Optimal transport theory in incomplete econometric models

Marc Henry

Kantorovich Initiative
24 February 2022

Pacific Northwest connection*

The work surveyed is joint with
Victor Chernozhukov, Arthur Charpentier, Arnaud Dupuy,
Ivar Ekeland*, Yanqin Fan*, Alfred Galichon, Marc Hallin,
Lixiong Li, Romuald Méango, Alexei Onatski, Brendan Pass*,
Maurice Queyranne*.

Prologue: OT in Econometrics

- Vector quantiles and vector copulas
  - Quantile: unique increasing map that pushes uniform to P.
  - Vector quantile: unique gradient of convex map that pushes the uniform on $[0,1]^d$ to any given distribution P on $\mathbb{R}^d$, from Brenier-McCann.
  - Vector copula: transform random vectors X and Y to multivariate uniform, in order to distinguish within and between vector dependence.
- Nonparametric identification
  - Identification of $Y = f(X, \epsilon)$ from Brenier-McCann.
  - More generally, uniqueness of OT solutions yields identification results in discrete choice and hedonic models
- Partial Identification and inference
  - Subject of this talk

Plan of Talk

1. Presentation of the model to analyze

2. Motivation and applications

3. A stylized example

4. OT and partial identification

5. OT and inference in partially identified models

6. Perspectives for future work

Empirical Model

- Data: $((Y_1, X_1), \ldots, (Y_n, X_n)))$ with true distribution $P_0^{(n)}$.

- Structural constraints on the data: There is a fixed vector θ and random vectors $\epsilon_1, \ldots, \epsilon_n$ such that the following holds.

  1. Support restriction: $(Y_i, X_i, \epsilon_i) \in G(\theta)$ a.s. for all $i \leq n$.

  2. The $\epsilon_1, \ldots, \epsilon_n$ are independently distributed, $\epsilon_i \sim Q_{\epsilon|X_i;\theta}$.

The map $G_y(\epsilon, x|\theta) := \{(y, x) : (y, x, \epsilon) \in G(\theta)\}$ is multi-valued.

$\Rightarrow$ Hence for a given θ, the model may predict more than one data generating process (incompleteness).

$\Rightarrow$ For a given data generating process $P_0^{(n)}$, there may be more than one value of θ, such that 1&2 hold (partial identification).

