Unobserved Heterogeneity in Matching Games

Jeremy T. Fox[1] Chenyu Yang[2]

[1]University of Michigan and NBER

[2]University of Michigan

BFI Matching Problems
June 2012

Outline

Matching Empirical Program

- Businesses form relationships with each other
- Data listing these relationships are sometimes available
  - Goodyear sold tires to Chrysler, etc.

Matching Empirical Program

- Businesses form relationships with each other
- Data listing these relationships are sometimes available
  - Goodyear sold tires to Chrysler, etc.
- What we can learn from data listing these relationships?

Matching Empirical Program

- Businesses form relationships with each other
- Data listing these relationships are sometimes available
  - Goodyear sold tires to Chrysler, etc.
- What we can learn from data listing these relationships?
- Matching games model relationship formation
  - Inputs: payoffs to matches
  - Outputs: stable matches
  - Firms on all sides of the market can be competing to match with the best partners

Matching Empirical Program

- Businesses form relationships with each other
- Data listing these relationships are sometimes available
  - Goodyear sold tires to Chrysler, etc.
- What we can learn from data listing these relationships?
- Matching games model relationship formation
  - Inputs: payoffs to matches
  - Outputs: stable matches
  - Firms on all sides of the market can be competing to match with the best partners
- What can we learn if we impose that the relationships in the data are a stable match?

Example of Matching for Car Parts

- Loosely inspired by Fox (2010a)
- Two suppliers of tires, Goodyear and Bridgestone
    - Upstream firms
- Two assemblers of cars, Chrysler and Hyundai
    - Downstream firms

Example of Matching for Car Parts

- Loosely inspired by Fox (2010a)
- Two suppliers of tires, Goodyear and Bridgestone
  - Upstream firms
- Two assemblers of cars, Chrysler and Hyundai
  - Downstream firms

- Matching game determines whether we see the assignment (list of matches)

$$\{\langle \text{Goodyear}, \text{Chrysler} \rangle, \langle \text{Bridgestone}, \text{Hyundai} \rangle\}$$

or the assignment

$$\{\langle \text{Goodyear}, \text{Hyundai} \rangle, \langle \text{Bridgestone}, \text{Chrysler} \rangle\}$$

What Matches Will Form?

- Matches occur according to *pairwise stability*
- Example *assignment*, a list of matches

$$\{\langle \text{Goodyear}, \text{Chrysler} \rangle, \langle \text{Bridgestone}, \text{Hyundai} \rangle\}$$

- Stability: Chrysler and Bridgestone could not both be better off by matching
- In transferable utility, money can compensate for a loss in direct structural profits

Available Data

- Assignment is

$$\{\langle \text{Goodyear}, \text{Chrysler} \rangle, \langle \text{Bridgestone}, \text{Hyundai} \rangle\}$$

- In terms of characteristics $(\text{experience}, \text{quality})$, assignment is

$$\{\langle (\text{low}, \text{low}), (\text{high}, \text{low}) \rangle, \langle (\text{high}, \text{high}), (\text{low}, \text{high}) \rangle\}$$

Available Data

- Assignment is

$$\{\langle \text{Goodyear}, \text{Chrysler}\rangle, \langle \text{Bridgestone}, \text{Hyundai}\rangle\}$$

- In terms of characteristics $(\text{experience}, \text{quality})$, assignment is

$$\{\langle (\text{low}, \text{low}), (\text{high}, \text{low})\rangle, \langle (\text{high}, \text{high}), (\text{low}, \text{high})\rangle\}$$

- **Quality not in data, observe only data**

$$\{\langle (\text{low}), (\text{high})\rangle, \langle (\text{high}), (\text{low})\rangle\}$$

- No data on rejections of partners, choice sets, transfers
- See hedonic models and labor panel literature for data on transfers (e.g, Heckman, Matzkin and Nesheim 2010, Chiappori, McCann, Nesheim 2010, Eeckhout and Kircher 2011)

Unobserved Characteristics

- Investigate the identification of objects such as distribution G of unobserved characteristics

$$G\,(\text{quality})$$

- Can we learn G from data on who matches with whom?

Literature Context for Unobserved Characteristics

- Matching empirical literature has modeled sorting on observed characteristics
  - Dozens of empirical papers by now
  - Including Choo & Siow (2006), Sorensen (2007), Fox (2010a)
  - Usually i.i.d. errors at match or type of matches level (or "rank order property")
  - Identification literature similar: Fox (2010b), Graham (2011), Galichon and Salanie (2011), etc.

