'MATH+ECON+CODE' MASTERCLASS COMPETITIVE EQUILIBRIUM: WALRASIAN EQUILIBRIUM WITH SUBSTITUTES

Alfred Galichon (NYU)

Spring 2018
Day 3, May 23 2018: matching with general transfers (1)
Block 8. Nonadditive random utility models

- Nonadditive random utility models

- Hotz and Miller (1993)
- Berry (1994)
- Berry, Levinsohn, and Pakes (1995)
- Aguirregabiria and Mira (2002)
- Arcidiacono and Miller (2011)
- Dube, Fox and Su (2012).
- Galichon and Salanié (2014)
- Chiong, Galichon and Shum (2016)
- Bonnet, Galichon, O'Hara and Shum (2018).

Section 1

SETTING

- Consider a (nonadditive) discrete choice problem where an agent ("consumer") draws a utility shock $\varepsilon \sim P$, and faces a choice between alternatives ("yoghurts") $j \in \mathcal{J}_0 = \mathcal{J} \cup \{0\}$, where $j = 0$ is the default option.
- Alternative $j \in \mathcal{J}_0$ brings the agent utility $\mathcal{U}_{\varepsilon j}(\delta_j)$, where δ_j is a systematic utility-shifter, and $\mathcal{U}_{\varepsilon j}(.)$ is continuous and increasing.
- The utility associated with alternative 0 is normalized to 0: $\delta_0 = 0$.
- The agent's problem is

$$\max_{j \in \mathcal{J}_0} \mathcal{U}_{\varepsilon j}(\delta_j), \tag{1}$$

and the agent chooses random variable $\tilde{j}(\varepsilon) \in \arg\max_{j \in \mathcal{J}_0} \mathcal{U}_{\varepsilon j}(\delta_j)$.

Examples of additive models (ARUMS)

- In additive random utility models (ARUMs), $\mathcal{U}_{\varepsilon j}(\delta_j)$ is quasilinear in δ_j.
- If $\varepsilon \in \mathbb{R}^{\mathcal{J}}$, and δ_j=systematic utility, ARUMs have

$$\mathcal{U}_{\varepsilon j}(\delta_j) = \delta_j + \varepsilon_j$$

- Specifications include:
  - logit model: (ε_j) are i.i.d. Gumbel
  - pure characteristics: $\varepsilon_j = \epsilon^{\intercal}\zeta_j$ with ϵ a consumer-specific random taste vector of $\mathbb{R}^d$ and $\zeta_j \in \mathbb{R}^d$ is the vector of characteristics of alternative j.

Examples of nonadditive models (NARUMs)

- In nonadditive random utility models (NARUMs), $\mathcal{U}_{\varepsilon j}\left(\delta_j\right)$ is increasing and continuous, but not quasilinear in δ_j.
- Investments with taxes: $\varepsilon \in \{0,1\} \times \mathbb{R}^{\mathcal{J}}$; $\varepsilon^1 = 1$ if tax-exempt individual, $\varepsilon^1 = 0$ if tax-liable; $\delta_j + \varepsilon_j^2 =$ project j's pre-tax earnings, tax rate $\tau \in (0,1)$, so

$$\mathcal{U}_{\varepsilon j}\left(\delta_j\right) = \varepsilon^1 \left(\delta_j + \varepsilon_j^2\right) + \left(1 - \varepsilon^1\right)(1-\tau)\left(\delta_j + \varepsilon_j^2\right).$$

- Product choice with waiting lines: $\delta_j = T - T_j$ is the time spared in line waiting for product j; price is p_j; $\varepsilon \in \mathbb{R} \times \mathbb{R}^{\mathcal{J}}$, where ε^1 is the valuation of time and ε_j^2 some value shock, so

$$\mathcal{U}_{\varepsilon j}\left(\delta_j\right) = \varepsilon^1 \delta_j - p_j + \varepsilon_j^2.$$

- Other examples include: risk aversion; cap on compensation; etc.