Examples

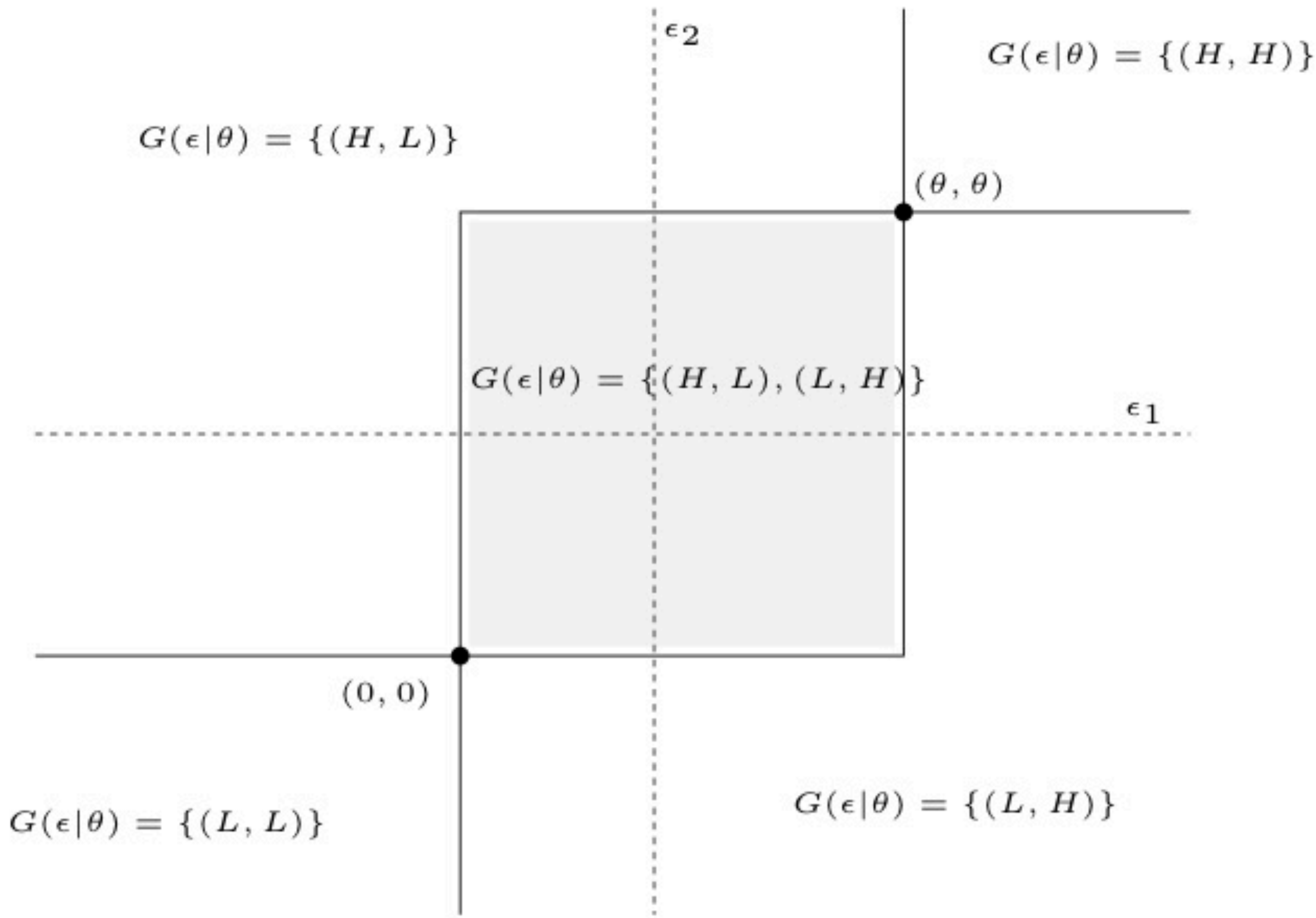
- Game theoretic models of imperfect competition in industrial organization:
  - Economic agents characterized by their unobserved type.
  - Agents maximize payoff, function of the vector of types ϵ, observed X and profile of actions.
  - Chosen profile Y of actions is constrained by an equilibrium concept (e.g., Nash equilibrium).