Literature Context for Unobserved Characteristics

- Matching empirical literature has modeled sorting on observed characteristics
  - Dozens of empirical papers by now
  - Including Choo & Siow (2006), Sorensen (2007), Fox (2010a)
  - Usually i.i.d. errors at match or type of matches level (or "rank order property")
  - Identification literature similar: Fox (2010b), Graham (2011), Galichon and Salanie (2011), etc.
- Ackerberg and Botticini (2002) study matching between farmers and landlords
  - Matching-like IV's correct an outcome regression for bias from sorting on tenant risk aversion and landlord monitoring ability
  - Finds substantial bias, consistent with sorting on unobservables

Real-Time Literature Review

- Compared to Bernard's talk this morning
- Finite number of agents per market (firms in IO)
- Many different matching markets (say component categories)
- At least one continuous characteristic per match / agent (not finite number of observed types)
- Nonparametric on the joint distribution of unobservables
- No restriction on joint dependence of unobservables within a market (no i.i.d. errors)

Unobserved, Heterogeneous Preferences

- Agents may also have unobserved, heterogeneous preferences
  - Like random coefficients in demand models
- Chrysler cares more about experience than Hyundai?
- Unobserved preferences may be important in marriage
  - Observationally identical men married to observationally different women

Paper's Contribution

- Data on many matching markets
  - Who matches with whom (dependent variable)
  - Observed agent characteristics (independent variables)

Paper's Contribution

- Data on many matching markets
  - Who matches with whom (dependent variable)
  - Observed agent characteristics (independent variables)
- Explore (non)-identification of distributions of

1. Unobserved characteristics
2. Unobserved preferences
3. **Unobserved complementarities**

Paper's Contribution

- Data on many matching markets
  - Who matches with whom (dependent variable)
  - Observed agent characteristics (independent variables)
- Explore (non)-identification of distributions of

1. Unobserved characteristics
2. Unobserved preferences
3. **Unobserved complementarities**

- Mathematical similarities to multinomial choice models
- Emphasize unique aspects of matching

Analogy to Regression Models

- Analog to $y = x'\beta_i + \epsilon_i$
- **Assignment** (list of matches) dependent variable, y in regression
- **Observed characteristics** independent variables, x's in regression
- **Unobserved characteristics** (quality) like error ϵ_i in regression
- **Unobserved preferences** like random coefficients, β_i
- Want to learn $G(\epsilon_i, \beta_i)$

Outline

Scope of Baseline Model

- Baseline model
  - One-to-one, two-sided matching (marriage?)
  - Equal numbers of upstream, downstream firms
  - All firms must be matched
  - One observed characteristic per match
  - No random coefficients

Scope of Baseline Model

- Baseline model
  - One-to-one, two-sided matching (marriage?)
  - Equal numbers of upstream, downstream firms
  - All firms must be matched
  - One observed characteristic per match
  - No random coefficients
- Paper / project / end of talk
  - Number of firms can differ across sides
  - Unmatched firms in data
  - Multiple observed characteristics per match
  - Characteristics at firm, not match level
  - Heterogeneous coefficients on characteristics
  - Many-to-many matching

Physical and Full Matches

- One-to-one matching
  - Upstream firms u_1, u_2; downstream firms d_1, d_2

Physical and Full Matches

- One-to-one matching
  - Upstream firms u_1, u_2; downstream firms d_1, d_2
- Upstream firm u and downstream firm d can form **physical match** $\langle u, d \rangle$
  - Upstream firm listed first
  - Have data listing the matches that form

Physical and Full Matches

- One-to-one matching
  - Upstream firms u_1, u_2; downstream firms d_1, d_2
- Upstream firm u and downstream firm d can form **physical match** $\langle u, d \rangle$
  - Upstream firm listed first
  - Have data listing the matches that form
- In game solution, u and d form **full match** $\left\langle u, d, t_{\langle u,d \rangle} \right\rangle$
  - $t_{\langle u,d \rangle}$ transfers d pays to u
  - No data on transfers: often confidential

Match Production

- Total production from match $\langle u, d \rangle$ is

$$z_{\langle u,d \rangle} + e_{\langle u,d \rangle}$$

  - $z_{\langle u,d \rangle}$ regressor specific to match $\langle u, d \rangle$
  - $e_{\langle u,d \rangle}$ **unobservable** for match $\langle u, d \rangle$

Match Production

- Total production from match $\langle u, d \rangle$ is

$$z_{\langle u,d \rangle} + e_{\langle u,d \rangle}$$

  - $z_{\langle u,d \rangle}$ regressor specific to match $\langle u, d \rangle$
  - $e_{\langle u,d \rangle}$ **unobservable** for match $\langle u, d \rangle$

- $e_{\langle u,d \rangle}$ nests $e_{\langle u,d \rangle} = e_u \cdot e_d$
- Match production is sum of upstream, downstream profits

Matching Production

- N firms on each side of market

$$\begin{pmatrix} z_{\langle 1,1\rangle} + e_{\langle 1,1\rangle} & \cdots & z_{\langle 1,N\rangle} + e_{\langle 1,N\rangle} \\ \vdots & \ddots & \vdots \\ z_{\langle N,1\rangle} + e_{\langle N,1\rangle} & \cdots & z_{\langle N,N\rangle} + e_{\langle N,N\rangle} \end{pmatrix}$$

