Analysis with data on final matches

Nikhil Agarwal

MIT and NBER

Role of Using Choice Models

✓ Preferences are the primitives in the theory of matching markets

- Positive Analysis:
  - Quantifying preferences/ammenities
  - Effects of market interventions are intermediated through agent choices
    - ✓ Taxes, tuition subsidies, free tuition, quotas
    - ✓ Preference estimates facilitate General Equilibrium policy analysis

- Normative Analysis:
  - Welfare and distributional consequences

✓ Complementary to theory in evaluation of trade-offs

- Magnitudes of effects identified in the theory
- Analysis when theory is intractible or ambiguous

Revealed Preference Approach

- Traditional revealed preference approach
  - ✓ Use data on consumer decisions to deduce most preferred option (given price)

- Matching Markets: Cannot choose your preferred option $\rightarrow$ must also be chosen
  - Cannot decide to enroll at any university
  - Your partner needs to agree to marry you
  - Cannot show up at work at Google
  - Peer-to-peer platforms require mutual consent (eg. AirBnb)

Revealed Preference Approach

✓ Rules of the market determine the interpretation of the data

- Matched partner need not be preferred to others
- College application decisions consider chances of admission
- Agents need not submit a truthful ranking

✓ Organized marketplaces present a unique opportunity for analysis

- Administrative data on outcomes and/or submitted rankings
- Well understood rules of the game assist modeling choices

Using Final Match Data

- School choice models use data on reported preferences

1. Many well-functioning markets do not use centralized systems
2. Barriers to obtaining reported preferences
   - Rank-ordered data are not collected (e.g. decentralized implementation)
   - Confidentiality concerns (NRMP)

- Early work was on marriage markets [Chiappori et al, 2012; Dagsvik, 2000]

- Key problem: Final matches depend on two sets of preferences

Outline

Outline

1. Non-Transferable Utility Models: Empirical Framework
   - An aspirational framework
   - “Double-Vertical” Model
   - Separable and Idiosyncratic Heterogeneity
2. Application: The Medical Match
   - Equilibrium Model and Data
   - Empirical Strategy
   - Results and Counterfactual
   - Application: Salary Competition
3. TU Models

(Aspirational) Empirical Framework

Borrows from survey in Agarwal and Somaini (2021)

- Two-sided matching market
  - Agents indexed by $i \in \mathcal{I}$ on side 1 and $j \in \mathcal{J}$ on side 2
  - Agents i may be matched with at most one agent in $\mathcal{J}$
  - Agents j may be matched with up to q_j agents
- Preferences (in their most general form)
  - Indirect utility of i for matching with j is given by u_{ij}. e.g.

$$u_{ij} = u(\boldsymbol{x}_j, \boldsymbol{z}_i, \xi_j, \varepsilon_i) - d_{ij}$$

  - Similarly, utility of j for matching with i is given by

$$v_{ji} = v(\boldsymbol{x}_j, \boldsymbol{z}_i, \eta_i) - w_{ji}$$

  - Typically assume the independence condition

$$(\varepsilon_i, \eta_i) \perp (\boldsymbol{d}_i, w_i) | \boldsymbol{z}_i, \boldsymbol{x}, (\xi_j)_{j=1}^{J}$$

- Equilibrium: Pairwise stable matching

Important Assumptions

1. No externalities
   - ✓ Utility only depends on who you match with
   - Difficulty in ensuring existence of stable matching [Extensions in Sasaki and Toda, 1996; Pycia and Yenmez, 1997]
   - Rules out peer-effects and preferences based on post-match competition
   - Recent advances make some progress [Uetake and Watanabe, 2019; Vissing, 2018]
2. No frictions in matching
   - Full information
   - Well-formed preferences [see Narita, 2018, for an exception]
3. Exogeneity of observables (orthogonality)
   - Problematic if counterfactuals that affect incentives for chosing characteristics
4. Transfers, if any, are fixed/not negotiated
   - Related models with fully or imperfectly transferable utility [Choo and Siow, 2006; Galichon and Salanie, 2021; Galichon et al., 2019]