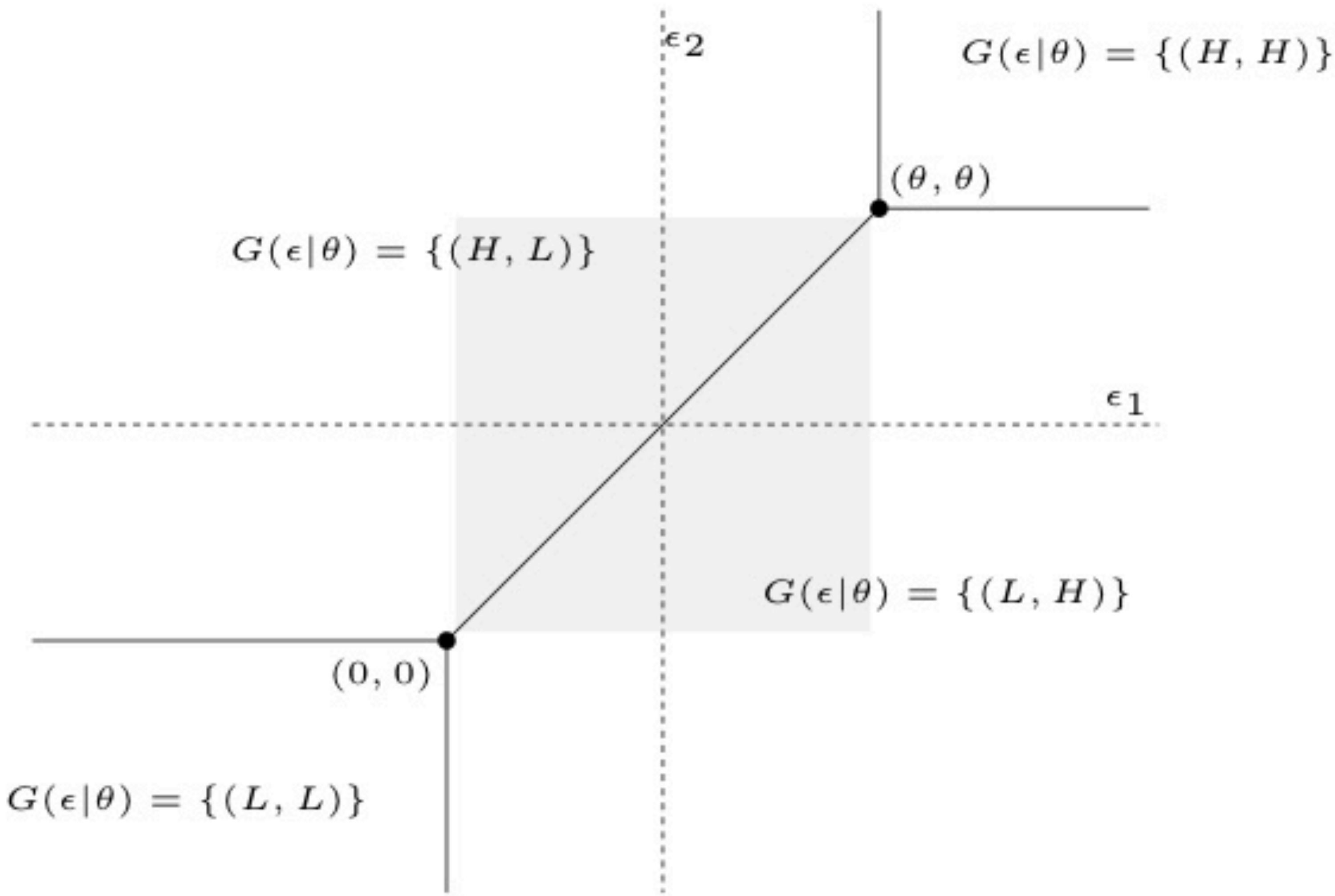
  $\Rightarrow$ Multiplicity of equilibrium causes incompleteness.

- Models of network formation with stability constraints.
- Auctions, where bids only partially reveal unobserved valuations.
- Consumer choice between discrete objects, when the effective choice set in partially known.
- Sample selection, censoring, truncation.

Example of a 2×2 game

		Player 2	
		H	L
Player 1	H	$(\epsilon_1 - \theta, \epsilon_2 - \theta)$	$(\epsilon_1, 0)$
	L	$(0, \epsilon_2)$	$(0, 0)$


ϵ_2
$G(\epsilon|\theta) = \{(H, H)\}$
$G(\epsilon|\theta) = \{(H, L)\}$
(θ, θ)
$G(\epsilon|\theta) = \{(H, L), (L, H)\}$
ϵ_1
$(0, 0)$
$G(\epsilon|\theta) = \{(L, L)\}$
$G(\epsilon|\theta) = \{(L, H)\}$



ϵ_2
$G(\epsilon|\theta) = \{(H, H)\}$
(θ, θ)
$G(\epsilon|\theta) = \{(H, L)\}$
ϵ_1
$G(\epsilon|\theta) = \{(L, H)\}$
$(0, 0)$
$G(\epsilon|\theta) = \{(L, L)\}$

Compatibility of model and observables

- The object of inference is the true value of the parameter θ.
- If more than one value of the vector θ can rationalize the data, it is called *partially identified*.
- The set of values of θ that rationalizes the data is called *identified set* Θ_I.
  - If Θ_I is empty, the model is *misspecified*.
  - If Θ_I is reduced to a point, the model is *point identified*.