- Rows: upstream firms
- Columns: downstream firms

E and Z Matrices

$$E = \begin{pmatrix} e_{\langle 1,1 \rangle} & \cdots & e_{\langle 1,N \rangle} \\ \vdots & \ddots & \vdots \\ e_{\langle N,1 \rangle} & \cdots & e_{\langle N,N \rangle} \end{pmatrix}, Z = \begin{pmatrix} z_{\langle 1,1 \rangle} & \cdots & z_{\langle 1,N \rangle} \\ \vdots & \ddots & \vdots \\ z_{\langle N,1 \rangle} & \cdots & z_{\langle N,N \rangle} \end{pmatrix}$$

- Z in data
- E not in data, observed to agents

Assignments

- **Assignment** A selects one cell from each row, each column
- $A = \{\langle u_1, d_1\rangle, \dots, \langle u_N, d_N\rangle\}$

$$\begin{pmatrix} \times & \circ & \dots & \circ \\ \circ & \times & \dots & \circ \\ \vdots & \vdots & \ddots & \vdots \\ \circ & \circ & \dots & \times \end{pmatrix}$$

Solution Concept: Pairwise Stability

- **Outcome** list of full matches
  - $\left\{\left\langle u_1, d_1, t_{\langle \mathbf{u_1}, \mathbf{d_1} \rangle}\right\rangle, \ldots, \left\langle u_N, d_N, t_{\langle \mathbf{u_N}, \mathbf{d_N} \rangle}\right\rangle\right\}$
- Outcome **pairwise stable** if robust to deviations by pairs of two firms
- Again, **assignment** A list of physical matches
  - $\{\langle u_1, d_1\rangle, \ldots, \langle u_N, d_N\rangle\}$
- Call assignment **pairwise stable** if underlying outcome pairwise stable

Existence and Uniqueness

- Roth and Sotomayor (1990, Chapter 8)
- Existence of pairwise stable assignment guaranteed
- Pairwise stable outcome is fully stable
  - Robust to deviation by any coalition of firms
  - One such coalition is set of all firms
- Let $S(A, E, Z) = \sum_{\langle u,d \rangle \in A} \left(z_{\langle u,d \rangle} + e_{\langle u,d \rangle} \right)$
- Pairwise stable assignment A maximizes $S(A, E, Z)$
- Maximizes sum of production across all assignments
- Uniqueness of assignment with probability 1 if E, Z arguments have continuous support

Data Across Markets

- Data (A, Z) from many markets
- Assignment $A = \{\langle u_1, d_1 \rangle, \ldots, \langle u_N, d_N \rangle\}$
- Observed characteristics

$$Z = \begin{pmatrix} z_{\langle 1,1 \rangle} & \cdots & z_{\langle 1,N \rangle} \\ \vdots & \ddots & \vdots \\ z_{\langle N,1 \rangle} & \cdots & z_{\langle N,N \rangle} \end{pmatrix}$$

Full Support on Z

$$Z = \begin{pmatrix} z_{\langle 1,1 \rangle} & \cdots & z_{\langle 1,N \rangle} \\ \vdots & \ddots & \vdots \\ z_{\langle N,1 \rangle} & \cdots & z_{\langle N,N \rangle} \end{pmatrix}$$

- Limiting data are $\Pr(A \mid Z)$
- Let Z have **full and product support**
- Any $Z \in \mathbb{R}^{N^2}$ is observed
- **Special regressor** used for point identification in binary/multinomial choice
  - Ichimura and Thompson (1998), Lewbel (2000), Matzkin (2007), Berry and Haile (2011), Fox and Gandhi (2010), etc.

$G(E)$: Key Primitive in the Model

- Unknown primitive to estimate is the **distribution** $G(E)$ of

$$E = \begin{pmatrix} e_{\langle 1,1 \rangle} & \cdots & e_{\langle 1,N \rangle} \\ \vdots & \ddots & \vdots \\ e_{\langle N,1 \rangle} & \cdots & e_{\langle N,N \rangle} \end{pmatrix}$$

- Different markets have different unobservable realizations E
- $G(E)$: distribution across markets
- Assume Z independent of E

Identification

- Data generation process

$$\Pr(A \mid Z;\ G) = \int 1\left[A \text{ stable} \mid Z, E\right] dG(E)$$

- $G(E)$ **identified** if true G only distribution that generates data $\Pr(A \mid Z)$ for all (A, Z)

Location Normalizations

- Add a constant to the production of all matches involving firm 1
    - Relative production of all assignments remains the same
    - Already non-identification result
- Location normalizations: $e_{\langle i,i \rangle} = 0 \, \forall \, i = 1, \ldots, N$

$$E = \begin{pmatrix} 0 & e_{\langle 1,2 \rangle} & \cdots & e_{\langle 1,N \rangle} \\ e_{\langle 2,1 \rangle} & 0 & \cdots & e_{\langle 2,N \rangle} \\ \vdots & \vdots & \ddots & \vdots \\ e_{\langle N,1 \rangle} & e_{\langle N,2 \rangle} & \cdots & 0 \end{pmatrix}$$

