonic function m. The goal is to show that 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 6.3 produces $\vec{x}_1$ and $\vec{x}_2$ such that $f^0(\vec{x}_1) > f^0(\vec{x}_2)$ and $f^1(\vec{x}_1) < f^1(\vec{x}_2)$ or $f^0(\vec{x}_1) < f^0(\vec{x}_2)$ and $f^1(\vec{x}_1) > f^1(\vec{x}_2)$. Focus on the first case. An inequality such as $f^0(\vec{x}_1) > f^0(\vec{x}_2)$ considers a group of matches centered around an upstream firm on the left and another group of matches centered around an upstream firm on the right. This is not a local production maximization inequality (Definition 3.4), which would require at least two groups, each centered on an upstream firm, on both the left and the right.

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

$$f(\vec{x}_1) + f(\vec{x}_3) > f(\vec{x}_2) + f(\vec{x}_3). \tag{34}$$

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 3.4. Let $B_1 = \{\langle u_1, d_1\rangle, \langle u_2, d_2\rangle\}$ and $B_2 = \{\langle u_1, d_2\rangle, \langle u_2, d_1\rangle\}$. Also let there be sufficiently large (to handle $\vec{y}$) sets D_{u_1} and D_{u_2}, where $d_1 \in D_{u_1}$, $d_1 \notin D_{u_2}$, $d_2 \in D_{u_2}$, and $d_2 \notin D_{u_1}$. Also define $\vec{x}(u_1, D_{u_1}) = \vec{x}_1$, $\vec{x}(u_2, D_{u_2}) = \vec{x}_3$, $\vec{x}(u_1, (D_{u_1}\backslash\{d_1\}) \cup \{d_2\}) = \vec{x}_2$, and $\vec{x}(u_2, (D_{u_2}\backslash\{d_2\}) \cup \{d_1\}) = \vec{x}_3$. With $\pi\langle u_1, d_1\rangle = d_2$ and $\pi\langle u_2, d_2\rangle = d_1$, inspection shows (34) satisfies Definition 3.4 for the case of match-specific characteristics.

The remainder of the proof uses quite similar arguments to those in Theorem 5.2. The four groups are embedded into a larger matching market. One can 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 $f^0(\vec{x}_1) < f^0(\vec{x}_2)$ and $f^1(\vec{x}_1) > f^1(\vec{x}_2)$ is similar: just reverse the local production maximization inequalities.

A.8 *Theorem 6.5: Identification up to a positive monotonic transformation, 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 $f^1(\vec{x}) \neq m \circ f^0(\vec{x})$ for all $\vec{x}$ and for any positive monotonic function m. Lemma 6.3 produces $\vec{x}_1$ and $\vec{x}_2$ such that $f^0(\vec{x}_1) > f^0(\vec{x}_2)$ and $f^1(\vec{x}_1) < f^1(\vec{x}_2)$ or $f^0(\vec{x}_1) < f^0(\vec{x}_2)$ and $f^1(\vec{x}_1) > f^1(\vec{x}_2)$. Focus on the first case. We will now construct a local production maximization inequality.

For match-specific characteristics, $\vec{x}(u, D_u) = \text{cat}((x^{\text{match}}_{\langle u,d_1\rangle,1}, \dots, x^{\text{match}}_{\langle u,d_1\rangle,K}), \dots, (x^{\text{match}}_{\langle u,d_m\rangle,1}, \dots, x^{\text{match}}_{\langle u,d_m\rangle,K}))$ for $D_u = \{d_1, \dots, d_m\}$, where $m = |D_u|$ is the number of downstream firms matched to upstream firm u and K is the number of scalar characteristics for each match. An alternative representation of $\vec{x}(u, D_u)$ is as a tuple of vectors rather than a concatenation of vectors (one long vector). For this proof only, let $\vec{x}_1 = (\vec{x}^{\text{match}}_{\langle 1,1\rangle}, \dots, \vec{x}^{\text{match}}_{\langle 1,m\rangle})$, where each $\vec{x}^{\text{match}}_{\langle 1,d\rangle}$ for $d = 1, \dots, m$ is itself potentially a vector. Likewise, let $\vec{x}_2 = (\vec{x}^{\text{match}}_{\langle 2,1\rangle}, \dots, \vec{x}^{\text{match}}_{\langle 2,n\rangle})$, where upstream firm 2 has n matches, each with a vector of characteristics. To further simplify notation, expand the shorter of the two characteristics collections $\vec{x}_1$ and $\vec{x}_2$ to have the same number of component matches by adding empty sets to the production vector. Call the common number of component matches $h = \max\{m, n\}$. If $m = 2$ and $n = 3$, $\vec{x}_1$ is expanded to be $(\vec{x}^{\text{match}}_{\langle 1,1\rangle}, \vec{x}^{\text{match}}_{\langle 1,2\rangle}, \emptyset)$.