$\Rightarrow$ The first objective is an operational characterization of Θ_I.

OT and the characterization of Θ_I

Drop X from notation for simplicity of exposition.

$\Rightarrow$ 1&2 become: $Y \in G_y(\epsilon|\theta)$ and $\epsilon \sim Q_\theta$.

Ignore sampling uncertainty and assume $Y \sim P$ known.

Formalization of the identified set:

$$\Theta_I \;=\; \left\{\theta : \exists \pi \in \mathcal{M}(Q_{\epsilon|\theta}, P) \text{ s.t. } Y \in G_y(\epsilon|\theta), \pi\text{-a.s.}\right\}.$$

Characterization with OT:

- Consider the OT problem with cost $c(\epsilon, Y) := 1\{Y \notin G_y(\epsilon|\theta)\}$,

$$V(\theta) := \min_{\pi \in \mathcal{M}(Q_{\epsilon|\theta}, P)} \mathbb{E}_\pi \left[c(\epsilon, Y)\right].$$

- We see that $\Theta_I = \{\theta : V(\theta) = 0\}$.

Kantorovich duality and Choquet capacity functionals

By the Kantorovich Duality Theorem,

$$\begin{aligned} V(\theta) &= \sup_{f,g} \mathbb{E}_{Q_{\epsilon|\theta}}[f(\epsilon)] + \mathbb{E}_P[g(Y)], \\ &\qquad\qquad \text{s.t.} f(\epsilon) + g(y) \le 1\{y \notin G_y(\epsilon|\theta)\}. \end{aligned}$$

With indicator cost functions, the supremum is achieved with indicator f and g, and it can be shown that

$$V(\theta) = \sup_B \big(c_\theta(B) - P(B)\big),$$

where

$$c_\theta(B) := Q_{\epsilon|\theta}\big(G_y(\epsilon|\theta) \cap B \ne \varnothing\big)$$

is the Choquet capacity functional of the random set $G_y(\epsilon|\theta)$.