Outline

1. Non-Transferable Utility Models: Empirical Framework
   - An aspirational framework
   - "Double-Vertical" Model
   - Separable and Idiosyncratic Heterogeneity

2. Application: The Medical Match
   - Equilibrium Model and Data
   - Empirical Strategy
   - Results and Counterfactual
   - Application: Salary Competition

3. TU Models

Empirical Content of Pairwise Stability

- How do we learn about preferences from the data?
  - Key problem: Final matches depend on two sets of preferences

- Intuition using simple model with no preference heterogeneity
  → Identification with "double-vertical" preferences

$$u_j = u(x_j) + \xi_j$$
$$v_i = v(z_i) + \eta_i$$

  → Perfect assortative matching on u and h
  1. Information in sorting patterns [Chiappori et al., 2012: Diamond and Agarwal, 2017]
  2. Necessity of using many-to-one matching structure [Diamond and Agarwal, 2017]

- Extension to heterogeneity in preferences recently studied [He, Sinha and Sun, 2022; Agarwal and Somaini, 2022]
  ✓ Stay tuned during the conference

Sign Restriction

- Sorting patterns summarized by F_{XZ}: Contingency table w/ binary characteristics
  - z denotes large or small hospital; x denotes high or low funding

Resident	Program Characteristic	
Characteristic	Large	Small
High	30%	20%
Low	20%	30%

- Need a sign restriction on one characteristic
  - Without this restriction, both characteristics could be undesirable
- Assumption: Residents from medical schools with higher NIH funding are more likely to have higher human capital ($\alpha_{NIH} > 0$)
  - Sorting indicates that larger hospitals are prefered ($\beta_{LARGE} > 0$)

Limitation of Sorting Patterns

Resident Characteristic	Program Characteristic: Large	Program Characteristic: Small
High	30%	20%
Low	20%	30%

- Cannot learn about preferences on both sides from sorting patterns alone
  - Consistent a strong preference for large hospitals + moderate association between high NIH funding and resident skill
  - Cannot distinguish from the reverse

$$u_j = x_j\beta + \xi_j$$
$$v_i = z_i\alpha + \varepsilon_i$$

  - Degree of sorting on observables increases with both α and β
    - Large β and small α vs. large α and small β

✓ Non-parametric version: quantile-quantile matching implies

$$u(\boldsymbol{x}_j) = F_U^{-1}(F_V^{-1}(v(z_i) + \eta_i) - \xi_j$$

Usefulness of Data from Many-to-One Matching

- Do residents matched at the same program have similar characteristics?

- Two residents matched at the same program must be similarly qualified
  - Otherwise, program or resident can find a better match

- Residents at a program have similar values of z if it strongly predicts human capital $\Rightarrow$ small within-program variation

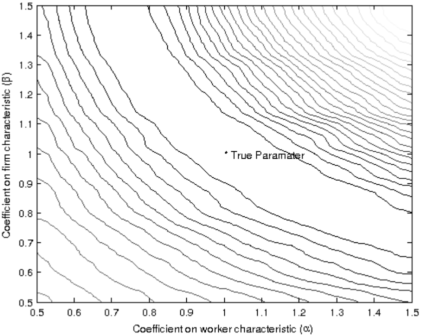
- Provides crucial information that is not available in one-to-one matching
  - Combine with sorting patterns to learn about preferences on both sides

✓ Multiple matches can be seen as noisy measures [Hu and Schennach, 2008; Diamond and Agarwal, 2017]

$$u(\boldsymbol{x}_j) = F_U^{-1}(F_V^{-1}(v(z_i) + \eta_i) - \xi_j$$
$$u(\boldsymbol{x}_j) = F_U^{-1}(F_V^{-1}(v(z_{i'}) + \eta_{i'}) - \xi_j$$