$G(E)$ is Not Identified

- Recall $S(A, E, Z) = \sum_{\langle u,d\rangle \in A} \left(e_{\langle u,d\rangle} + z_{\langle u,d\rangle}\right)$ governs pairwise stable assignment
- Compare

$$E_1 = \begin{pmatrix} 0 & e_{\langle 1,2\rangle} & \cdots & e_{\langle 1,N\rangle} \\ e_{\langle 2,1\rangle} & 0 & \cdots & e_{\langle 2,N\rangle} \\ \vdots & \vdots & \ddots & \vdots \\ e_{\langle N,1\rangle} & e_{\langle N,2\rangle} & \cdots & 0 \end{pmatrix}$$

$$E_2 = \begin{pmatrix} 0 & e_{\langle 1,2\rangle}+1 & \cdots & e_{\langle 1,N\rangle} \\ e_{\langle 2,1\rangle}-1 & 0 & \cdots & e_{\langle 2,N\rangle}-1 \\ \vdots & \vdots & \ddots & \vdots \\ e_{\langle N,1\rangle} & e_{\langle N,2\rangle}+1 & \cdots & 0 \end{pmatrix}$$

- E_1 and E_2 have same sums of unobserved production for all assignments

Non-Identification Theorem

- $S(A, E_1, Z) = S(A, E_2, Z) \; \forall A, Z$
- Frequencies of E_1 and E_2 cannot be distinguished
- Cannot identify if firms tend to be high quality from these data on matched firms

Theorem

*The distribution $G(E)$ of market-level unobserved match characteristics is **not** identified.*

Complementarities Drive Matching

- If distribution of E not identified, what distribution is?
- Becker (1973): marriage with heterogeneous schooling levels
- **Assortative matching** when male and female schooling are complements in production
- **Complementarities:** positive *cross partial derivative* of production with respect to schooling
- *Increasing differences* if schooling discrete

Unobserved Complementarities

- Let

$$c(u_1, u_2, d_1, d_2) \equiv e_{\langle u_1, d_1 \rangle} + e_{\langle u_2, d_2 \rangle} - e_{\langle u_1, d_2 \rangle} - e_{\langle u_2, d_1 \rangle}$$

- **Unobserved complementarity** between the matches $\langle u_1, d_1 \rangle$ and $\langle u_2, d_2 \rangle$
  - Relative to exchange of partners $\langle u_1, d_2 \rangle$ and $\langle u_2, d_1 \rangle$
- One unobserved complementarity for each of two upstream, two downstream firms
- How much matches $\langle u_1, d_1 \rangle$ and $\langle u_2, d_2 \rangle$ gain in unobserved quality over matches $\langle u_1, d_2 \rangle$ and $\langle u_2, d_1 \rangle$

Market-Level Unobserved Complementarities

- Match-specific unobservables for each market

$$E = \begin{pmatrix} 0 & e_{\langle 1,2 \rangle} & \cdots & e_{\langle 1,N \rangle} \\ e_{\langle 2,1 \rangle} & 0 & \cdots & e_{\langle 2,N \rangle} \\ \vdots & \vdots & \ddots & \vdots \\ e_{\langle N,1 \rangle} & e_{\langle N,2 \rangle} & \cdots & 0 \end{pmatrix}$$

- Change variables

$$C = \left(c \left(u_1, u_2, d_1, d_2 \right) \mid u_1, u_2, d_1, d_2 \in N \right)$$

- Each valid C must be formed from a valid E

Market-Level Unobserved Complementarities

Lemma

There is a random vector

$$B = (c\,(u_1, u_2, d_1, d_2) \mid u_1 = d_1 = 1,\ u_2, d_2 \in \{2, \ldots, N\})$$

of $(N-1)^2$ *unobserved complementarities such that any unobserved complementarity* $c\,(u_1, u_2, d_1, d_2)$ *in* C *is equal to a* (u_1, u_2, d_1, d_2)*-specific sum and difference of terms in* B*. The indices* (u_1', u_2', d_1', d_2') *of the terms in* B *in the sum do not depend on the realization of* E*.*

Ex: $N = 3$ Agents Per Side

$$E = \begin{pmatrix} 0 & e_{\langle 1,2 \rangle} & e_{\langle 1,3 \rangle} \\ e_{\langle 2,1 \rangle} & 0 & e_{\langle 2,3 \rangle} \\ e_{\langle 3,1 \rangle} & e_{\langle 3,2 \rangle} & 0 \end{pmatrix}$$

Ex: $N = 3$ Agents Per Side

- 12 items in C

$$C = \left(c\left(u_1, u_2, d_1, d_2\right) \mid u_1, u_2, d_1, d_2 \in \{1,2,3\}\right)$$

- Definition of B, 4 items in B

$$\begin{aligned} B = (c\,(1,2,1,2)\,, c\,(1,2,1,3)\,, c\,(1,3,1,2)\,, c\,(1,3,1,3)) = \\ \left(-\left(e_{\langle 1,2\rangle} + e_{\langle 2,1\rangle}\right), e_{\langle 2,3\rangle} - \left(e_{\langle 1,3\rangle} + e_{\langle 2,1\rangle}\right),\right. \\ \left. e_{\langle 3,2\rangle} - \left(e_{\langle 1,2\rangle} + e_{\langle 3,1\rangle}\right), -\left(e_{\langle 1,3\rangle} + e_{\langle 3,1\rangle}\right)\right) \end{aligned}$$