- Assume for now no indifference:

$$\Pr\left(\mathcal{U}_{\varepsilon j}\left(\delta_j\right) = \mathcal{U}_{\varepsilon j'}\left(\delta_{j'}\right)\right) = 0 \text{ for every } \delta \text{ and } j \neq j'$$

then the *demand map* $\tilde{\sigma} : \mathbb{R}^{\mathcal{J}} \to \mathbb{R}^{\mathcal{J}}$ associates the market share (=vector of choice probabilities) to vector of systematic utilities δ, i.e.

$$s = \tilde{\sigma}\left(\delta\right) \iff s_j = \Pr\left(\mathcal{U}_{\varepsilon j}\left(\delta_j\right) \geq \max_{j' \in \mathcal{J}_0} \mathcal{U}_{\varepsilon j'}\left(\delta_{j'}\right)\right)$$

- Our focus is *inverse demand*, namely

$$\tilde{\sigma}^{-1}\left(s\right) = \left\{\delta \in \mathbb{R}^{\mathcal{J}} : s = \tilde{\sigma}\left(\delta\right)\right\}$$

a.k.a. "CCP inversion".
- In the logit case (additive random utility with Gumbel distribution), done in closed form, but in general, need to resort to various numerical methods.

- Demand inversion: literature started by Hotz and Miller (1993), Berry (1994). See among others Aguirregabiria and Mira (2002), Arcidiacono and Miller (2011), Kristensen, Nesheim, and de Paula (2014)...
- Fixed-point approach: Berry (1994) and Berry, Levinsohn, and Pakes (1995).
- MCMC approach: Rossi, Allenby and McCulloch (2005).
- MPEC approach: Dube, Fox and Su (2012).
- Optimal transport/linear programming approach (additive models): G and Salanié (2014), Chiong, G and Shum (2016), Chernozhukov, G, Henry and Pass (2014).

Section 2

The demand inversion problem

- Recall:
  - Demand computation (σ): **fix** $\varepsilon \sim P$ and $\delta \in \mathbb{R}^{\mathcal{J}}$, **compute** $\tilde{j} \sim s$.
  - Demand inversion (σ^{-1}): **fix** $\varepsilon \sim P$ and $\tilde{j} \sim s$, **compute** $\delta \in \mathbb{R}^{\mathcal{J}}$.
- Our result shows equivalence between the latter and:
  - Matching equilibrium: **fix** $\varepsilon \sim P$ and $\tilde{j} \sim s$, **compute** $\delta \in \mathbb{R}^{\mathcal{J}}$.
  - Here, $\varepsilon \sim P$=distribution of workers' types; $j \sim s$=distribution of firms' types; δ=wage offered by firm j.
- 2 sets of implications:
  - Computational: a new class of algorithms (coming from matching theory) for performing demand inversion.
  - Theoretical: (1) lattice structure; (2) inverse isotonicity; (3) desirable properties of the identified set (connectedness, point-identication, stability)

- Consider a market where workers characteristics is $\varepsilon \in \mathcal{X}$ (continuous) and firms' types are $j \in \mathcal{J}_0$ (discrete). Assume $\varepsilon \sim P$, and $j \sim s$, i.e. workers are distributed as P and the mass of firm's type j is s_j. One has

$$\int dP(\varepsilon) = 1 \text{ and } \sum_{j \in \mathcal{J}_0} s_j = 1,$$

so that there is the same mass (normalized to one) of workers and firms.

- Assume that if worker ε works with firm j at wage u, then the firm's profit is $\Phi_{\varepsilon j}(u)$, decreasing and continuous in u.
- An outcome is the specification of $(\pi(\varepsilon, j), u(\varepsilon), v_j)$, where
  - $\pi(\varepsilon, j)$ is a matching: it is the probability of drawing a matched pair of worker ε and firm j. It should have margins P and s, i.e.

$$\pi \in M(P, s) \Leftrightarrow_{def} \left\{ \begin{array}{l} \int \pi(\varepsilon, j)\, d\varepsilon = s_j \\ \sum_{j \in \mathcal{J}_0} \pi(\varepsilon, j) = P(\varepsilon) \end{array} \right. .$$

  - $(u(\varepsilon), v_j)$ are payoffs: $u(\varepsilon)$ is the wage of worker ε, and v_j is the profit of a firm of type j.