Sorting Patterns: Objective Function

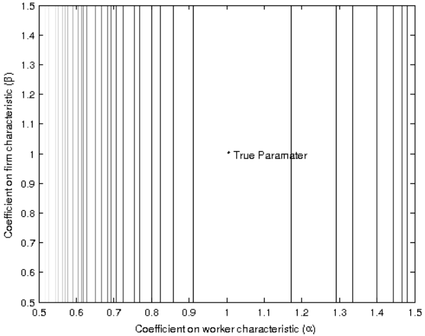
$$u_j = z_j\beta + \xi_j$$
$$v_i = x_i\alpha + \varepsilon_i$$



- Level sets of sorting moments across large β and small α

Within Program Moments: Objective Function

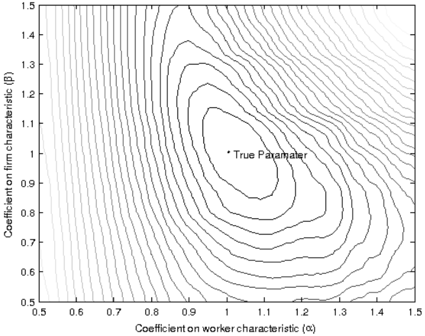
$$u_j = x_j\beta + \xi_j$$
$$v_i = z_i\alpha + \varepsilon_i$$



- Within program variation changes only with α

Within Program + Sorting: Objective Function

$$u_j = x_j\beta + \xi_j$$
$$h_i = z_i\alpha + \varepsilon_i$$



Outline

Separable and Idiosyncratic Heterogeneity

- One to one matching model of Menzel (2015)

$$
\begin{aligned}
u_{ij} &= u(x_i, z_j) + \varepsilon_{ij} \\
v_{ji} &= v(x_i, z_j) + \eta_{ji} \\
\frac{f(x,z)}{f(*,z)f(x,*)} &= \exp(u(x,z) + v(x,z))
\end{aligned}
$$

where $\varepsilon_{ij}, \eta_{ji} \sim$ Type 1 EV

- Notes
  - ✓ Tractable!
  - Generalizes to distributions with tail behavior similar to Type 1 EV
  - Reinforces the idea that one-to-one models are under-identified

Extensions and Variations

- Restricted transfers/moment inequalities [Uetake and Watanabe, 2019]
  - Revealed preference inequality derived from no blocking conditions
  - No "structural" errors
- Political mergers (one-sided matching) [Weese, 2015; Gordon and Knight, 2009]
- Matching with Nash Bargaining over surplus [Sorenson, 2007]

$$S(x_j, z_i) + \eta_{ij} \text{ split } \lambda, 1 - \lambda$$

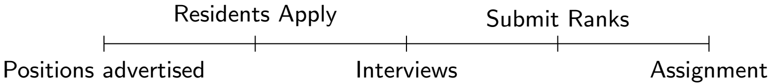
✓ Likelihood based methods

Outline

- Clearinghouse uses rank-order lists and the Roth-Peranson algorithm
- Outcomes in matching markets result from two-sided preferences

National Residency Matching Program

- Centralized assignment process (Roth-Peranson algorithm)



Research Objectives

Methods:

1. Develop a method for estimating preferences using only final matches
   - Employer-employee matched data or school enrollment records are common

Policy Analysis:

2. How do government regulations affect the assignments in rural programs?
   - Study both supply regulations and financial incentives
   - ✓ Estimating primitives allows analyzing important general equilibrium effects

3. Why are medical residents' salaries lower than substitute labor?
   - An antitrust lawsuit and research have questioned the role of the match
   - Analyze salary depression in a counterfactual without the match

Outline

1. Non-Transferable Utility Models: Empirical Framework
   - An aspirational framework
   - "Double-Vertical" Model
   - Separable and Idiosyncratic Heterogeneity
2. Application: The Medical Match
   - Equilibrium Model and Data
   - Empirical Strategy
   - Results and Counterfactual
   - Application: Salary Competition
3. TU Models

