Unobserved Heterogeneity in Matching Games

Jeremy T. Fox\textsuperscript{1} Chenyu Yang\textsuperscript{2}

\textsuperscript{1}University of Michigan and NBER
\textsuperscript{2}University of Michigan

BFI Matching Problems
June 2012
Outline

1 Matching Empirical Program

2 Baseline Model

3 Model Variants
   - Other Observed Characteristics
   - Data on Unmatched Firms
   - Agent-Specific Characteristics
   - One-Sided Matching
   - Many-to-Many Matching
Businesses form relationships with each other

Data listing these relationships are sometimes available
  - Goodyear sold tires to Chrysler, etc.
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?
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
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)
  \[
  \{\langle \text{Goodyear, Chrysler} \rangle, \langle \text{Bridgestone, Hyundai} \rangle\}
  \]
  or the assignment
  \[
  \{\langle \text{Goodyear, Hyundai} \rangle, \langle \text{Bridgestone, Chrysler} \rangle\}\]
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 \}\n
- 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
Assignment is

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

In terms of characteristics (experience, quality), assignment is

$$\{\langle \langle \text{low, low} \rangle, \langle \text{high, low} \rangle \rangle, \langle \langle \text{high, high} \rangle, \langle \text{low, high} \rangle \rangle \}$$
Available Data

- Assignment is
  \[
  \{\langle Goodyear, Chrysler \rangle, \langle Bridgestone, Hyundai \rangle \}\]

- In terms of characteristics (experience, quality), assignment is
  \[
  \{\langle (low, low) \rangle, \langle (high, low) \rangle, \langle (high, high) \rangle, \langle (low, high) \rangle \}\}
  
- Quality not in data, observe only data
  \[
  \{\langle (low) \rangle, \langle (high) \rangle, \langle (high) \rangle, \langle (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?
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.
Unobserved Heterogeneity in Matching Games
Matching Empirical Program

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
Unobserved Heterogeneity in Matching Games
Matching Empirical Program

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 $\epsilon_i$ in regression
- **Unobserved preferences** like random coefficients, $\beta_i$
- Want to learn $G(\epsilon_i, \beta_i)$
Unobserved Heterogeneity in Matching Games
Baseline Model

Outline

1. Matching Empirical Program
2. Baseline Model
3. Model Variants
   - Other Observed Characteristics
   - Data on Unmatched Firms
   - Agent-Specific Characteristics
   - One-Sided Matching
   - Many-to-Many Matching
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
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 $\langle u, d, t_{(u,d)} \rangle$
- $t_{(u,d)}$ transfers $d$ pays to $u$
- No data on transfers: often confidential
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
$N$ 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

\[ E = \begin{pmatrix} e_{1,1} & \cdots & e_{1,N} \\ \vdots & \ddots & \vdots \\ e_{N,1} & \cdots & e_{N,N} \end{pmatrix}, \quad Z = \begin{pmatrix} z_{1,1} & \cdots & z_{1,N} \\ \vdots & \ddots & \vdots \\ z_{N,1} & \cdots & z_{N,N} \end{pmatrix} \]

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

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

\[
\begin{pmatrix}
\times & \circ & \cdots & \circ \\
\circ & \times & \cdots & \circ \\
\vdots & \vdots & \ddots & \vdots \\
\circ & \circ & \cdots & \times 
\end{pmatrix}
\]
Solution Concept: Pairwise Stability

- **Outcome** list of full matches
  \[ \{ \langle u_1, d_1, t_{\langle u_1, d_1 \rangle} \rangle, \ldots, \langle u_N, d_N, t_{\langle u_N, d_N \rangle} \rangle \} \]

- 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 $S(A, E, Z) = \sum_{(u,d) \in A} (z_{(u,d)} + e_{(u,d)})$
- 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
Unobserved Heterogeneity in Matching Games
Baseline Model

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

\[
Z = \begin{pmatrix}
Z_{\langle 1,1 \rangle} & \cdots & Z_{\langle 1,N \rangle} \\
\vdots & \ddots & \vdots \\
Z_{\langle N,1 \rangle} & \cdots & Z_{\langle N,N \rangle}
\end{pmatrix}
\]
Unobserved Heterogeneity in Matching Games
Baseline Model

Full Support on $Z$

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

- Limiting data are $Pr(A | Z)$
- 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
$G(E)$: 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 $E$
Data generation process

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

- \( G(E) \) 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: \( e_{\langle i,i \rangle} = 0 \ \forall \ i = 1, \ldots, N \)

\[
E = \begin{pmatrix}
0 & e_{\langle 1,2 \rangle} & \cdots & e_{\langle 1,N \rangle} \\
e_{\langle 2,1 \rangle} & 0 & \cdots & e_{\langle 2,N \rangle} \\
\vdots & \vdots & \ddots & \vdots \\
e_{\langle N,1 \rangle} & e_{\langle N,2 \rangle} & \cdots & 0
\end{pmatrix}
\]
Unobserved Heterogeneity in Matching Games
Baseline Model