- Example of constructing item in C from B

$$\begin{aligned} c\,(2,3,2,3) &= e_{\langle 2,2\rangle} + e_{\langle 3,3\rangle} - \left(e_{\langle 2,3\rangle} + e_{\langle 3,2\rangle}\right) \\ &= -\left(e_{\langle 2,3\rangle} + e_{\langle 3,2\rangle}\right) \\ &= c\,(1,2,1,2) - c\,(1,2,1,3) - c\,(1,3,1,2) + c\,(1,3,1,3) \end{aligned}$$

Distribution of Unobserved Complementarities

- Recall

$$C = (c\,(u_1, u_2, d_1, d_2) \mid u_1, u_2, d_1, d_2 \in N)$$

- **Try to identify joint distribution** $F\,(C)$

Unobserved Complementarities and Assignments

- Recall $S(A,E,Z) = \sum_{\langle u,d\rangle \in A} \left(e_{\langle u,d\rangle} + z_{\langle u,d\rangle}\right)$ governs pairwise stable assignment
- Let $\tilde{S}(A,E) = \sum_{\langle u,d\rangle \in A} e_{\langle u,d\rangle}$ be **unobserved production** from assignment A

Lemma

For each A, $\tilde{S}(A,E)$ is equal to an A-specific sum and difference of unobserved complementarities in C. The indices (u_1, u_2, d_1, d_2) of the terms in the sum do not depend on the realization of E.

- Use the overloaded notation $\tilde{S}(A,C)$ for $\tilde{S}(A,E)$
- Can calculate optimal assignment from C and Z
- Hence, assignment probabilities from $F(C)$

Ex: $N = 3$ Agents Per Side

$$
\begin{aligned}
A_1 &= \{\langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle\} \\
A_2 &= \{\langle 1,2\rangle, \langle 2,1\rangle, \langle 3,3\rangle\} \\
A_3 &= \{\langle 1,3\rangle, \langle 2,2\rangle, \langle 3,1\rangle\} \\
A_4 &= \{\langle 1,2\rangle, \langle 2,3\rangle, \langle 3,1\rangle\} \\
A_5 &= \{\langle 1,1\rangle, \langle 2,3\rangle, \langle 3,2\rangle\} \\
A_6 &= \{\langle 1,3\rangle, \langle 2,1\rangle, \langle 3,2\rangle\}
\end{aligned}
$$

Ex: $N = 3$ Agents Per Side

- Write sum of unobserved production as sum of elements in C

$$\begin{pmatrix} \tilde{S}(A_1, E) \\ \tilde{S}(A_2, E) \\ \tilde{S}(A_3, E) \\ \tilde{S}(A_4, E) \\ \tilde{S}(A_5, E) \\ \tilde{S}(A_6, E) \end{pmatrix} =$$

$$\begin{pmatrix} 0 \\ e_{\langle 1,2\rangle} + e_{\langle 2,1\rangle} \\ e_{\langle 1,3\rangle} + e_{\langle 3,1\rangle} \\ e_{\langle 1,2\rangle} + e_{\langle 2,3\rangle} + e_{\langle 3,1\rangle} \\ e_{\langle 2,3\rangle} + e_{\langle 3,2\rangle} \\ e_{\langle 1,3\rangle} + e_{\langle 2,1\rangle} + e_{\langle 3,2\rangle} \end{pmatrix} = \begin{pmatrix} 0 \\ -c(1,2,1,2) \\ -c(1,3,1,3) \\ c(1,2,2,3) - c(1,3,1,3) \\ -c(2,3,2,3) \\ -c(1,3,1,3) + c(2,3,1,2) \end{pmatrix}$$

Unobserved Complementarities Empirically Distinguishable

- Recall

$$C = (c(u_1, u_2, d_1, d_2) \mid u_1, u_2, d_1, d_2 \in N)$$

Lemma

Consider two realizations C_1 and C_2 of the random vector C. $C_1 = C_2$ if and only if $\tilde{S}(A, C_1) = \tilde{S}(A, C_2)$ for all assignments A.