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

$$f(\vec{x}_1) + \sum_{c=1}^{h-1} f(\vec{w}_c) > f(\vec{x}_2) + \sum_{c=1}^{h-1} f(\vec{w}_c). \tag{35}$$

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

A local production maximization inequality must satisfy Definition 3.4. 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 (35) is indeed a local production maximization inequality, we need to show that we can pick $\{\vec{w}_c\}_{c=1}^{h-1}$ so that each term on the right side is only one match-specific characteristic vector separate from a term on the left side. The general construction of a $\vec{w}_c$ for $c \leq h-1$ is

$$\vec{w}_c = (\vec{x}^{\text{match}}_{\langle 1,1\rangle}, \dots, \vec{x}^{\text{match}}_{\langle 1,h-c\rangle}, \vec{x}^{\text{match}}_{\langle 2,1\rangle}, \dots, \vec{x}^{\text{match}}_{\langle 2,c\rangle}).$$

The construction is motivated as follows. From Definition 3.4, let $B_1 = \{\langle u_0, d_0\rangle, \langle u_1, d_1\rangle, \dots, \langle u_{h-1}, d_{h-1}\rangle\}$ and $B_2 = \{\langle u_0, d_1\rangle, \langle u_1, d_2\rangle, \dots, \langle u_{h-1}, d_0\rangle\}$. The group centered around upstream firm u_0, with characteristics $\vec{x}(u_0, D_{u_0}) = \vec{x}_1$, replaces one downstream firm, $d_0 \in D_{u_0}$, with a new firm, $d_1 \in D_{u_1}$. A valid new match-specific value for u_0's new partner d_1 is, by intentional choice, $\vec{x}(u_0, \{d_1\}) = \vec{x}^{\text{match}}_{\langle 2,1\rangle}$, the first vector in $\vec{x}_2$.[36] This results in a group of matches with characteristics $\vec{w}_1 = \vec{x}(u_0, (D_{u_0}\backslash\{d_0\}) \cup \{d_1\}) = (\vec{x}^{\text{match}}_{\langle 1,1\rangle}, \dots, \vec{x}^{\text{match}}_{\langle 1,h-1\rangle}, \vec{x}^{\text{match}}_{\langle 2,1\rangle})$ appearing on the right side of (35). Recall that we need to add the same terms on the left and right sides to move from $f(\vec{x}_1) > f(\vec{x}_2)$ to (35). So we add $f(\vec{w}_1) = f(\vec{x}(u_1, D_{u_1}))$ on the left side. The group centered around upstream firm u_1 replaces one downstream firm, $d_1 \in D_{u_1}$, with $d_2 \in D_{u_2}$. On the right side, $\vec{w}_2 = \vec{x}(u_1, (D_{u_1}\backslash\{d_1\}) \cup \{d_2\}) = (\vec{x}^{\text{match}}_{\langle 1,1\rangle}, \dots, \vec{x}^{\text{match}}_{\langle 1,h-2\rangle}, \vec{x}^{\text{match}}_{\langle 2,1\rangle}, \vec{x}^{\text{match}}_{\langle 2,2\rangle})$. As before, $f(\vec{w}_2) = f(\vec{x}(u_2, D_{u_2}))$ appears on the left side as well.