$\Rightarrow$ The identified set is characterized by

$$\Theta_I = \{\theta : P(B) \le c_\theta(B), \text{ all } B \text{ Borel}\}.$$

OT and Max flow characterization of Θ_I

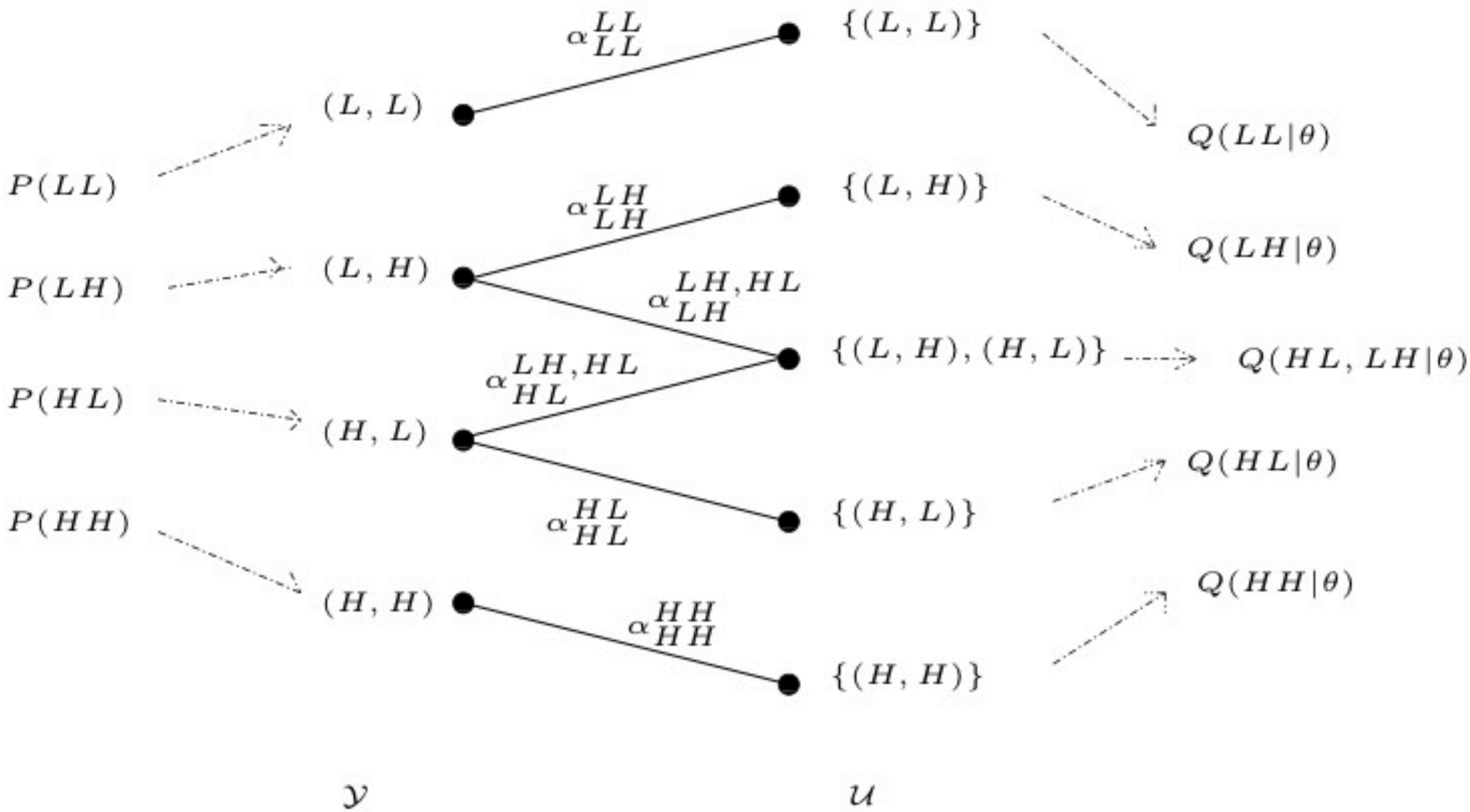
Call $\mathcal{U}$ the set of possible values of $G_y(\epsilon|\theta)$,

Define $Q(u|\theta) = \mathbb{P}(G_y(\epsilon|\theta) = u|\theta)$.

In the 2×2 game,

$$\mathcal{U} = \{\{(L,L)\},\{(L,H)\},\{(H,L)\},\{(H,H)\},\{(L,H),(H,L)\}\},$$

The relation between model and observables can be represented graphically in a bipartite graph.


$P(LL)$
$P(LH)$
$P(HL)$
$P(HH)$
(L, L)
(L, H)
(H, L)
(H, H)
α_{LL}^{LL}
α_{LH}^{LH}
$\alpha_{LH}^{LH,HL}$
$\alpha_{HL}^{LH,HL}$
α_{HL}^{HL}
α_{HH}^{HH}
$\{(L, L)\}$
$\{(L, H)\}$
$\{(L, H), (H, L)\}$
$\{(H, L)\}$
$\{(H, H)\}$
$Q(LL|\theta)$
$Q(LH|\theta)$
$Q(HL, LH|\theta)$
$Q(HL|\theta)$
$Q(HH|\theta)$
$\mathcal{Y}$
$\mathcal{U}$

OT and Max flow characterization of Θ_I (continued)

The following statements are equivalent:

- Parameter value θ rationalizes the data (i.e., $\theta \in \Theta_I$),
- For all subset A,

$$P(A) \leq \sum_{u \cap A \neq \varnothing} Q(u|\theta).$$

- There exists a joint probability on $\mathcal{G} = \{(y,u) : y \in u\}$ and with marginal probabilities $P(.)$ and $Q(.|\theta)$,
- A mass of 1 can flow through the directed network above.