- Cf. Roth-Sotomayor, ch. 9. Outcome (π, u, v) if *pairwise stable* when:
  - *no blocking pair*: for all ε and j,

$$v_j \geq \Phi_{\varepsilon j}(u(\varepsilon))$$

    otherwise the pair would be "blocking" i.e. it would be possible for firm j to hire ε at her current wage and get more than v_j.
  - *feasibility*: if ε and j are matched, i.e. if ,

$$(\varepsilon, j) \in Supp(\pi) \implies v_j = \Phi_{\varepsilon j}(u(\varepsilon))$$

    otherwise the payoffs $u(\varepsilon)$ and δ_j would not be sustainable for matched pair (ε, j).
- Walrasian interpretation of pairwise stability: if $(u(\varepsilon), v_j)$ are stable payoffs, then no blocking pair+feasibility imply that

$$v_j = \max_{\varepsilon \in \mathcal{X}} \Phi_{\varepsilon j}(u(\varepsilon)) \text{ and } u(\varepsilon) = \max_{j \in \mathcal{J}_0} \Phi_{\varepsilon j}^{-1}(v_j). \tag{2}$$

  - This is expresses that the problem of workers and the problem of firms are dual of one another.
  - Equivalence result stems from associating $\Phi_{\varepsilon j}^{-1}(v_j)$ with $\mathcal{U}_{\varepsilon j}(\delta_j)$.

The following result is proven in Bonnet, G, O'Hara and Shum (2018):

THEOREM

One has

$$\delta \in \tilde{\sigma}^{-1}(s)$$

if and only if (u, v) *is a stable outcome in the matching problem, where*

$$\delta_j = -v_j \text{ and } \mathcal{U}_{\varepsilon j}(\delta_j) = \Phi_{\varepsilon j}^{-1}(-\delta_j)$$

and

$$u(\varepsilon) = \max_{j \in \mathcal{J}_0} \{\mathcal{U}_{\varepsilon j}(\delta_j)\}.$$

- Note that Φ is obtained from $\mathcal{U}$ by

$$\Phi_{\varepsilon j}(u) = -\mathcal{U}_{\varepsilon j}^{-1}(u).$$

- In fact, there is an equivalent – dual – problem where "yoghurts choose consumers" – just as in the matching model, the problem of the workers and the problem of the firms are dual one another.
- Yoghurt j's problem is

$$v_j = -\delta_j = \max_{\varepsilon \in \mathcal{X}} \left\{ -\mathcal{U}_{\varepsilon j}^{-1} \left(u \left(\varepsilon \right) \right) \right\},$$

  and these problems are observationally equivalent: an econometrician could not find out from the data whether consumers chose yoghurts or yoghurts chose consumers.
- This duality is an instance of a Galois connection (Singer 1997). It generalizes convex duality (Rockafellar 1970), as well as Φ-convex duality (Villani 2009).
- Bottom line: the distinction between "one-sided" vs. "two-sided" problem is purely arbitrary when considering an equilibrium problem. This has other implications, see e.g. the Becker-Coase theorem.

- Consider the pure characteristics model:

$$\mathcal{U}_{\varepsilon j}\left(\delta_j\right) = \delta_j + \epsilon^{\intercal}\xi_j$$

(w.l.o.g. the argument extends to any ARUM). Then $\mathcal{U}_{\varepsilon j}^{-1}\left(u\right) = u - \epsilon^{\intercal}\xi_j$, so

$$\Phi_{\epsilon\xi_j}\left(u\right) = -u + \epsilon^{\intercal}\xi_j.$$

- $\left(\pi, u, v\right)$ is a stable outcome in the matching problem if for all ε and j,

$$u\left(\epsilon\right) + \delta_j \geq \epsilon^{\intercal}\xi_j$$

with equality if $\left(\epsilon, j\right) \in Supp\left(\pi\right)$.

- The problem of demand inversion in ARUMS can thus be recast as a matching problem with transferable utility, or optimal transport problem (Monge-Kantorovich, Koopmans-Beckman, Dantzig, Shapley-Shubik, Beker):

$$\min_{u,\delta} \int u(\epsilon)\, dP(\epsilon) + \sum_{j \in \mathcal{J}_0} \delta_j s_j$$
$$s.t.\ u(\epsilon) + \delta_j \geq \epsilon^\intercal \xi_j$$

which has primal formulation

$$\max_{\pi \in M(P,s)} \mathbb{E}_\pi \left[\epsilon^\intercal \xi_j \right]$$

- Very active area of research in mathematics and applications; see my recent *Optimal Transport Methods in Economics* monograph for some economic applications. Many recent algorithms for solving this problem; see Peyré and Cuturi's forthcoming book *Numerical Optimal Transport* (arxiv 1803.00567).

- In the case of NARUMs, the problem of demand inversion is equivalent to a problem of matching with imperfectly transferable utility, or equilibirum transport problem–a far reaching generalization of optimal transport problems.
- Unlike optimal transport problem, equilibrium transport problem do not have a natural optimization formulation.
- Emerging literature; see Trudinger (2013), Noeldeke and Samuelson (2015), McCann and Zhang (2017).