Equilibrium Concept

Residents

Programs — Interviews, Ranks — Stable Match

Pairwise Stability

1. IR: Each program is assigned no more than its capacity
2. IC: No resident prefers a program that prefers that resident to an assigned resident (at fixed salaries)

Discussion

- Mechanism implements a stable match with respect to reported ranks
- Strategic interviewing/ranking can result in violations
  - Low frictions in this market: ~ 8 interviews per position

Family Medicine: Data

- Data from annual census of programs matched with residents (AMA/AAMC)
  - Estimation: 2003 - 2004 to 2010 - 2011; Out-of-sample: 2011 - 2012
  - Multiple years are used only to improve precision → data from large markets
- Residents birth location and medical school
  - For MDs, merge with medical school characteristics
- Extensive set of characteristics for programs
  - Program setting, affiliated hospitals and medical schools and location

Residents	All
	Mean
Allopathic/MD	45%
Osteopathic/DO	14%
Foreign Graduate	41%

Programs	All		Rural
	Mean	Std	Mean
First Year Salary	\$46,394	\$3,239	\$46,259
Positions	7.57	2.77	5.25
Matches	7.01	2.92	4.72

Pure Characteristics Model: Berry and Pakes (2007)

$$u_{ijt} = z_{ijt}\beta_i + \delta w_{jt} + \xi_{jt}$$

i :resident j :program t :market

- Utility micro-founded on finitely many program and resident characteristics
  - Unobserved heterogeneity through research focus, size and diagnostic mix
  - Allows for unmeasured program quality (faculty and resources)

z_{ijt}	Pgm. Chars.	$\rightarrow$	NIH Funding (Major and Minor affiliates), Beds Case Mix, Rent, Wage Index, Program Types
	Geo. Het.	$\rightarrow$	Birth/Med school state and Rural-born x Rural Program
β_i	Unob. Het.	$\rightarrow$	NIH, Beds and Case Mix via normally distributed random-coefficients with estimated variance
w_{jt}	Salary		$\xi_{jt} \perp w_{jt}$ (relaxed in paper)
ξ_{jt}	Unob. Quality		$\xi_{jt} \sim N(0,1)$

Program Preferences: Human Capital

- Model program preferences using human capital index

$$h_i = x_i\alpha + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma_{x_i})$$

x_i	Medical school characteristics, degree type, US Born (Foreign Grads)
σ_{x_i}	Depends on degree type, normalized to 1 for MD

- Program directors refer to "pecking order"

  Additional Benefits/Properties

- Implies uniqueness of stable match (Clark, 2006; Niederle and Yariv, 2009)
  - Multiplicity may not be empirically important (Roth and Peranson, 1999)
  - Computational simplicity is an additional benefit

Outline

Identification

- Addresses salary endogeneity using control function [Imbens and Newey, 2009]
- Heterogeneity in resident preferences identified via exclusion restrictions
  - Preference shifters for only one side of the market
  - ✓ Isolates source of sorting patterns
  - Instrumental variables intuition for simultaneous equations

Empirical model uses two restrictions

1. Birth/medical school state excluded from resident desirability
   - Learn about geographic preferences using sorting of medical school classmates born in different locations
2. Determinants of human capital index excluded from resident preferences
   - Higher quality residents choose ahead of those with less desirable traits

Estimation: Moments

- Data is a match $\mu \to$ Resident i's match: $\mu(i)$; Program j's matches: $\mu^{-1}(j)$

1. Moments from sorting patterns

$$\frac{1}{N}\sum_i x_i z_{\mu(i)}$$

2. Within-program variance of resident characteristics

$$\frac{1}{N}\sum_i \left(x_{1,i} - \overbrace{\frac{1}{\left|\mu^{-1}(\mu(i))\right|} \sum_{i' \in \mu^{-1}(\mu(i))} x_{1,i'}}^{\text{Mean } x_1,\ i'\text{s program}} \right)^2$$

