Choices and Outcomes

August 14, 2022

Motivation

- So far: focus on efficiency and utilitarian welfare
  - ✓ Motivates revealed preference approach

- Policymakers also care about other objectives
  - School choice: educational achievement
  - Public housing: socio-economic outcomes, mobility
  - Organs: health care costs, life-years from transplantation (LYFT)

✓ How do choices made in matching mechanisms align with outcomes?

1. Alternative to revealed preference metric for evaluation
2. Incentives for market participants to invest in outcomes (e.g. school choice)

Outline

Empirical Framework

- Notation
  - Objects indexed by $j \in \{1, \ldots, J\}$ assigned to agents indexed by i
  - Outcomes and preferences

$$Y_{i0}, Y_{i1}, \ldots, Y_{iJ}$$
$$V_{i0}, Y_{i1}, \ldots, V_{iJ}$$

  - Agents take actions $D_i = D(V_i)$ in the mechanism
  - Assignment $T_i \in \{0,1\}^J$, with $\sum_j T_{ij} \leq 1$

- Data on $Y_i = \sum_j T_{ij} Y_{ij}$, T_i, D_i and covariates X_i

- Potential for selection into assignment on
  1. Average Y_{ij} across j relative to no assignment, Y_{i0}
  2. Value-added of j: expectation of Y_{ij} given j relative to $E[Y_{i0}]$
  3. Match-specific gains: Y_{iij} relative to the expectation given j

✓ Closely related to generalizations of Roy (1951) and Heckman (1973) selection

Empirical Framework

- Common assumptions used for solving selection problems
  1. Randomness in assignment given preferences

$$Y_{i0}, Y_{i1}, \ldots, Y_{iJ} \perp T_i | V_i, X_i$$

  - Sources of randomness include
    1.1 Reports of other agents in the mechanism
    1.2 Lotteries/tie-breakers
    ✓ Knowing the mechanism is useful for verifying the assumptions

  2. Exclusion restrictions on (some) observables

$$Y_{i0}, Y_{i1}, \ldots, Y_{iJ} | V_i, X_i = x \sim Y_{i0}, Y_{i1}, \ldots, Y_{iJ} | V_i, X_i = x'$$

Outline

One School Lottery Example

Simplified from Walters (2018)

- Do parents demand effective charter schools?

- Applicants to charter schools need to win a lottery
  - Only those that demand charter schools will apply and enroll if they get an offer
  - ✓ Are these students selected?

- Assignment to a charter school is given by

$$T_{i1} = \overbrace{1\{V_i - X_i > 0\}}^{A_i} \times Z_i,$$

where X_{i1} is distance to (closest) charter school and $Z_i = 1$ if i gets an offer

- Important simplifications relative to the paper
  - No enrollment stage
  - Single charter as opposed to portfolio choice
  - No other covariates

✓ Single spell version of Heckman and Navarro (2007)

What does the offer IV estimate?

- Assume that
  1. Distance is excludable from potential outcomes and unobserved preferences

$$(Y_{i1}, Y_{i0}, V_i) \perp X_i$$

  2. Offers are randomly assigned $Z_i \perp (X_i, V_i, Y_{i1}, Y_{i0})$

- Second assumption implies that offer IV conditional on distance estimates

$$E[Y_{i1} - Y_{i0} | V_i > X_i]$$

  ✓ Effect for those that apply

- Extrapolate to $E[Y_{i1} - Y_{i0}]$ using
  - Parametric selection correction [Heckman, 1974]
  - Non-parametric identification at infinity argument [Heckman, 1990]

$$\lim_{x \to x^*} P(A_i = 1 | X_i = x) = 1$$

Outline

- Is the demand for high value-added schools higher?
  - Traditional rationale for school choice improving outcomes [Friedman, 1962; Hoxby, 2003]

- Parents may not be able to distinguish peer quality and school quality

✓ Demand may incentivize screening and cream-skimming instead of quality [e.g. Rothstein, 2006; Barseghyan et al. 2014]

Data and Objective

- Application and outcome data from NYC public HS match 2004 – 2007
  - Uses a variant of the Deferred Acceptance Algorithm [Abdulkadiroglu et al., 2005; 2017]
  - Rank-order list data
  - Merged with outcome data on test scores, graduation, college attendance
- Estimate demand and school value-added, accounting for selection
  - Assess whether peers or quality explains demand

Approach

1. Estimate random utility model of school demand
2. Estimate value-added while correcting for selection via a control function from step 1
3. Summarize relationship between demand and school value-added/peer quality