This iterative process truncates. A hypothetical $\vec{w}_h$ equals $\vec{x}_2$, one of the original two vectors from the beginning of the proof. Also, $\vec{x}_1$ equals a hypothetical $\vec{w}_0$, the beginning of the iterative process. The above construction shows that each $\vec{x}(u_c, D_{u_c}) = \vec{w}_c$ on the left side exchanges one downstream firm d_c to yield $\vec{x}(u_c, (D_{u_c}\backslash\{d_c\}) \cup \{d_{c+1}\}) = \vec{w}_{c+1}$ on the right side. By inspection, each collection of characteristics $\vec{x}(u_c, (D_{u_c}\backslash\{d_c\}) \cup \{d_{c+1}\})$ is different from $\vec{x}(u_c, D_{u_c})$ by the characteristics of one match: $\vec{x}(u_c, \{d_{c+1}\}) = \vec{x}^{\text{match}}_{\langle 2,c+1\rangle}$ instead of $\vec{x}(u_c, \{d_c\}) = \vec{x}^{\text{match}}_{\langle 1,h-c\rangle}$. Therefore, (35) is a valid local production maximization inequality according to Definition 3.4.

The remainder of the proof follows arguments similar to those in previous proofs.

A.9 *Theorem 6.6: Identification up to a positive monotonic transformation, 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 $f^1(\vec{x}) \neq m \circ f^0(\vec{x})$ for all $\vec{x}$ and for any positive monotonic function m. Lemma 6.3 produces $\vec{x}_1$ and $\vec{x}_2$ such that $f^0(\vec{x}_1) > f^0(\vec{x}_2)$ and $f^1(\vec{x}_1) < f^1(\vec{x}_2)$ or $f^0(\vec{x}_1) < f^0(\vec{x}_2)$ and $f^1(\vec{x}_1) > f^1(\vec{x}_2)$. Focus on the first case. We will now construct a local production maximization inequality.

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

The characteristics are firm-specific: $\vec{x}(u, D_u) = \text{cat}((x^{\text{up}}_{u,1}, \dots, x^{\text{up}}_{u,K^{\text{up}}}), (x^{\text{down}}_{d_1,1}, \dots, x^{\text{down}}_{d_1,K^{\text{down}}}), \dots, (x^{\text{down}}_{d_l,1}, \dots, x^{\text{down}}_{d_l,K^{\text{down}}}))$, where u is matched to l downstream firms, each upstream firm has K^{up} characteristics, and each downstream firm has K^{down} characteristics.

In this proof only, I will use the notation $f(\vec{x}_1) = f(\vec{x}^{\text{up}}_1, \vec{x}^{\text{do},1}_1, \dots, \vec{x}^{\text{do},1}_l)$ to represent the production of a group of matches with firm characteristics $\vec{x}_1$. Here, $\vec{x}^{\text{up}}_1 = (x^{\text{up}}_{1,1}, \dots, x^{\text{up}}_{1,K^u})$, $\vec{x}^{\text{do},1}_d = (x^{\text{do}}_{d,1}, \dots, x^{\text{do}}_{d,K^d})$ and "do" is simply short for "downstream firm." Each argument of $f(\vec{x}^{\text{up}}_1, \vec{x}^{\text{do},1}_1, \dots, \vec{x}^{\text{do},1}_l)$ is a vector of firm-specific characteristics.

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

I put the 1 superscript on these downstream firms to remind us that their characteristics are part of $\vec{x}_1$. Also, let l be the maximum of the number of downstream firms whose characteristics are in $\vec{x}_1$ and $\vec{x}_2$; vectors of empty sets can be added as arguments if the numbers of downstream firms in $\vec{x}_1$ and in $\vec{x}_2$ are not equal. Altogether, $f(\vec{x}_1) = f(\vec{x}_1^{\text{up}}, \vec{x}_1^{\text{do},1}, \dots, \vec{x}_l^{\text{do},1})$ and $f(\vec{x}_2) = f(\vec{x}_2^{\text{up}}, \vec{x}_1^{\text{do},2}, \dots, \vec{x}_l^{\text{do},2})$.