Implications of the Equivalence theorem

- Computational implications: use of matching algorithms to perform demand inversioin, i.e.
  - Kelso and Crawford (Consumer propose first vs. Yoghurts propose first).
  - Entropic regularization
  - Transportation algorithms: dual transportation simplex, Bertsekas' auction algorithm...
- Theoretical implications: structure of the identified payoff set, i.e.
  - Representation of the demand correspondence
  - Inverse isotonicity
  - Lattice structure of the identified payoff set
  - Nonemptyness of the identified payoff set
  - Consistency

Section 3

COMPUTATION

- Thanks to the Equivalence theorem, demand inversion in discrete choice models can be performed by matching algorithm (and conversely, any algorithm doing demand inversion can be used as an algorithm to determine stable outcomes in matching problems).
- In the specific case of additive random utility models (equivalent to a matching problem with transferable utility), there is a rich panel of algorithm. These include:
    - generic linear programming using a large-scale solver such as Gurobi a black box (see Chiong, G and Shum, 2016)
    - dual transportation simplex
    - the Hungarian algorithm for the optimal assignment
    - Bertsekas's auction algorithm with similar objects
    - Pseudo-flow algorithms (Orlin-Ahuja, Goldberg-Kennedy)
- This is especially useful for the pure characteristics model.

- In the case of ARUMs (say, the pure characteristics model), recall the primal formulation of the equivalent matching problem $\max_{\pi \in M(P,s)} \mathbb{E}_\pi \left[\epsilon^\intercal \xi_j\right]$.
- Now consider N i.i.d. draws $\epsilon_i, 1 \leq i \leq N$ sampled from the distribution P on $\mathbb{R}^d$. The problem becomes

$$\max_{\pi_{ij} \geq 0} \sum_{\substack{1 \leq i \leq N \\ j \in J_0}} \pi_{ij} \epsilon_i^\intercal \xi_j$$
$$s.t. \sum_{j \in J_0} \pi_{ij} = 1/N \;\; [u_i]$$
$$\sum_{1 \leq i \leq N} \pi_{ij} = s_j \;\; [-\delta_j]$$

which has dual

$$\min_{(u_i),(\delta_j)} \sum_{i=1}^{N} \frac{u_i}{N} - \sum_{j \in J_0} s_j \delta_j$$
$$s.t. \; u_i - \delta_j \geq \epsilon_i^\intercal \xi_j.$$

- Consider the previous problem with an entropic regularization term

$$
\begin{aligned}
\max_{\pi_{ij} \geq 0} & \sum_{\substack{1 \leq i \leq N \\ j \in J_0}} \pi_{ij} \epsilon_i^\intercal \xi_j - T \sum \pi_{ij} \ln \pi_{ij} \\
s.t. & \sum_{j \in J_0} \pi_{ij} = 1/N \;\; [u(\epsilon_i)] \\
& \sum_{1 \leq i \leq N} \pi_{ij} = s_j \;\; [-\delta_j]
\end{aligned}
$$

  which has dual

$$
\min_{(u_i),(\delta_j)} \sum_{i=1}^{N} \frac{u_i}{N} - \sum_{j \in J_0} s_j \delta_j + T \sum_{i,j} \exp\left(\frac{\delta_j + \epsilon_i^\intercal \xi_j - u_i}{T} \right) - T
$$

- Formulation very much used in numerical OT for fast and parallelizable computation of OT problems; see Peyré and Cuturi, ch. 4. In particular, can be computed by an algorithm called IPFP/RAS/Sinkhorn/Matrix scaling/Bregman iterated projection algorithm – call it IPFP in the sequel, which is coordinate descent on the dual.

- Descent on u_i (exact minimization wrt u_i for fixed δ_j's) yields

$$\sum_{j \in J_0} \exp\left(\frac{\delta_j + \epsilon_i^\intercal \xi_j - u_i}{T}\right) = \frac{1}{N}$$

that is

$$u_i^{2k+1} = T \log N \sum_{j \in J_0} \exp\left(\frac{\delta_j^{2k} + \epsilon_i^\intercal \xi_j}{T}\right)$$

- Descent on the δ_j (exact minimization wrt δ_j for fixed u_i's) yields

$$\sum_i \exp\left(\frac{\delta_j + \epsilon_i^\intercal \xi_j - u_i}{T}\right) = s_j$$

that is

$$\delta_j^{2k+2} = -T \log \frac{\sum_{i=1}^N \exp\left(\frac{\epsilon_i^\intercal \xi_j - u_i^{2k}}{T}\right)}{s_j}.$$