- If $C_1 \neq C_2$, there exists A such that $\tilde{S}(A, C_1) \neq \tilde{S}(A, C_2)$
- Distribution $F(C)$ is potentially identifiable

Ex: $N = 3$ Agents Per Side

- Given two realizations C_1 and C_2, if $\tilde{S}(A, C_1) = \tilde{S}(A, C_2)$ for all A, then $C_1 = C_2$

$$\begin{aligned} c(1,2,1,2) &= \tilde{S}(A_1, C) - \tilde{S}(A_2, C) \\ c(1,2,1,3) &= \tilde{S}(A_5, C) - \tilde{S}(A_6, C) \\ c(1,3,1,2) &= \tilde{S}(A_5, C) - \tilde{S}(A_4, C) \\ c(1,3,1,3) &= \tilde{S}(A_1, C) - \tilde{S}(A_3, C) \end{aligned}$$

- If $C_1 = C_2$, then $\tilde{S}(A, C_1) = \tilde{S}(A, C_2)$ for all A
  - Follows from formulas for $\tilde{S}(A, E)$
  - Recall $\tilde{S}(A, C)$ and $\tilde{S}(A, E)$ overloaded notation for same sum

Main Result: $F(C)$ is Identified

- First identify the distribution of $\tilde{S}$ by varying Z across markets
  - Sums of unobserved production for all assignments in a market
- Then change variables to get distribution $F(E)$
  - Change of variables is *one-to-one* by previous lemma
  - So $F(C)$ is identified

Theorem

The distribution $F(C)$ of market-level unobserved complementarities is identified in a matching game where all agents must be matched.

The Distribution of $\tilde{S}$

- $\tilde{S}(A, C)$ sum of unobserved production for assignment A
- $N!$ assignments A in a market
- Differences in assignment production govern pairwise stable assignment
  - Use $A_1 = \{\langle 1, 1\rangle, \ldots, \langle N, N\rangle\}$ as a baseline assignment
  - $\tilde{S}(A_1, C) = 0 \,\forall\, C$ by earlier location normalization
- $\tilde{S} = \left(\tilde{S}(A_i, C)\right)_{i=2}^{N!}$ vector of random variables

Lemma

The CDF $H\left(\tilde{S}\right)$ of unobserved production for all assignments is identified.

Proof: Identifying $H\left(\tilde{S}\right)$ Using Z

- Recall $S(A, E, Z) = \sum_{\langle u,d\rangle \in A} \left(e_{\langle u,d\rangle} + z_{\langle u,d\rangle}\right)$ governs pairwise stable assignment

Proof: Identifying $H\left(\tilde{S}\right)$ Using Z

- Recall $S(A, E, Z) = \sum_{\langle u,d \rangle \in A} \left(e_{\langle u,d \rangle} + z_{\langle u,d \rangle}\right)$ governs pairwise stable assignment
- Each $E^\star$ gives one C^* & one $\tilde{S}^\star = \tilde{S}(A, C^\star)$, set

$$z^\star_{\langle u,d \rangle} = -e^\star_{\langle u,d \rangle}$$

Proof: Identifying $H\left(\tilde{S}\right)$ Using Z

- Recall $S(A,E,Z)=\sum_{\langle u,d\rangle\in A}\left(e_{\langle u,d\rangle}+z_{\langle u,d\rangle}\right)$ governs pairwise stable assignment
- Each $E^{\star}$ gives one C^{*} & one $\tilde{S}^{\star}=\tilde{S}(A,C^{\star})$, set

$$z^{\star}_{\langle u,d\rangle}=-e^{\star}_{\langle u,d\rangle}$$

- Then $S(A,E^{\star},Z^{\star})=\tilde{S}(A,C^{\star})+\sum_{\langle u,d\rangle\in A}z^{\star}_{\langle u,d\rangle}=0\,\forall\,A$

Proof: Identifying $H\left(\tilde{S}\right)$ Using Z

- Recall $S(A,E,Z)=\sum_{\langle u,d\rangle\in A}\left(e_{\langle u,d\rangle}+z_{\langle u,d\rangle}\right)$ governs pairwise stable assignment
- Each $E^{\star}$ gives one C^{*} & one $\tilde{S}^{\star}=\tilde{S}(A,C^{\star})$, set

$$z^{\star}_{\langle u,d\rangle}=-e^{\star}_{\langle u,d\rangle}$$

- Then $S(A,E^{\star},Z^{\star})=\tilde{S}(A,C^{\star})+\sum_{\langle u,d\rangle\in A}z^{\star}_{\langle u,d\rangle}=0\,\forall\,A$
- Definition of the CDF

$$H\left(\tilde{S}^{\star}\right)=\Pr\left(\tilde{S}(A,C)\leq\tilde{S}(A,C^{\star}),\forall\,A\neq A_1\right)$$

Proof: Identifying $H\left(\tilde{S}\right)$ Using Z

$$
\begin{aligned}
H\left(\tilde{S}^{\star}\right) &= \Pr\left(\tilde{S}(A,C) \leq \tilde{S}(A,C^{\star}), \forall A \neq A_1\right) \\
&= \Pr\left(S(A,E,Z^{\star}) \leq S(A_1,E,Z^{\star}), \forall A \neq A_1\right) \\
&= \Pr\left(S(A,E,Z^{\star}) \leq 0, \forall A \neq A_1\right) \\
&= \Pr\left(A_1 \mid Z^{\star}\right)
\end{aligned}
$$

- Third equality uses choice of $Z^{\star}$:
- Uses $\Pr(A_1 \mid Z^{*})$ for arbitrary assignment A_1, many Z^{*}