**G (E) is Not Identified**

- Recall \( S (A, E, Z) = \sum_{\langle u,d \rangle \in A} (e_{\langle u,d \rangle} + z_{\langle u,d \rangle}) \) 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.
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 Heterogeneity in Matching Games
Baseline Model

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 $\langle u_2, d_1 \rangle$
Match-specific unobservables for each market

\[ E = \begin{pmatrix}
0 & e_{1,2} & \cdots & e_{1,N} \\
e_{2,1} & 0 & \cdots & e_{2,N} \\
\vdots & \vdots & \ddots & \vdots \\
e_{N,1} & e_{N,2} & \cdots & 0
\end{pmatrix} \]

Change variables

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

Each valid \( C \) must be formed from a valid \( E \)
Unobserved Heterogeneity in Matching Games
Baseline Model

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

- Definition of $B$, 4 items in $B$

  \[ B = (c(1, 2, 1, 2), c(1, 2, 1, 3), c(1, 3, 1, 2), c(1, 3, 1, 3)) = (-(e_{2, 1} + e_{2, 1}), e_{2, 3} - (e_{1, 3} + e_{2, 1}), e_{3, 2} - (e_{1, 2} + e_{3, 1}), -(e_{1, 3} + e_{3, 1})) \]

- Example of constructing item in $C$ from $B$

  \[ c(2, 3, 2, 3) = e_{2, 2} + e_{3, 3} - (e_{2, 3} + e_{3, 2}) \]
  \[ = -(e_{2, 3} + e_{3, 2}) \]
  \[ = c(1, 2, 1, 2) - c(1, 2, 1, 3) - c(1, 3, 1, 2) + c(1, 3, 1, 3) \]
Recall

\[ C = (c(u_1, u_2, d_1, d_2) \mid u_1, u_2, d_1, d_2 \in \mathbb{N}) \]

Try to identify joint distribution \( F(C) \)
Recall $S(A, E, Z) = \sum_{(u,d) \in A} (e_{(u,d)} + z_{(u,d)})$ governs pairwise stable assignment.

Let $\tilde{S}(A, E) = \sum_{(u,d) \in A} e_{(u,d)}$ 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

$A_1 = \{\langle 1, 1 \rangle, \langle 2, 2 \rangle, \langle 3, 3 \rangle\}$
$A_2 = \{\langle 1, 2 \rangle, \langle 2, 1 \rangle, \langle 3, 3 \rangle\}$
$A_3 = \{\langle 1, 3 \rangle, \langle 2, 2 \rangle, \langle 3, 1 \rangle\}$
$A_4 = \{\langle 1, 2 \rangle, \langle 2, 3 \rangle, \langle 3, 1 \rangle\}$
$A_5 = \{\langle 1, 1 \rangle, \langle 2, 3 \rangle, \langle 3, 2 \rangle\}$
$A_6 = \{\langle 1, 3 \rangle, \langle 2, 1 \rangle, \langle 3, 2 \rangle\}$
Unobserved Heterogeneity in Matching Games
Baseline Model

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

\[ C = (c(u_1, u_2, d_1, d_2) \mid u_1, u_2, d_1, d_2 \in \mathcal{N}) \]

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 $C_1 = C_2$

  \[
  c(1, 2, 1, 2) = \tilde{S}(A_1, C) - \tilde{S}(A_2, C) \\
  c(1, 2, 1, 3) = \tilde{S}(A_5, C) - \tilde{S}(A_6, C) \\
  c(1, 3, 1, 2) = \tilde{S}(A_5, C) - \tilde{S}(A_4, C) \\
  c(1, 3, 1, 3) = \tilde{S}(A_1, C) - \tilde{S}(A_3, C)
  \]

- 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 $H(\tilde{S})$ of unobserved production for all assignments is identified.
Proof: Identifying $H (\tilde{S})$ Using $Z$

- Recall $S (A, E, Z) = \sum_{\langle u, d \rangle \in A} (e_{\langle u, d \rangle} + z_{\langle u, d \rangle})$ governs pairwise stable assignment
Proof: Identifying $H(\tilde{S})$ Using $Z$

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

- Each $E^*$ gives one $C^*$ & one $\tilde{S}^* = \tilde{S}(A, C^*)$, set $Z_{\langle u, d \rangle}^* = -e_{\langle u, d \rangle}^*$.
Proof: Identifying $H\left(\tilde{S}\right)$ Using $Z$

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

- Each $E^*$ gives one $C^*$ & one $\tilde{S}^* = \tilde{S}\left(A, C^*\right)$, set $z_{\langle u, d \rangle}^* = -e_{\langle u, d \rangle}^*$.

- Then $S\left(A, E^*, Z^*\right) = \tilde{S}\left(A, C^*\right) + \sum_{\langle u, d \rangle \in A} z_{\langle u, d \rangle}^* = 0 \forall A$. 