- Outcome equation

$$Y_{ij} = \alpha_j + X_i'\beta_j + \varepsilon_{ij},$$

with normalizations $E[X_i] = 0$ and $E[\varepsilon_{ij}] = 0$

  ✓ Observe $Y_i = \sum_j T_{ij} Y_{ij}$

- Random utility model with Type I EV

$$U_{ij} = \delta_{c(X_i),j} - \tau_{c(X_i)} dist_{ij} + \eta_{ij},$$

where $c(X_i)$ denotes covariate cells and $dist_{ij}$ is distance

- Rank order list based on truthful reporting

$$R_{ik} = \arg \max_{j \in \mathcal{J} / R_{i,<k}} U_{ij}$$

✓ Correlation between ε_i and $(dist_i, \eta_i)$ can create selection

VAM vs Rank-ordered Control Function

- Approach 1: Standard value-added model assuming only selection on observables

$$E[Y_{ij}|X_i, T_i] = \alpha_j + X_i'\beta_j$$

- Approach 2: Control for selection using rank-order list data

$$E[Y_{ij}|X_i, dist_i, \eta_i, T_{ij}] = \alpha_j + X_i'\beta_j + g_j(dist_i, \eta_i)$$

where $g_j(dist_i, \eta_i) \equiv E[\varepsilon_{ij}|X_i, dist_i, \eta_i, T_{ij}] = E[\varepsilon_{ij}|dist_i, \eta_i]$

  - Assumes that assignment is random conditional on rank-order lists
  - Tie-breakers and priorities unrelated to outcomes

Rank-ordered Control Function

- Recall

$$E[Y_{ij}|X_i, dist_i, \eta_i, T_{ij}] = \alpha_j + X_i'\beta_j + g_j(dist_i, \eta_i)$$

where $g_j(dist_i, \eta_i) \equiv E[\varepsilon_{ij}|X_i, dist_i, \eta_i, T_{ij}] = E[\varepsilon_{ij}|dist_i, \eta_i]$

- Parametrize as

$$E[Y_{ij}|X_i, dist_i, \eta_i, T_{ij}] = \alpha_j + X_i'\beta_j + dist_i'\gamma + \sum_k \psi_k \times (\eta_{ik} - \mu_\eta) + \varphi \times (\eta_{ij} - \mu_\eta)$$

  - Follows Dubins and McFadden (1984)
  - Preference for school k correlated with general ability ψ_k
  - Idiosyncratic preferences correlated with outcomes via φ

$$E[Y_{ij}|X_i, dist_i, \eta_i, T_{ij}] = \alpha_j + X_i'\beta_j + dist_i'\gamma + \sum_k \psi_k \times (\eta_{ik} - \mu_\eta) + \varphi \times (\eta_{ij} - \mu_\eta)$$

- Re-write by taking expectations conditional on R_i

$$E[Y_{ij}|X_i, dist_i, R_i] = \alpha_j + X_i'\beta_j + dist_i'\gamma + \sum_k \psi_k \times \lambda_k(X_i, dist_i, R_i) + \varphi \times \lambda_j(X_i, dist_i, R_i)$$

  where $\lambda_k(X_i, dist_i, R_i) = E[\eta_{ik} - \mu_\eta | X_i, dist_i, R_i]$
  - Two-step estimation with $\lambda_k(X_i, dist_i, R_i)$ estimated using demand model

- Obtain estimates of value-added
  - Bayesian shrinkage to reduce the effect of sample noise