3. Peer based moments

$$\frac{1}{N}\sum_i x_{1,i} \overbrace{\frac{1}{\left|\mu^{-1}(\mu(i))\right| - 1} \sum_{i' \in \mu^{-1}(\mu(i)) \setminus \{i\}} x_{2,i'}}^{\text{Mean } x_2,\ i'\text{s peers}}$$

Simulating Matches/Moments

- SMD needs a procedure for simulating equilibrium match for any parameter θ

1. Simulate human capital index $\{h_i\}$

$$h_i = x_i \alpha + \varepsilon_i$$

2. Simulate preferences of residents

$$u_{ij} = z_{ij} \beta_i + \delta w_j + \xi_j$$

3. Calculate the (simulated) pairwise stable match
    - Step 1 : Assign top resident to their first choice
    - Step k : Assign k-th resident to most preferred choice with unfilled positions

- Pairwise Stable: A resident can only envy the assignment of a more qualified resident

- Use S simulated matches to compute simulated moments $\hat{m}^S(\theta)$

Simulated Minimum Distance

- The simulated minimum distance estimate, $\hat{\theta}_{SMD}$ minimizes criterion

$$\left\| \hat{m} - \hat{m}^S(\theta) \right\|_W = \left(\hat{m} - \hat{m}^S(\theta) \right)' W \left(\hat{m} - \hat{m}^S(\theta) \right)$$

$\hat{m}$	Sample moment
$\hat{m}^S$	Simulated counterpart
W	Positive definite weight matrix

  - $\hat{m}^S$ and $\hat{m}$ are averaged across years and individual matches
  - $\hat{\theta}_{SMD}$: Parameter generating the best fit for sorting and many-to-one moments

- Confidence sets need to account for dependence of matches
  - Parametric bootstrap used to compute covariance of moments
  - Delta method to get standard errors in parameters

Outline

Estimates: Resident Preferences

Select Variables	Full Heterogeneity (1)	Geographic Heterogeneity (2)	Geo. Het. w/ Instrument (3)
Case Mix Index (1 sd.)	\$4,792	\$2,320	\$6,088
Random Coeff. (sigma)	\$4,503		
Log NIH Fund (Major) (1 sd.)	\$491	\$6,499	\$4,402
Random Coeff. (sigma)	\$5,498		
Log Beds (1 sd.)	\$6,900	\$3,528	\$8,837
Random Coeff. (sigma)	\$11,107		
Log NIH Fund (Minor) (1 sd.)	\$4,993	\$5,560	\$7,620
Medical School State	\$9,820	\$2,302	\$4,529
Birth State	\$6,342	\$1,320	\$2,451
Rural Birth x Rural Program	\$1,189	\$109	\$233

Estimates: Willingness to Pay

- Large willingness to pay for more desirable programs
  - Estimated standard deviation in utility of $\sim \$14,000$ to $\sim \$28,000$
- Larger for models using wage instruments, but imprecisely estimated
  - Decline in co-efficient on salaries $\rightarrow$ indicates positive correlation between w_{jt} and ξ_{jt}
- Mean utility from rural hospitals is lower, but not economically large

	Full Heterogeneity (1)	Geographic Heterogeneity (2)	Geo. Het. w/ Instrument (3)
Std. Dev in Utility (Across Programs)	\$21,937	\$14,088	\$28,577
	(5,215)	(1,880)	(8,166)
Mean Utility of Rural Programs	-\$7,292	-\$4,692	-\$8,066
	(3,101)	(967)	(4,044)
Mean Utility of Urban Programs	\$1,259	\$810	\$1,392
	(535)	(167)	(698)

Estimates: Human Capital

- Similar coefficient estimates on medical school prestige indicators
- Unobserved characteristics have larger variance for foreign medical graduates

$$h_i = \alpha x_i + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma_{x_i})$$