Definition of the CDF $H\left(\tilde{S}\right)$:

$H\left(\tilde{S}\right) = \Pr\left(\tilde{S}\left(A, C\right) \leq \tilde{S}\left(A, C^*\right), \forall A \neq A\right)$.
Proof: Identifying $H\left(\tilde{S}\right)$ Using $Z$

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

- Each $E^*$ gives one $C^*$ & one $\tilde{S}^* = \tilde{S}\left(A, C^*\right)$, set
  $$z_{\langle u,d \rangle}^* = -e_{\langle u,d \rangle}^*$$

- Then $S\left(A, E^*, Z^*\right) = \tilde{S}\left(A, C^*\right) + \sum_{\langle u,d \rangle \in A} z_{\langle u,d \rangle}^* = 0 \forall A$

- Definition of the CDF
  $$H\left(\tilde{S}^*\right) = \Pr\left(\tilde{S}\left(A, C\right) \leq \tilde{S}\left(A, C^*\right), \forall A \neq A_1\right)$$
Proof: Identifying $H(\tilde{S})$ Using $Z$

\[
H(\tilde{S}^*) = \Pr(\tilde{S}(A, C) \leq \tilde{S}(A, C^*), \forall A \neq A_1)
\]
\[
= \Pr(S(A, E, Z^*) \leq S(A_1, E, Z^*), \forall A \neq A_1)
\]
\[
= \Pr(S(A, E, Z^*) \leq 0, \forall A \neq A_1)
\]
\[
= \Pr(A_1 | Z^*)
\]

- Third equality uses choice of $Z^*$:
- Uses $\Pr(A_1 | Z^*)$ for arbitrary assignment $A_1$, many $Z^*$
Large and product support on $Z$ traces CDF of sums of unobserved production of assignments

- Special regressors
- Failure of large and product support gives partial identification of $H\left(\tilde{S}\right)$ and hence $F\left(C\right)$
Large and product support on $Z$ traces CDF of sums of unobserved production of assignments

- Special regressors
- Failure of large and product support gives partial identification of $H(\tilde{S})$ and hence $F(C)$

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

1. Matching Empirical Program
2. Baseline Model
3. Model Variants
   - Other Observed Characteristics
   - Data on Unmatched Firms
   - Agent-Specific Characteristics
   - One-Sided Matching
   - Many-to-Many Matching
Other Observed Characteristics

- 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 $X$
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 \(d\)

\[X = \left( N, (x_u)_{u \in N}, (x_d)_{d \in N}, (x_{\langle u,d \rangle})_{u,d \in N} \right)\]
More on Example with $X$

- **Total match production**

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

- **Now define**

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

and

\[
c(u_1, u_2, d_1, d_2) \equiv e_{u_1,d_1} + e_{u_2,d_2} - e_{u_1,d_2} - e_{u_2,d_1}
\]
Previous theorems did not use $X$, can condition on $X$

Example model makes the distribution $F(C \mid X)$ of

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

depend on $X$

- Still require independence of $Z$ and $\psi$

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

\[ \{ \langle u_1, 0 \rangle, \langle 0, d_1 \rangle, \langle u_2, d_2 \rangle \} \]
Unobserved Heterogeneity in Matching Games
Model Variants
Data on Unmatched Firms

Unmatched Has 0 Production

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

$$E = \begin{pmatrix}
  e_{1,1} & \cdots & e_{1,N_d} \\
  \vdots & \ddots & \vdots \\
  e_{N_u,1} & \cdots & e_{N_u,N_d}
\end{pmatrix}$$
Unobserved Heterogeneity in Matching Games

Model Variants

Data on Unmatched Firms

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 $E^*$, set $z^*_{ud} = -e^*_{ud}$
  - Then the production of all assignments is 0
  - All agents indifferent between being unmatched and matched
Proof: $G(E)$ is Identified

- Fix $E^*$, set $z^*_{u,d} = -e^*_{u,d}$
  - 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 $e_{u,d} \leq e^*_{u,d} \forall (u, d)$
Proof: $G(E)$ is Identified

- Fix $E^*$, set $z^*_{\langle u, d \rangle} = -e^*_{\langle u, d \rangle}$
  - 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 $e_{\langle u, d \rangle} \leq e^*_{\langle u, d \rangle}$ $\forall \langle u, d \rangle$

- Then
  $$G(E^*) = Pr(E \leq E^* \text{ elementwise}) = Pr(A_0 | Z^*)$$
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 $z_{u,d}$
- Now *agent-specific* regressors $z_u$ and $z_d$
- $2 \cdot N$ such regressors

$$Z = ((z_u)_{u \in N}, (z_d)_{d \in N})$$
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
With data on unmatched firms, can get at distribution $G(E)$ of

$$E = \left( (e_u)^N_{u=3}, (e_d)^N_{d=2} \right).$$

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.*
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
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