Figure 1: Comparison of value-added and control function estimates

A. Regents math scores

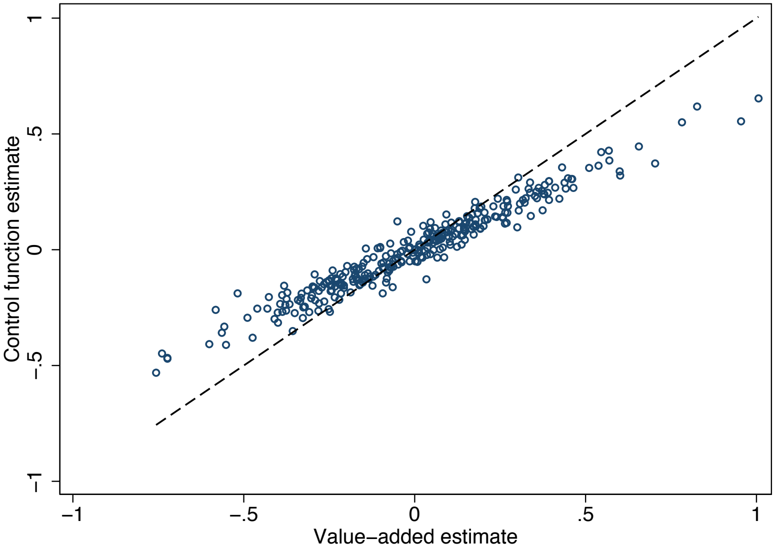


Figure 1: Comparison of value-added and control function estimates

D. College attendance

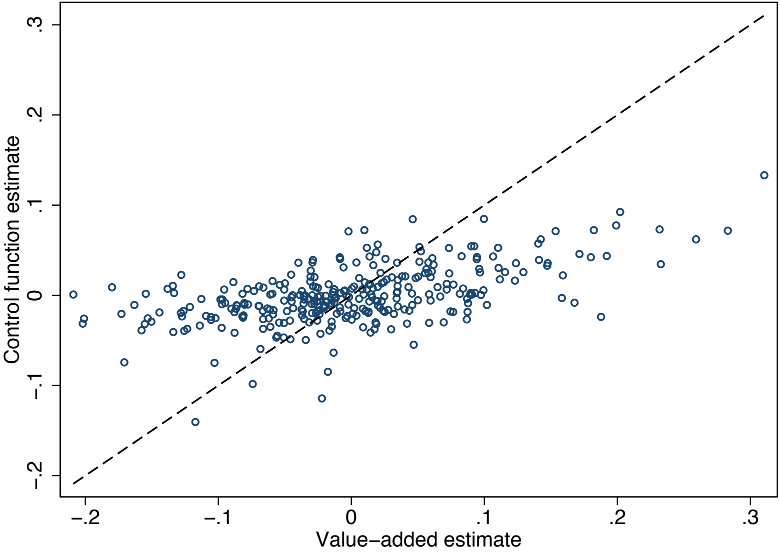


Table 5. Correlations of peer quality and treatment effect parameters for Regents math scores

	Peer quality	Control function parameters							
		ATE	Female	Black	Hispanic	Sub. lunch	Log tract inc.	Math score	Reading score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ATE	0.587								
	(0.052)								
Female	0.079	0.290							
	(0.078)	(0.101)							
Black	0.013	0.106	-0.202						
	(0.076)	(0.104)	(0.139)						
Hispanic	0.005	0.073	-0.301	0.995					
	(0.075)	(0.106)	(0.157)	(0.007)					
Subsidized lunch	0.046	-0.155	0.059	-0.068	-0.073				
	(0.085)	(0.115)	(0.138)	(0.150)	(0.148)				
Log census tract income	0.040	0.049	-0.061	-0.222	-0.210	-0.219			
	(0.099)	(0.134)	(0.161)	(0.174)	(0.173)	(0.182)			
Eighth grade math score	-0.076	0.039	-0.078	-0.029	-0.020	0.051	0.017		
	(0.064)	(0.083)	(0.098)	(0.101)	(0.100)	(0.111)	(0.130)		
Eighth grade reading score	-0.414	-0.455	-0.190	-0.054	-0.032	-0.006	0.094	0.242	
	(0.068)	(0.093)	(0.116)	(0.127)	(0.128)	(0.132)	(0.153)	(0.098)	
Preference coefficient (ψ_j)	0.427	0.245	0.200	-0.093	-0.112	-0.108	0.311	-0.236	-0.281
	(0.063)	(0.092)	(0.104)	(0.104)	(0.105)	(0.116)	(0.131)	(0.084)	(0.099)

Demand for School Quality

- Finally, relate preferences to school effectiveness

$$\hat{\delta}_{cj} = \kappa_c + \rho_1 Q_j^* + \rho_2 ATE_j^* + \rho_3 M_{cj}^* + \xi_{cj}$$