OT again: inference on θ

- Back to our empirical model. For each θ, call $\mathcal{P}_\theta^{(n)}$ the set of distributions $P^{(n)}$ of samples $(Y_1, \ldots, Y_n)$ such that there exist random vectors $\epsilon_1, \ldots, \epsilon_n$ which satisfy:

  1. Support restriction: $(Y_i, \epsilon_i) \in G(\theta)$ a.s. for all $i \leq n$.
     - Equivalently: $Y_i \in G_y(\epsilon|\theta)$ or $\epsilon_i \in G_\epsilon(y|\theta)$.
  2. The $\epsilon_1, \ldots, \epsilon_n$ are independently distributed, $\epsilon_i \sim Q_{\epsilon|\theta}$.

- We want to learn about θ: more precisely, we want a confidence region

$$CR_n \quad := \quad \{\theta : T_n(\theta) \leq c_{n,1-\alpha}(\theta)\},$$

where $T_n(\theta)$ is a statistic and $c_{n,1-\alpha}(\theta)$ a critical value, such that

$$\inf_{P^{(n)} \in \mathcal{P}_\theta^{(n)}} P^{(n)}\left(T_n(\theta) \leq c_{n,1-\alpha}(\theta)\right) \quad = \quad 1 - \alpha,$$

Statistic $T_n(\theta)$

The test statistic is

$$T_n(\theta) = \min_{\pi \in \Pi_n} \sum_{i,j=1}^{n} \pi_{ij} C_{ij}(\theta) \left(\text{also} = \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{n} C_{i\sigma(i)}(\theta) \right),$$

where:

1. Π_n is the set of $n \times n$ non negative matrices π such that $\Sigma_i \pi_{ij} = \Sigma_j \pi_{ij} = 1/n$, for all $i, j \leq n$ (i.e., $n\pi$ is bi-stochastic).

2. The cost matrix $C(\theta)$ has entries

$$C_{ij}(\theta) \;\; := \;\; d\Big(\tilde{\epsilon}_i, G_\epsilon(Y_j|\theta) \Big), \text{ for each } i, j \leq n,$$

3. $\tilde{\epsilon}^{(n)} := (\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_n)$ is an iid simulated sample from $Q_{\epsilon|\theta}$.

Sample analogue of the OT problem from before

$$\min_{\pi \in \mathcal{M}(Q_{\epsilon|\theta}, P)} \mathbb{E}\Big[d\Big(\epsilon_i, G_\epsilon(Y_j|\theta) \Big) \Big]$$

Critical value $c_{n,1-\alpha}(\theta)$

The critical values are the quantiles of the distribution

$$\tilde{T}_n(\theta) \;=\; \sup_{C \in \mathcal{C}_\theta} \min_{\pi \in \Pi_n} \sum_{i,j=1}^{n} \pi_{ij} C_{ij},$$

where

- $\mathcal{C}_\theta$ is the class of all cost matrices with elements C_{ij} satisfying

$$C_{ij} \;=\; d((\tilde{\epsilon}_i, G_\epsilon(y|\theta)), \text{ for some } y \in G_y(\tilde{\epsilon}'_j|\theta).$$

- $(\tilde{\epsilon}'_1, \ldots, \tilde{\epsilon}'_n)$ is another iid simulated sample from $Q_{\epsilon|\theta}$.

In practice, the following is much faster to compute (finite sequence of OT and small dimensional LP problems):

$$\tilde{T}'_n(\theta) \;=\; \sup_{C \in co(\mathcal{C}_\theta)} \min_{\pi \in \Pi_n} \sum_{i,j=1}^{n} \pi_{ij} C_{ij},$$

OT and econometrics, future directions

1. Analysis of a class of incomplete models, where the support restriction is individual specific.

2. Beyond the parametric case: multi marginal OT and independence restrictions.

3. OT formulation of the problem of inference on a low dimensional function of the parameter, which is relevant to policy.

That's it for now.