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

$$\begin{aligned} f(\vec{x}_1) &+ \sum_{d=1}^{l-1} f(\vec{x}_1^{\text{up}}, \vec{x}_1^{\text{do},1}, \dots, \vec{x}_d^{\text{do},1}) + f(\vec{x}_1^{\text{up}}) + \sum_{d=1}^{l} f(\vec{x}_d^{\text{do},1}) \\ &+ \sum_{d=1}^{l-1} f(\vec{x}_2^{\text{up}}, \vec{x}_1^{\text{do},2}, \dots, \vec{x}_d^{\text{do},2}) + f(\vec{x}_2^{\text{up}}) + \sum_{d=1}^{l} f(\vec{x}_d^{\text{do},2}) \\ &> \sum_{d=1}^{l-1} f(\vec{x}_1^{\text{up}}, \vec{x}_1^{\text{do},1}, \dots, \vec{x}_d^{\text{do},1}) + f(\vec{x}_1^{\text{up}}) + \sum_{d=1}^{l} f(\vec{x}_d^{\text{do},1}) \\ &+ \sum_{d=1}^{l-1} f(\vec{x}_2^{\text{up}}, \vec{x}_1^{\text{do},2}, \dots, \vec{x}_d^{\text{do},2}) + f(\vec{x}_2^{\text{up}}) + \sum_{d=1}^{l} f(\vec{x}_d^{\text{do},2}) + f(\vec{x}_2). \end{aligned} \tag{36}$$

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

By inspection, one can loosely verify that (36) is almost, but not quite, a local production maximization inequality, Definition 3.4, with firm-specific characteristics. The term $\vec{x}_1$ on the left exchanges $\vec{x}_l^{\text{do},1}$ for the option of being unmatched, 0, to add $f(\vec{x}_1^{\text{up}}, \vec{x}_1^{\text{do},1}, \dots, \vec{x}_{l-1}^{\text{do},1}) + f(\vec{x}_l^{\text{do},1})$ on the right side. Following a pattern, each term $f(\vec{x}_1^{\text{up}}, \vec{x}_1^{\text{do},1}, \dots, \vec{x}_d^{\text{do},1})$ on the left side splits away the term $\vec{x}_d^{\text{do},1}$ to leave a $f(\vec{x}_1^{u}, \vec{x}_1^{\text{do},1}, \dots, \vec{x}_{d-1}^{\text{do},1}) + f(\vec{x}_d^{\text{do},1})$ on the right. Each term on the left involving the characteristics originally from $\vec{x}_2$, for example, $f(\vec{x}_2^{\text{up}}, \vec{x}_1^{\text{do},2}, \dots, \vec{x}_d^{\text{do},2})$, combines with an unmatched $\vec{x}_{d+1}^{\text{do},1}$ to form $f(\vec{x}_2^{\text{up}}, \vec{x}_1^{\text{do},2}, \dots, \vec{x}_{d+1}^{\text{do},2})$ on the right side.

The inequality (36) is not a local production maximization inequality. For example, look at the terms $f(\vec{x}_1^{\text{up}}) + \sum_{d=1}^{l} f(\vec{x}_d^{\text{do},1})$ on the left side. These unmatched firms do not combine with other firms to make pairings on the right side of (36). Therefore, as written, (36) is not a local production maximization inequality according to Definition 3.4. However, the statement of the theorem imposes a noninnocuous localization normalization, which gives $f(\vec{x}_1^{\text{up}}) + \sum_{d=1}^{l} f(\vec{x}_d^{\text{do},1}) = 0$ on the left and $f(\vec{x}_2^{\text{up}}) + \sum_{d=1}^{l} f(\vec{x}_d^{\text{do},2}) = 0$ on the right. With this change, (36) becomes

$$\begin{aligned} f(\vec{x}_1) &+ \sum_{d=1}^{l-1} f(\vec{x}_1^{\text{up}}, \vec{x}_1^{\text{do},1}, \dots, \vec{x}_d^{\text{do},1}) + \sum_{d=1}^{l-1} f(\vec{x}_2^{\text{up}}, \vec{x}_1^{\text{do},2}, \dots, \vec{x}_d^{\text{do},2}) \\ &+ f(\vec{x}_2^{\text{up}}) + \sum_{d=1}^{l} f(\vec{x}_d^{\text{do},2}) \end{aligned} \tag{37}$$