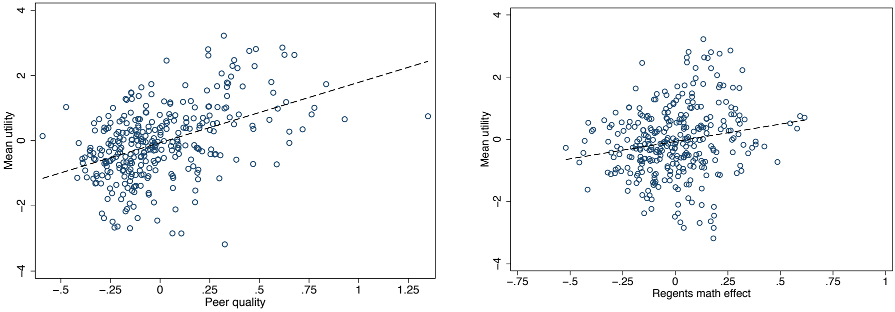
  - κ_c is a covariate cell fixed effect
  - Q_j^* and ATE_j^* are predictions of peer quality and school average treatment effects
  - M_{cj}^* is an estimate of the match effect

Table 7. Preferences for peer quality and Regents math effects

	Control function models			
	(5)	(6)	(7)	(8)
Peer quality	0.410		0.467	0.426
	(0.057)		(0.061)	(0.061)
ATE		0.216	-0.084	-0.081
		(0.047)	(0.046)	(0.045)
Match effect				-0.157
				(0.050)
N	21684			

Figure 3: Relationships between preferences, peer quality, and Regents math effects

A. Bivariate relationships



Outline

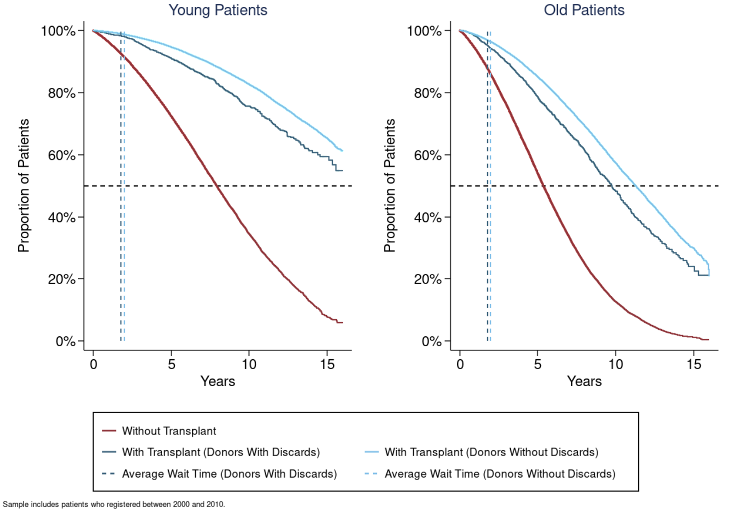
What about the effects on life-years?

Agarwal, Hodgson and Somaini, 2021

- Can we optimize mechanisms to improve matches on alternative metrics?
- Goal: Evaluate a mechanism on the basis of outcomes
  - LYFT from current system
  - Does choice help?
  - Scope for increasing LYFT
  - Distributional consequences of maximizing LYFT
- Challenge: Unlike school seats, every kidney is different
  1. Operationalize a characteristic space approach
  2. Use both randomness in offers and a choice shifter

  ✓ Choice shifter is low-dimensional

Survival Curves



- 15 years of data on patient survival with detailed characteristics

Assignment Mechanism and Observed Outcomes

- (Transplant-specific) Potential outcomes framework. Observe

$$Y_i = \begin{cases} Y_{ij} & \text{if } T_{ij} = 1 \\ Y_{i0} & \text{if } T_{ij} = 0 \quad \forall j \end{cases}$$

  ✓ In our context Y_i is (censored) survival pre- and post-transpant

Assignment Mechanism and Observed Outcomes

- (Transplant-specific) Potential outcomes framework. Observe

$$Y_i = \begin{cases} Y_{ij} & \text{if } T_{ij} = 1 \\ Y_{i0} & \text{if } T_{ij} = 0 \quad \forall j \end{cases}$$

  ✓ In our context Y_i is (censored) survival pre- and post-transpant

- Decisions and Assignments:

time/offers

$D_{i1} = 0$

Assignment Mechanism and Observed Outcomes

- (Transplant-specific) Potential outcomes framework. Observe

$$Y_i = \begin{cases} Y_{ij} & \text{if } T_{ij} = 1 \\ Y_{i0} & \text{if } T_{ij} = 0 \quad \forall j \end{cases}$$

  ✓ In our context Y_i is (censored) survival pre- and post-transpant

- Decisions and Assignments:

time/offers

$D_{i1} = 0 \quad D_{i2} = 0 \quad D_{i3} = 1, T_{i3} = 1$

Assignment Mechanism and Observed Outcomes

- (Transplant-specific) Potential outcomes framework. Observe

$$Y_i = \begin{cases} Y_{ij} & \text{if } T_{ij} = 1 \\ Y_{i0} & \text{if } T_{ij} = 0 \quad \forall j \end{cases}$$