Special Regressors and Tracing CDFs

1. Large and product support on Z traces CDF of sums of unobserved production of assignments
   - Special regressors
   - Ichimura and Thompson (1998), Lewbel (2000), Matzkin (2007), Berry and Haile (2011), Fox and Gandhi (2010)
   - Failure of large and product support gives partial identification of $H\left(\tilde{S}\right)$ and hence $F\left(C\right)$

Special Regressors and Tracing CDFs

1. Large and product support on Z traces CDF of sums of unobserved production of assignments
   - Special regressors
   - Ichimura and Thompson (1998), Lewbel (2000), Matzkin (2007), Berry and Haile (2011), Fox and Gandhi (2010)
   - Failure of large and product support gives partial identification of $H\left(\tilde{S}\right)$ and hence $F\left(C\right)$

1. Given $H\left(\tilde{S}\right)$, change of variables completes proof of identification of $F\left(C\right)$

Recap of Main Results

- Negative identification result

Theorem

*The distribution $G(E)$ of market-level unobserved match characteristics is **not** identified in a matching game where all agents must be matched.*

- Positive identification result

Theorem

The distribution $F(C)$ of market-level unobserved complementarities is identified in a matching game where all agents must be matched.

Economic Intuition for Unobserved Complementarities

- Transferable utility matching games
- Becker (1973) shows complementarities govern sorting
  - One characteristic (schooling) per agent
- Positive assortative matching could occur if men want to marry women with
  - Same level of schooling (horizontal preferences)
  - Highest level of schooling (vertical preferences)
- Have both match-specific observables and unobservables
- Nevertheless, can learn about the distribution of unobserved complementarities

Outline

Other Observed Characteristics X

- Researcher observes other market-level characteristics X
- In addition to special regressors in Z
- Firm or agent specific characteristics
- Number of firms could vary, be in X

Example of Production with X

- Total match production

$$(x_u \cdot x_d)' \beta_{\langle u,d\rangle,1} + x'_{\langle u,d\rangle} \beta_{\langle u,d\rangle,2} + \mu_{\langle u,d\rangle} + z_{\langle u,d\rangle}$$

  - x_u vector of upstream firm characteristics
  - x_d vector of downstream firm characteristics
  - $x_u \cdot x_d$ all interactions between upstream, downstream characteristics
  - $x_{\langle u,d\rangle}$ vector of match-specific characteristics
  - $\beta_{\langle u,d\rangle,1}$, $\beta_{\langle u,d\rangle,2}$ random coefficients specific to match
    - Can be sum of **random preferences** of upstream, downstream firms
  - $\mu_{\langle u,d\rangle}$ random intercept
    - Can capture unobserved characteristics of both u and d

$$X = \left(N, (x_u)_{u \in N}, (x_d)_{d \in N}, \left(x_{\langle u,d\rangle}\right)_{u,d \in N}\right)$$

More on Example with X

- Total match production

$$(x_u \cdot x_d)' \beta_{\langle u,d\rangle,1} + x'_{\langle u,d\rangle}\beta_{\langle u,d\rangle,2} + \mu_{\langle u,d\rangle} + z_{\langle u,d\rangle}$$

- Now define

$$e_{\langle u,d\rangle} = (x_u \cdot x_d)' \beta_{\langle u,d\rangle,1} + x'_{\langle u,d\rangle}\beta_{\langle u,d\rangle,2} + \mu_{\langle u,d\rangle}$$

and

$$c(u_1, u_2, d_1, d_2) \equiv e_{\langle u_1,d_1\rangle} + e_{\langle u_2,d_2\rangle} - e_{\langle u_1,d_2\rangle} - e_{\langle u_2,d_1\rangle}$$

Condition on X

- Previous theorems did not use X, can condition on X
- Example model makes the distribution $F(C \mid X)$ of

$$C = (c(u_1, u_2, d_1, d_2) \mid u_1, u_2, d_1, d_2 \in N)$$

  depend on X
    - Still require independence of Z and ψ
- Prior arguments identify $F(C \mid X)$

Data on Unmatched Firms

- Full matching model allows firms to be unmatched in stable assignments
- In some IO applications, data on these unmatched firms
    - Potential merger partners, single people in marriage
- **Say we can have data on unmatched firms**
- Let $\langle u, 0\rangle$ be a physical match for an unmatched upstream firm
    - Also, use $\langle 0, d\rangle$
- Assignments like this allowed

$$\{\langle u_1, 0\rangle, \langle 0, d_1\rangle, \langle u_2, d_2\rangle\}$$

Unmatched Has 0 Production

- No special regressor for single matches
- $e_{\langle u,0\rangle} = 0$ for single matches as a location normalization, so

$$E = \begin{pmatrix} e_{\langle 1,1\rangle} & \cdots & e_{\langle 1,N_d\rangle} \\ \vdots & \ddots & \vdots \\ e_{\langle N_u,1\rangle} & \cdots & e_{\langle N_u,N_d\rangle} \end{pmatrix}$$

Unmatched Has 0 Production

- No special regressor for single matches
- $e_{\langle u,0\rangle} = 0$ for single matches as a location normalization, so

$$E = \begin{pmatrix} e_{\langle 1,1\rangle} & \cdots & e_{\langle 1,N_d\rangle} \\ \vdots & \ddots & \vdots \\ e_{\langle N_u,1\rangle} & \cdots & e_{\langle N_u,N_d\rangle} \end{pmatrix}$$

- Without unmatched firms, could not identify $G(E)$
- Only distribution $F(C)$ of unobservable complementarities

Theorem

The distribution $G(E \mid X)$ of market-level unobservables is constructively identified with data on unmatched agents.