TABLE 2. Proof of Theorem 6.6: Demonstrating that (37) is a local production maximization inequality.

Index c	(1) $\vec{x}(u_c, \{0\})$	(2) $\vec{x}(u_c, D_{u_c})$	(3) $\vec{x}(0, \{d_c\})$	(4) $\pi\langle u_c, d_c\rangle$	(5) $\vec{x}(0, \{\pi\langle u_c, d_c\rangle\})$	(6) $\vec{x}(u_c, (D_{u_c}\backslash\{d_c\})\cup\{\pi\langle u_c, d_c\rangle\})$
1	$\vec{x}_1^{\text{up}}$	$\vec{x}_1$	$\vec{x}_l^{\text{do},1}$	d_{l+1}	$\emptyset$	$(\vec{x}_1^{\text{up}}, \vec{x}_1^{\text{do},1}, \dots, \vec{x}_{l-1}^{\text{do},1})$
2	$\vec{x}_1^{\text{up}}$	$(\vec{x}_1^{\text{up}}, \vec{x}_1^{\text{do},1})$	$\vec{x}_1^{\text{do},1}$	d_{l+2}	$\emptyset$	$(\vec{x}_1^{\text{up}})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
l	$\vec{x}_1^{\text{up}}$	$(\vec{x}_1^{\text{up}}, \vec{x}_1^{\text{do},1}, \dots, \vec{x}_{l-1}^{\text{do},1})$	$\vec{x}_{l-1}^{\text{do},1}$	d_{2l}	$\emptyset$	$(\vec{x}_1^{\text{up}}, \vec{x}_1^{\text{do},1}, \dots, \vec{x}_{l-2}^{\text{do},1})$
$l+1$	$\vec{x}_2^{\text{up}}$	$(\vec{x}_2^{\text{up}}, \vec{x}_1^{\text{do},2})$	$\emptyset$	d_{2l+2}	$\vec{x}_2^{\text{do},2}$	$(\vec{x}_2^{\text{up}}, \vec{x}_1^{\text{do},2}, \vec{x}_2^{\text{do},2})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2l-1$	$\vec{x}_2^{\text{up}}$	$(\vec{x}_2^{\text{up}}, \vec{x}_1^{\text{do},2}, \dots, \vec{x}_{l-1}^{\text{do},2})$	$\emptyset$	d_{3l}	$\vec{x}_l^{\text{do},2}$	$\vec{x}_2 = (\vec{x}_2^{\text{up}}, \vec{x}_1^{\text{do},2}, \dots, \vec{x}_l^{\text{do},2})$
$2l$	$\vec{x}_2^{\text{up}}$	$(\vec{x}_2^{\text{up}})$	$\emptyset$	d_{2l+1}	$\vec{x}_1^{\text{do},2}$	$(\vec{x}_2^{\text{up}}, \vec{x}_1^{\text{do},2})$
$2l+1$	$\emptyset$	$(\vec{x}_1^{\text{do},2})$	$\vec{x}_1^{\text{do},2}$	d_1	$\vec{x}_l^{\text{do},1}$	$(\vec{x}_l^{\text{do},1})$
$2l+2$	$\emptyset$	$(\vec{x}_2^{\text{do},2})$	$\vec{x}_2^{\text{do},2}$	d_2	$\vec{x}_1^{\text{do},1}$	$(\vec{x}_1^{\text{do},1})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$3l$	$\emptyset$	$(\vec{x}_l^{\text{do},2})$	$\vec{x}_l^{\text{do},2}$	d_l	$\vec{x}_{l-1}^{\text{do},1}$	$(\vec{x}_{l-1}^{\text{do},1})$