  ✓ In our context Y_i is (censored) survival pre- and post-transpant

- Decisions and Assignments:

$$\xrightarrow{\hspace{12cm}} \text{time/offers}$$

$D_{i1} = 0$ $D_{i2} = 0$ $D_{i3} = 1, T_{i3} = 1$ fourth potential offer

Latent Outcomes and Decisions

- Outcomes from assignment

$$\begin{aligned} Y_{i0} &= g_0(x_i, \nu_{i,0}) \\ Y_{ij} &= g_1(q_j, x_i, \nu_{i,1}, \varepsilon_{ij,1}), \end{aligned}$$

  - generated from survival models with hazard rates
  - ✓ Assumption: no unobserved time-evolving frailty

- Agents are offered objects sequentially and must accept or reject

$$D_{ij} = g_D(q_j, x_i, z_i, \nu_{i,D}, \varepsilon_{ij,D}) \in \{0, 1\}$$

  - E.g. Optimal Stopping (see Agarwal et. al., 2021)

$$D_{ij} = 1\{\Gamma(q_j, x_i, \nu_{i,D}, \varepsilon_{ij,D}) > V(x_i, z_i)\}$$

- Assume: z_i is indep. of unobservables conditional on x_i and $\{q_j\}_j$
  - ✓ One dimensional rather than J-dimensional

Selection

- Allow correlations:
  1. between $\nu_{i,0}$, $\nu_{i,1}$, and $\nu_{i,D}$
  2. between $\varepsilon_{ij,1}$ and $\varepsilon_{ij,D}$

$$
\begin{aligned}
Y_{i0} &= g_0(x_i, \nu_{i,0}) \\
Y_{ij} &= g_1(q_j, x_i, \nu_{i,1}, \varepsilon_{ij,1}), \\
D_{ij} &= g_D(q_j, x_i, z_i, \nu_{i,D}, \varepsilon_{ij,D})
\end{aligned}
$$

- Selection into treatment on
  1. baseline Y_{i0}
  2. average treated outcome $\bar{Y}_i = \frac{1}{J}\sum_j Y_{ij}$
  3. match-specific outcomes $Y_{ij} - \bar{Y}_i$

- Selection due to:
  1. mechanism and choices
  2. observables or unobservables

Two Sources of Variation

1. (Conditional) independence of offers

$$\boldsymbol{\nu}_i, (\varepsilon_{ij})_{j=1,\ldots,J} \perp J_i | x_i$$

   since x_i includes priority type

   - Source: arrival of donors; decisions of other patients

2. Choice shifter z_i excluded from outcomes

   - Source: variation in scarcity of kidneys across region/time

- We combine both sources identify the model:

- Identify the function $g_D(\cdot)$ and the marginal distributions of $Y_{i,j}$ and $Y_{i,0}$ conditional on the vector $(x_i, q_j, z_i, \varepsilon_{i,j,D}, \nu_{i,D})$.

✓ Two dimensions of selection on unobservables - two instruments.

1: Variation in Offers

- Conditional on x_i, J_i is a function of random kidney arrival and the choices of other patients.
- Patients that recieve more good kidney offers are more likely to be transplanted.
- To illustrate this source of variation, regress transplantation on functions of J_i
- In particular, use number of kidneys arriving in the first 2 years after a patient registers for which that patient would be in the top 10 offers.

First Stage: Offer Sequence

	Transplant			
	Any Kidney (1)	Any Kidney (2)	KDPI <= 50% (3)	KDPI > 50% or Missing (4)
log(1 + # Top 10 Offers in 2 Years)				
KDPI <= 50%	0.0322***	0.0334***	0.0439***	-0.0105***
	(0.00441)	(0.00441)	(0.00306)	(0.00287)
KDPI > 50% or Missing	0.0303***	0.0297***	-0.0128***	0.0425***
	(0.00475)	(0.00478)	(0.00314)	(0.00294)
DSA FE, year FE, and blood type FE	x	x	x	x
Control for Pediatric at Listing	x	x	x	x
CPRA Category Controls	x	x	x	x
Patient Characteristics		x	x	x
F-statistic	93.20	92.23	108.0	130.6
Number of Observations	132715	131105	131105	131105

2: Choice Shifter

- Affects choices but not outcomes
  - Use variation in (beliefs about) kidney availability

- Operationalize with a proxy for expected offers/donors
  - Offers made to same blood group in same DSA in the last year.