	Full Heterogeneity (1)		Geographic Heterogeneity (2)		Geo. Het. w/ Instrument (3)	
	Est.	(s.e.)	Est.	(s.e.)	Est.	(s.e.)
Log NIH Fund (MD)	0.115	(0.016)	0.127	(0.014)	0.094	(0.013)
Median MCAT Score	0.081	(0.007)	0.067	(0.004)	0.041	(0.003)
σ_{MD}	1	—	1	—	1	—
σ_{DO}	0.884	(0.036)	0.794	(0.029)	0.728	(0.029)
$\sigma_{Foreign}$	3.619	(0.110)	3.071	(0.072)	2.821	(0.072)
Parameters	25		22		24	
Moments	106		106		118	

Outline

Is the match responsible for low salaries?

- In 2002 former residents alleged a price-fixing conspiracy
  - "The NRMP matching program has the purpose and effect of depressing, standardizing and stabilizing compensation [...] below competitive levels." - Jung et al. v. AAMC et al. (2002).
  - Reasoned that inflexible salaries is a restraint to competition $\rightarrow$ Residents cannot use multiple offers and wage bargaining
- Plaintiffs suggested perfect competition as the alternative
  - Used salaries of physician assistants as a proxy for resident productivity
  - Ignores entry barriers (accreditation, fixed costs) and program heterogeneity

Frictionless Decentralized Market

Competitive Equilibrium

- Assignment of residents to programs and resident-program specific salaries
- Equilibria correspond to core allocations: Shapley and Shubik (1971)
    1. The allocation is individually rational
    2. No program-resident pair would prefer recontracting (with flexible salaries)
    - Further negotiations cannot be mutually beneficial

Frictionless Decentralized Market

Competitive Equilibrium

- Assignment of residents to programs and resident-program specific salaries
- Equilibria correspond to core allocations: Shapley and Shubik (1971)
  1. The allocation is individually rational
  2. No program-resident pair would prefer recontracting (with flexible salaries)
  - Further negotiations cannot be mutually beneficial

Illustrative Model

- N residents with human capital h_i and N programs with quality q_j
  - Resident value program quality and wage $\rightarrow$ extends Bulow and Levin (2006)

$$u = aq + w$$

  - Program-resident pair produce output $f(h, q) \geq 0$, where $f_h,\ f_q,\ f_{hq} \geq 0$. Profit is

$$f(h, q) - w$$

  - Each program hires at most one resident
- Important features are capacity constrains and heterogeneity in types
  - Entry barriers include accreditation requirements and fixed costs

Implicit Tuition

- If $\mu(i)$ is i's equilibrium match, salaries are bounded above by

$$\overbrace{f\left(h_i, q_{\mu(i)}\right)}^{\text{Output net of costs}} \quad - \quad \overbrace{a q_{\mu(i)}}^{\text{Implicit Tuition}}$$

1. Results due to residents' willingness to pay for quality and capacity constraints
2. Higher at more desirable programs $\rightarrow$ compensating differentials
3. Lower bound for depression of salaries from marginal productivity

Implicit Tuition

- If $\mu(i)$ is i's equilibrium match, salaries are bounded above by

$$\overbrace{f\left(h_i, q_{\mu(i)}\right)}^{\text{Output net of costs}} - \overbrace{aq_{\mu(i)}}^{\text{Implicit Tuition}}$$

  1. Results due to residents' willingness to pay for quality and capacity constraints
  2. Higher at more desirable programs $\rightarrow$ compensating differentials
  3. Lower bound for depression of salaries from marginal productivity

- Lowest markdown from output when $f(h, q) = \bar{f}(h) \rightarrow$ Salaries: $w_i = \bar{f}(h_i) - aq_{\mu(i)}$
  - Program profits equal the implicit tuition $\rightarrow$ residents "own" productive input
  - Invariant to choice of $\bar{f}(h_i) \rightarrow$ need not estimate productivity of residents
  - Depends only on resident willingness to pay for programs and positions offered