$$
\begin{aligned}
&> \sum_{d=1}^{l-1} f(\vec{x}_1^{\text{up}}, \vec{x}_1^{\text{do},1}, \dots, \vec{x}_d^{\text{do},1}) + f(\vec{x}_1^{\text{up}}) + \sum_{d=1}^{l} f(\vec{x}_d^{\text{do},1}) \\
&\quad + \sum_{d=1}^{l-1} f(\vec{x}_2^{\text{up}}, \vec{x}_1^{\text{do},2}, \dots, \vec{x}_d^{\text{do},2}) + f(\vec{x}_2),
\end{aligned}
$$

which, by the above informal arguments, is a local production maximization inequality. I intentionally do not remove from (36) all production functions with zero production. Even though the production $f(\vec{x}_1^{\text{up}})$ of singleton matches is zero, these production functions are needed to show that (37) satisfies the definition of a local production maximization inequality, Definition 3.4.

The above arguments were informal. I will now formally show that (37) satisfies Definition 3.4. There are $3l$ terms on the left side of (37). 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 $B_1 = \{\langle u_1, d_1\rangle, \dots, \langle u_{3l}, d_{3l}\rangle\}$, where the indexing $\langle u_c, d_c\rangle$ follows the order on the left side of (37), from left to right. As I will show, many of these match partners will be 0, representing being unmatched. Now let

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

The match $\langle u_1, d_{l+1}\rangle \in B_2$ means that the upstream firm u_1, which on the left side has characteristics $\vec{x}(u_1, D_{u_1}) = \vec{x}_1$, exchanges a downstream firm d_1 for the downstream firm $d_{l+1} \in D_{u_{l+1}}$. In this case, downstream firm d_1 has the characteristics $\vec{x}_l^{\text{do},1}$, while d_{l+1} is actually a dummy partner, 0, representing being unmatched. For each index $c = 1, \ldots, 3l$, Table 2 lists the upstream firm characteristics, the characteristics for the group of all firms u_c and D_{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 an upstream firm 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 column 5 are always the same as the characteristics of that downstream firm in column 3. This is the key idea behind showing that (37) 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 previous proofs.

References

Ackerberg, D. A. and M. Botticini (2002), "Endogenous matching and the empirical determinants of contract form." *Journal of Political Economy*, 110, 564–591. [211]

Akkus, O. and A. Hortacsu (2007), "The determinants of bank mergers: A revealed preference analysis." Working paper, University of Chicago. [206]

Andrews, D., S. T. Berry, and P. Jia (2004), "Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount store locations." Working paper, Yale University. [228]

Baccara, M., A. Imrohoroglu, A. Wilson, and L. Yariv (2009), "A field study on matching with network externalities." Working paper, New York University. [206]

Bajari, P., V. Chernozhukov, H. Hong, and D. Nekipelov (2009), "Identification and efficient semiparametric estimation of a dynamic discrete game." Working paper, University of Minnesota. [203]

Bajari, P., H. Hong, and S. Ryan (2010), "Identification and estimation of a discrete game of complete information." *Econometrica*, 78 (5), 1529–1568. [228, 230]

Becker, G. S. (1973), "A theory of marriage: Part I." *Journal of Political Economy*, 81 (4), 813–846. [203, 204, 207, 208, 209, 212, 231, 236]

Beresteanu, A., I. Molchanov, and F. Molinari (2008), "Sharp identification regions in games." Working paper, University of Pittsburgh. [228]

Berry, S. and E. Tamer (2006), "Identification in models of oligopoly entry." In *Advances in Economics and Econometrics: Theory and Applications: Ninth World Congress* (W. Newey, R. K. Blundell, and T. Persson, eds.), Chapter 2, p. 46, Cambridge University Press. [203]