- Needs to satisfy two requirements:
  - Exclusion: Should not affect outcomes except through choice
    - ✓ Depends on donor arrivals for comparison group
  - Relevance: Should affect which kidney is assigned
    - ✓ Correlated with beliefs about donor arrival frequency

✓ Data shows: better prospects reduce acceptance probability

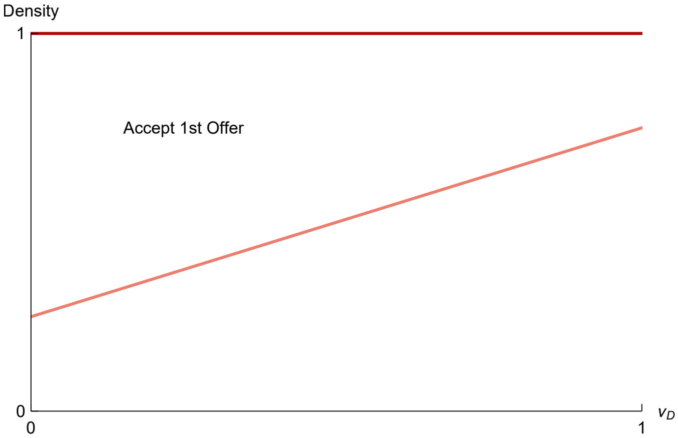
First Stage: Past Donors

	Acceptance						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(1 + No. Donors)	-0.0490***		-0.0479***		-0.0365***		-0.0360***
	(0.00341)		(0.00338)		(0.00324)		(0.00323)
Log(1 + No. Offers)		-0.0536***		-0.0528***		-0.0439***	
		(0.00185)		(0.00183)		(0.00183)	
Offer Year FE	x	x	x	x	x	x	x
Priority Type FE	x	x	x	x	x	x	x
DSA FE and blood type FE	x	x	x	x	x	x	x
Years Waited at Offer FE	x	x	x	x	x	x	x
Patient Characteristics			x	x			x
Donor Characteristics					x	x	x
Match Characteristics					x	x	x
F-statistic	205.8	842.1	200.5	829.8	126.7	575.2	124.2
Number of Observations	912889	912761	912889	912761	900794	900669	900794
R-Squared	0.166	0.172	0.169	0.174	0.263	0.233	0.265

Balance
Monotonicity Donors
Monotonicity Offers

- For simplicity, assume that all kidneys and patients are the same

- Consider the remaining density after first offer

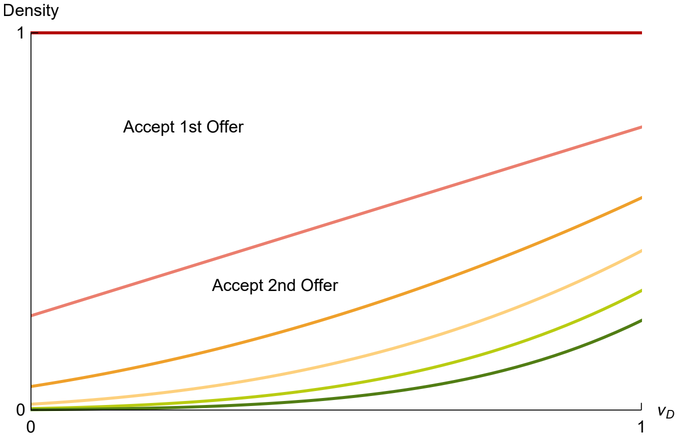


Identification

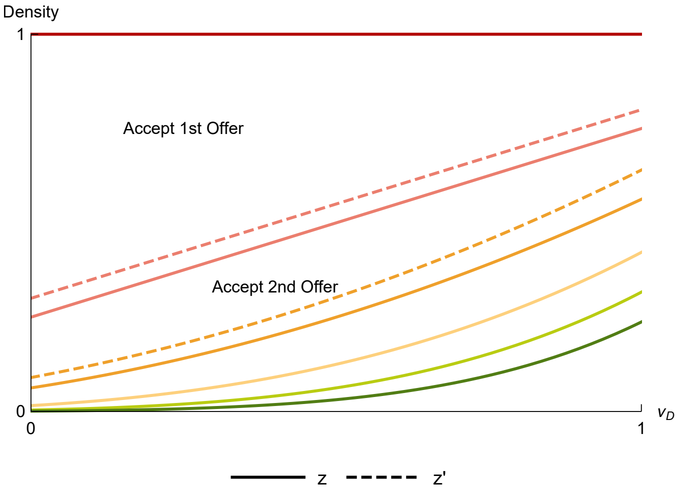
- Observe $E[Y_{ij}|T_{i1} = 1, z_i]$ and $P[T_{i1} = 1|z_i]$
- Infer $E[Y_{i0}|T_{i1} = 1, z_i]$ from those without offers