Implicit Tuition: Estimates

- Estimated average implicit tuition is between $22,500 and $43,500
  - Current salaries paid to residents is $47,000
  - Median pay for physician assistants is about $86,000

	Full Heterogeneity (1)	Geographic Heterogeneity (2)	Geo. Het. w/ Instruments (3)
Mean	$23,803	$22,628	$43,470
	(5,526)	(3,496)	(13,678)
Median	$21,263	$21,168	$40,607
	(5,077)	(3,266)	(12,848)
Standard Deviation	$16,661	$12,278	$24,792
25th Percentile	$11,649	$14,070	$24,853
75th Percentile	$31,467	$28,902	$58,355
95th Percentile	$55,280	$45,785	$92,343

- Salary depression may be caused by a limited supply of heterogeneous positions
  - Implicit tuition may explain low salaries without a match observed in Niederle and Roth (2003, 2009)

Outline

- Literature built on Choo and Siow (2006) with discrete observable types

$$\begin{aligned} V_{i,x,y} &= \alpha_{x,y} - \tau_{x,y} + \varepsilon_{i,x,y} \\ U_{j,x,y} &= \gamma_{x,y} + \tau_{x,y} + \varepsilon_{j,x,y}, \end{aligned}$$

where $\varepsilon \sim$ Type 1 EV

  - Demand and supply

$$\begin{aligned} \ln \mu^d_{x,y} - \ln \mu^d_{x,0} &= \alpha_{x,y} - \alpha_{x,0} - \tau_{x,y} \\ \ln \mu^s_{x,y} - \ln \mu^s_{0,y} &= \gamma_{x,y} - \gamma_{0,y} + \tau_{x,y} \end{aligned}$$

  - Equilibrium match probabilities $\mu_{x,y}$

$$\ln \Pi_{x,y} \equiv \frac{\alpha_{x,y} - \alpha_{x0} + \gamma_{x,y} - \gamma_{0y}}{2} = \ln \mu_{xy} - \frac{\ln \mu_{0y} + \ln \mu_{x0}}{2}$$

- Extensions by Galichon and Salanie (2010) to other functional forms
- Imperfectly transferable utility by Galichon, Kominers, Weber (2019)

Semi-parametric approaches

- Fox (2010; 2018) develops a semi-parametric approach
  - Upstream/downstream firm pair j and i receive payoffs of $\pi_{ij}^{d} - t_{ij}$ and $\pi_{ij}^{u} + t_{ij}$
  - Total surplus $f_{ij} = \pi_{ij}^{u} + \pi_{ij}^{d}$

- Stability implies efficiency

$$\sum_{ij} \mu'_{ij} f_{ij} \leq \sum_{ij} \mu_{ij} f_{ij}$$

where $\mu_{ij} = 1$ if i is matched with j, and zero otherwise

  - Consider swapping the partners of i and i'. It must be that

$$f_{ij} + f_{i'j'} \geq f_{ij'} + f_{i'j}$$

  - Inequality above depends only on the joint surplus

Maximum Score Estimator

- Fox (2018) maximum score estimator akin to Manski (1975)
  - If there are no unobservables, $x_{ij}\theta + x_{i'j'}\theta \geq x_{ij'}\theta + x_{i'j}\theta$, where $x_{ij}\theta$ is an approximation for f_{ij}
  - Suggests maximizing the objective function

$$S(\theta) = \sum_{i=1}^{N-1} \sum_{i'>i}^{N} 1\left\{x_{ij}\theta + x_{i'j'}\theta \geq x_{ij'}\theta + x_{i'j}\theta\right\}$$

- Note that the unobservables are omitted
  - Need that value of θ that maximizes $S(\theta)$ also maximizes a version with unobservable terms
  - This rank-order property is shown for certain forms in Graham (2011; 2014) and in Fox et al. (2018)