Berry, S. T. (1992), "Estimation of a model of entry in the airline industry." *Econometrica*, 60 (4), 889–917. [228]

Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff (2003), "Analyzing the determinants of the matching public school teachers to jobs: Estimating compensating differentials in imperfect labor markets." Working paper, University of Buffalo. [206, 210]

Bresnahan, T. F. and P. C. Reiss (1991), "Entry and competition in concentrated markets." *The Journal of Political Economy*, 99 (5), 977–1009. [228]

Camiña, E. (2006), "A generalized assignment game." *Mathematical Social Sciences*, 52 (2), 152–161. [217]

Choo, E. and A. Siow (2006), "Who marries whom and why." *The Journal of Political Economy*, 114 (1), 175–201. [206, 210, 236]

Ciliberto, F. and E. Tamer (2009), "Market structure and multiple equilibria in airline markets." *Econometrica*, 77 (6), 1791–1828. [222, 228, 230]

Dagsvik, J. K. (2000), "Aggregation in matching markets." *International Economic Review*, 41 (1), 27–57. [206]

Demange, G., D. Gale, and M. A. O. Sotomayor (1986), "Multi-item auctions." *The Journal of Political Economy*, 94 (4), 863–872. [217]

Elyakime, B., J.-J. Laffont, P. Loisel, and Q. Vuong (1994), "First-price sealed-bid auctions with secret reservation prices." *Annales d'Economie et de Statistique*, 34, 115–141. [203]

Fox, J. T. (2007), "Semiparametric estimation of multinomial discrete-choice models using a subset of choices." *RAND Journal of Economics*, 38 (4), 1002–1019. [205, 210, 225, 227]

Fox, J. T. (2010), "Estimating matching games and automotive suppliers." Working paper, University of Michigan. [206, 211, 212, 217, 222, 229, 230]

Fox, J. T. and P. Bajari (2010), "Measuring the efficiency of an FCC spectrum auction." Working paper, University of Michigan. [206, 207, 217, 232]

Gale, D. and L. Shapley (1962), "College admissions and the stability of marriage." *American Mathematical Monthly*, 69, 9–15. [206]

Galichon, A. and M. Henry (2008), "Inference in models with multiple equilibria." Working paper, École Polytechnique. [228]

Garicano, L. and E. Rossi-Hansberg (2006), "Organization and inequality in a knowledge economy." *The Quarterly Journal of Economics*, 121 (4), 1383–1435. [218]

Goeree, J. K., C. A. Holt, and T. R. Palfrey (2005), "Regular quantal response equilibrium." *Experimental Economics*, 8, 347–367. [227]

Gordon, N. and B. Knight (2009), "A spatial merger estimator with an application to school district consolidation." *Journal of Public Economics*, 93, 752–765. [206]

Gourieroux, C. and A. Monfort (1995), *Statistics and Econometric Models*, volume 1 of Themes in Modern Econometrics. Cambridge University Press, Cambridge. [211]

Hatfield, J. W. and F. Kojima (2008), "Matching with contracts: Comment." *American Economic Review*, 98 (3), 1189–1194. [222]

Hatfield, J. W. and P. R. Milgrom (2005), "Matching with contracts." *American Economic Review*, 95 (4), 913–935. [222]

Hatfield, J. W. and S. D. Kominers (2010), "Many-to-many matching with contracts." Working paper, Stanford University. [222]

Horowitz, J. L. (1998), *Semiparametric Methods in Econometrics*, volume 131 of Lecture Notes in Statistics. Springer, New York. [234]

Jaume, D., J. Massó, and A. Neme (2009), "The multiple-partners assignment game with heterogeneous sales and multi-unit demands: Competitive equilibria." Working paper, Universidad Nacional de San Luis. [217]

Jehle, G. A. and P. J. Reny (2000), *Advanced Microeconomic Theory*, second edition. Addison Wesley, Boston. [246]

Jia, P. (2008), "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry." *Econometrica*, 76 (6), 1263–1316. [228, 230]