Identification

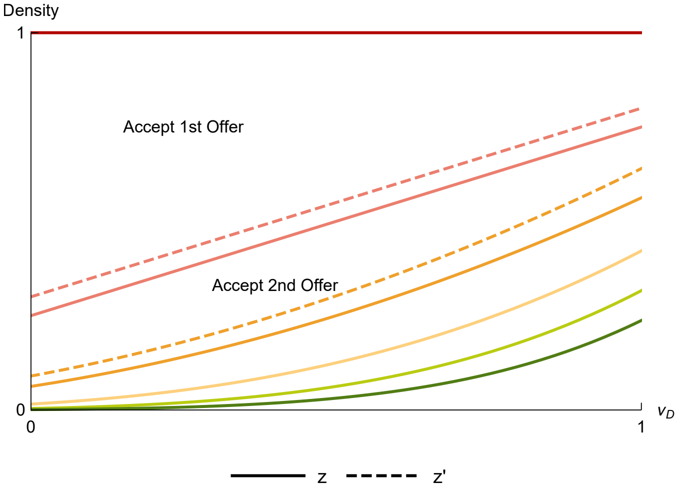
- Obtain $E[Y_{ij}|T_{ik}=1,z_i]$, $E[Y_{i0}|T_{ik}=1,z_i]$ and $P[T_{ik}=1|z_i]$ for any k
- Higher $k \implies$ composition of patients with higher selectivity



- Choice shifter moves set of compliers for each offer instrument



- Theorems combines z_i with offer instrument
- ✓ Disambiguate the effects of $\varepsilon_{ij,D}$ and $\nu_{i,D}$ on expected outcomes



- Parametric survival model and choice equation

$$
\begin{aligned}
y_{i0} &= x_i \beta_x + \nu_{i,0} \\
y_{ij} &= x_{ij} \alpha_x + \alpha_\eta \eta_j + \nu_{i,1} + \varepsilon_{ij,1} \\
D_{ij} &= 1\{x_{ij} \gamma_x + z_i \gamma_z + \eta_j + \nu_{i,D} + \varepsilon_{ij,D} > 0\}
\end{aligned}
$$