Proof: $G(E)$ is Identified

- Fix $E^\star$, set $z^\star_{\langle u,d\rangle} = -e^\star_{\langle u,d\rangle}$
  - Then the production of all assignments is 0
  - All agents indifferent between being unmatched and matched

Proof: $G(E)$ is Identified

- Fix $E^{\star}$, set $z^{\star}_{\langle u,d\rangle} = -e^{\star}_{\langle u,d\rangle}$
  - Then the production of all assignments is 0
  - All agents indifferent between being unmatched and matched
- Let A_0 be assignment *where all agents are unmatched*
  - $\tilde{S}(A_0, E) = 0$
  - Agents still unmatched if $e_{\langle u,d\rangle} \leq e^{\star}_{\langle u,d\rangle} \forall \langle u,d\rangle$

Proof: $G(E)$ is Identified

- Fix $E^\star$, set $z^\star_{\langle u,d\rangle} = -e^\star_{\langle u,d\rangle}$
  - Then the production of all assignments is 0
  - All agents indifferent between being unmatched and matched
- Let A_0 be assignment *where all agents are unmatched*
  - $\tilde{S}(A_0, E) = 0$
  - Agents still unmatched if $e_{\langle u,d\rangle} \leq e^\star_{\langle u,d\rangle} \forall \langle u, d\rangle$
- Then

$$G(E^\star) = \Pr(E \leq E^* \text{ elementwise}) = \Pr(A_0 \mid Z^\star)$$

Intuition for Identification of $G(E)$

- Without unmatched agents, can only identify distribution of unobserved complementarities
- With unmatched agents, introduces an element of individual rationality in the data
  - Agent can unilaterally decide to be single
  - Production of all non-single matches must be nonpositive when all other agents are available to match
- Look at probability all agents are single given Z
- Individual rationality makes identification similar to
  - Single agent multinomial choice
  - Nash games

Agent-Specific Characteristics in Z

- Results rely on *match-specific* special regressors $z_{\langle u,d \rangle}$
- Now *agent-specific* regressors z_u and z_d
- $2 \cdot N$ such regressors

$$Z = \left((z_u)_{u \in N}, (z_d)_{d \in N} \right)$$

Agent-Specific Characteristics in Z

- Only matched firms
- Functional form of production

$$e_u \cdot e_d + z_u \cdot z_d$$

- Only interactions matter in sorting if agents must be matched

Agent-Specific Characteristics

- With data on unmatched firms, can get at distribution $G(E)$ of
$$E = \left((e_u)_{u=3}^{N}, (e_d)_{d=2}^{N} \right).$$
- Normalizations: $e_u = 0$ for $u = 1$, $e_d = 0$ for $d = 1$, $e_u = 1$ for $u = 2$

Theorem

The distribution $G(E \mid X)$ is identified in the one-to-one matching model with agent-specific characteristics, agent-specific unobservables, and without unmatched agents.

One-Sided Matching

- Consider the example of mergers
- Which firm is a target and which is an acquirer is an endogenous outcome
- None of the previous theorems relied on dividing agents into two sides
- Our results automatically generalize to one-sided matching
- Existence issues (Chiappori, Galichon and Salanie 2012)

Many-to-Many, Two-Sided Matching

- Many-to-many matching: upstream firms can have multiple downstream firm partners
  - And downstream firms can have multiple upstream firm partners

Many-to-Many, Two-Sided Matching

- Many-to-many matching: upstream firms can have multiple downstream firm partners
    - And downstream firms can have multiple upstream firm partners
- Additive separability: production of matches $\langle u_1, d_1 \rangle$ and $\langle u_1, d_2 \rangle$

$$z_{\langle u_1, d_1 \rangle} + e_{\langle u_1, d_1 \rangle} + z_{\langle u_1, d_2 \rangle} + e_{\langle u_1, d_2 \rangle}$$

    - Sotomayor (1999)
- Results simply generalize when production is additively separable across multiple matches involving the same firm

Multiple Pairwise Stable Assignments

- Transferable utility matching games with production not additively separable across multiple matches may have multiple pairwise stable assignments
- Also may have existence issues
- Need to adopt some sort of solution to games with multiple equilibria
  - Parameterize selection rule
  - Broad assumptions about selection rule
  - Partial identification
  - Identify selection rule?

Conclusions

- Study identification in matching games
  - Data on assignments (lists of matches)
  - Observed agent, match characteristics
- **Without unmatched agents, can identify distribution of unobserved complementarities**
- With unmatched agents, can identify distribution of unobserved match *characteristics*