Kelso, A. S. and V. P. Crawford (1982), "Job matching, coalition formation, and gross substitutes." *Econometrica*, 50 (6), 1483–1504. [217]

Kojima, F., P. A. Pathak, and A. E. Roth (2010), "Matching with couples: Stability and incentives in large markets." Working paper, Stanford University. [222]

Koopmans, T. C. and M. Beckmann (1957), "Assignment problems and the location of economic activities." *Econometrica*, 25 (1), 53–76. [203, 207, 209, 216, 236]

Kovalenkov, A. and M. H. Wooders (2003), "Advances in the theory of large cooperative games and application to club theory: The side payments case." In *The Endogenous Formation of Economic Coalitions* (C. Carraro, ed.), Edward Elgar, Chelenham. [218]

Leonard, H. B. (1983), "Elicitation of honest preferences for the assignment of individuals to positions." *The Journal of Political Economy*, 91 (3), 461–479. [217]

Levine, A. A. (2009), "Licensing and scale economies in the biotechnology pharmaceutical industry." Working paper, Harvard University. [206]

Lewbel, A. (2000), "Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables." *Journal of Econometrics*, 97, 145–177. [234]

Magnac, T. and E. Maurin (2007), "Identification and information in monotone binary models." *Journal of Econometrics*, 139 (1), 76–104. [234]

Manski, C. F. (1975), "Maximum score estimation of the stochastic utility model of choice." *Journal of Econometrics*, 3, 205–228. [205, 210, 226, 227]

Manski, C. F. (1988), "Identification of binary response models." *Journal of the American Statistical Association*, 83 (403), 729–738. [226, 234]

Matzkin, R. L. (1993), "Nonparametric identification and estimation of polychotomous choice models." *Journal of Econometrics*, 58, 137–168. [204, 205, 210, 215, 216, 217, 233, 234, 236]

Mazzeo, M. J. (2002), "Product choice and oligopoly market structure." *The RAND Journal of Economics*, 33 (2), 221–242. [228]

Milgrom, P. and C. Shannon (1994), "Monotone comparative statics." *Econometrica*, 62 (1), 157–180. [231]

Mindruta, D. (2009), "Value creation in university-firm research collaborations: A matching approach." Working paper, HEC Paris. [206]

Ostrovsky, M. (2008), "Stability in supply chain networks." *American Economic Review*, 98 (3), 897–923. [218]

Pakes, A., J. Porter, K. Ho, and J. Ishii (2006), "Moment inequalities and their application." Working paper, Harvard University. [225]

Pycia, M. (2008), "Non-existence result for matching games with transfers." Working paper, UCLA. [222]

Roth, A. E. and M. A. O. Sotomayor (1990), *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press, Cambridge. [207, 229]

Seim, K. (2006), "An empirical model of firm entry with endogenous product-type choices." *The RAND Journal of Economics*, 37 (3), 619–640. [230]

Shapley, L. S. and M. Shubik (1972), "The assignment game I: The core." *International Journal of Game Theory*, 1, 111–130. [203, 207, 209, 216, 236]

Sørensen, M. (2007), "How smart is smart money? A two-sided matching model of venture capital." *Journal of Finance*, LXII (6), 2725–2762. [206, 210]

Sotomayor, M. (1992), "The multiple partners game." In *Equilibrium and Dynamics: Essays in Honour of David Gale* (M. Majumdar, ed.), Chapter 17, 322–336, Macmillan, New York. [217]

Sotomayor, M. (1999), "The lattice structure of the set of the stable outcomes of the multiple partners assignment game." *International Journal of Game Theory*, 28 (4), 567–583. [217]

Tamer, E. (2003), "Incomplete simultaneous discrete response model with multiple equilibria." *Review of Economic Studies*, 70 (1), 147–165. [228]

Yang, Y., M. Shi, and A. Goldfarb (2009), "Estimating the value of brand alliances in professional team sports." *Marketing Science*, 28 (6), 1095–1111. [206]

Submitted August, 2009. Final version accepted August, 2010.