where $\eta_j \sim \mathcal{N}(0, \sigma_\eta^2)$,

$$\boldsymbol{\nu}_i \sim \mathcal{N}(0, \Sigma_\nu)$$

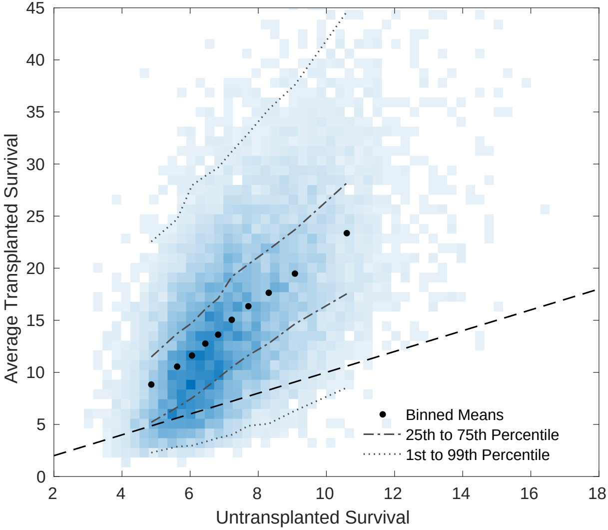
and

$$\boldsymbol{\varepsilon}_{ij} \sim \mathcal{N}(0, \Sigma_\varepsilon)$$

- Box-cox transformed outcomes with truncation

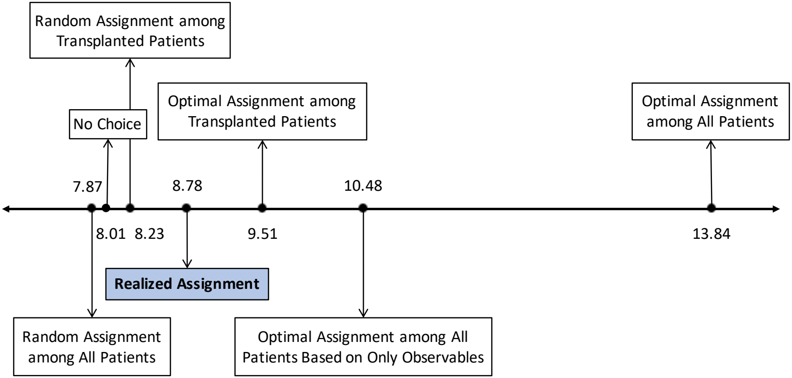
- Estimate parameters via Gibbs Sampler (equivalent to MLE)
  - Factor structure for correlations in $\boldsymbol{\nu}_i$ and $\boldsymbol{\varepsilon}_{ij}$
  - Rich set of donor characteristics $\rightarrow$ too many to include

Transplanted and Untransplanted Survival



Healthier patients have higher LYFT

Increasing LYFT



- Optimal Assignment: maximize total LYFT s.t. resource constraints

✓ Sample: patients that registered in 2005 to ease computation

Planner's Dilemma

	All Patients	Realized Assignment		Optimal Assignment	
		Transplanted Patients	LYFT	Transplanted Patients	LYFT
	(1)	(2)	(3)	(4)	(5)
Age < 18	3.1%	5.4%	14.89	5.4%	18.83
Age 18 - 35	11.6%	13.0%	12.45	16.8%	15.35
Age 36 - 59	54.8%	54.7%	8.86	57.5%	13.70
Age >= 60	30.5%	26.9%	5.61	20.3%	11.68
Diabetic	41.4%	33.3%	6.51	31.6%	12.14
Untransplanted Survival	6.68	6.81	-	7.27	-

- Increasing LYFT requires leaving sicker patients untransplanted
  - Transplanted patients are younger/healthier
  - ✓ Optimal assignment exacerbates differences
- Equity-efficiency trade-off
  - Moral imperative to transplant sicker patients who may soon die
  - Concerns about discriminating based on patient characteristics

Outline

1. Diagnosing Market Failures
   - Market power [e.g. supply reduction in spectrum auctions – Doraszelski et al., 2017]
   - Collusion [e.g. medical match, auctions – Porter and Zona, 1993; Asker, 2010]
   - Rent seeking [e.g. high-frequency trading – Budish et al., 2015]
   - Participation [Bulow and Klemperer, 1996]
   - Market fragmentation [e.g. kidney exchange – Agarwal et al., 2019]
   - Strategic misreporting

Designing Markets

1. Diagnosing Market Failures

2. Evaluating and Comparing Designs
   - School choice design [see Pathak, 2017; Agarwal and Somaini 2020 for surveys]
   - Organ waiting lists [Agarwal et al., 2021]
   - Public housing [Leshno, 2019; Waldinger, 2020]
   - Natural resources [Reeling and Verdier, 2020]
   - Course allocation [Budish and Cantillon, 2012]
   - Food banks [Prendergast, 2020]

Designing Markets

1. Diagnosing Market Failures

2. Evaluating and Comparing Designs

3. Proposing New Designs
   - Medical matching markets [Roth, 1984; Roth and Peranson, 1999]
   - School choice [Abdulkadiroglu and Sonmez, 2003; Pathak, 2011]
   - Course allocation [Budish and Cantillon, 2012]
   - Food banks [Prendergast, 2020]
   - Financial Exchanges [Budish et al., 2015]

The Market Design Agenda

- The goal of a market designer is to analyze rules and develop new ones
  - ✓ A priori unclear which rules are important [Roth, 2002; Klemperer, 2004]

- Theory, data and practice play complementary roles
  - Theory:
    - Identify qualitative trade-offs, analyze new designs
  - Data:
    - Quantify trade-offs, test hypotheses
  - Practice:
    - Identify new issues, bring in new data, uncover new designs

The Market Design Agenda

- The goal of a market designer is to analyze rules and develop new ones
  - ✓ A priori unclear which rules are important [Roth, 2002; Klemperer, 2004]

- Theory, data and practice play complementary roles
  - Theory:
    - Identify qualitative trade-offs, analyze new designs
  - Data:
    - Quantify trade-offs, test hypotheses
  - Practice:
    - Identify new issues, bring in new data, uncover new designs

- Ultimate Goal: Remedy market failures

✓ Partnerships with practitioners is essential at all stages of this research