- Note that if we combine the two steps, we get,

$$\delta_j^{2k+2} = T \log s_j - T \log \sum_{i=1}^{N} \frac{1}{N} \frac{\exp\left(\frac{\epsilon_i^\intercal \xi_j}{T}\right)}{\sum_{j \in J_0} \exp\left(\frac{\delta_j^{2k} + \epsilon_i^\intercal \xi_j}{T}\right)}$$

- Hence the IPFP coincides exactly the contraction mapping algorithm of Berry, Levinsohn and Pakes. However as we'll see the coincidence will cease in the NARUM case.

- As is well-known, the contraction mapping algorithm behaves poorly when T gets small. This is due to the fact there may be large positive numbers in the exponentials, causing blowups.
- That, however, can be fixed using the log-sum-exp trick, familiar in maching learning:

$$T \log \sum_i \exp \frac{a_i}{T} = \max_i \{a_i\} + T \log \sum_i \exp \frac{a_i - \max_i \{a_i\}}{T}$$

  where substracting the maximum ensures there are only nonpositive numbers in the arguments of the exponentials.
- Here, the algorithm yields

$$\begin{cases} \bar{\delta}_j^{k+1} = \min_i \left\{ \frac{u_i^k - \nu_i^\intercal x_j + T \log s_j}{T} \right\} \\ \delta_j^{k+1} = \bar{\delta}_j^{k+1} - T \log \left(\sum_{i=1}^N \frac{1}{N} \exp \left(\frac{-u_i^k + \nu_i^\intercal x_j - T \log s_j + \bar{\delta}_j^{k+1}}{T} \right) \right) \\ \bar{u}_i^{k+1} = \max_{j \in \mathcal{J}_0} \frac{\delta_j^{k+1} + \nu_i^\intercal x_j}{T} \\ u_i^{k+1} = \bar{u}_i^{k+1} + T \log \left(\sum_{j \in \mathcal{J}_0} \exp \left(\frac{\delta_j^{k+1} + \nu_i^\intercal x_j - \bar{u}_i^{k+1}}{T} \right) \right). \end{cases}$$

- In the pure characteristics model, $\sigma = 0$, and the problem becomes

$$\max_{\pi \in M(P,s)} \mathbb{E}_\pi \left[\epsilon^\intercal \xi_j \right]$$

which has dual formulation

$$\min_{u,\delta} \int u(\epsilon)\, dP(\epsilon) + \sum_{j \in \mathcal{J}_0} \delta_j s_j$$
$$s.t.\ u(\epsilon) + \delta_j \geq \epsilon^\intercal \xi_j$$

- This problem can be reformulated as a finite-dimensional convex optimization problem as

$$\min_\delta F(\delta) := \int \max_{j \in \mathcal{J}_0} \left\{ \epsilon^\intercal \xi_j - \delta_j \right\} dP(\epsilon) + \sum_{j \in \mathcal{J}_0} \delta_j s_j$$

where F is convex and strictly convex (after imposing $\delta_0 = 0$), and

$$\nabla F(\delta)_j = P\left(j \in \arg\max_{j \in \mathcal{J}_0} \left\{ \epsilon^\intercal \xi_j - \delta_j \right\} \right) - s_j$$

so provided we can compute the latter, we should be able to do gradient descent on δ; Newton descent if we can compute the Hessian.

- The problem hence boils down to evaluating the mass assigned by ϵ to the region
$$\epsilon^{\mathsf{T}}\left(\xi_j - \xi_{j'}\right) \geq \delta_j - \delta_{j'} \;\forall j \in \mathcal{J}_0$$
which is a polyhedron. When P is piecewise constant on polyhedral regions, this is something that can be achieved by computational geometry operations. The basic difficulty is to determining the extreme points of these regions.

- Recentely, this problem has become extremely popular in image analysis, where optimal transport methods have been used extensively over the last decade. This problem (where ϵ is continuous and j is discrete) is called the semi-discrete transport problem.
- Two teams of researchers, one led by Quentin Mérigot (Université Paris-Sud) and the other one by Bruno Lévy (INRIA), have managed to write libraries that solve the problem for millions of j's and up to three characteristic dimensions (for now).

- In the case of nonadditive random utility models, linear programming cannot be used. One has therefore to resort to more general matching algorithms.
- Matching algorithms are well developped in the case when $\mathcal{X}$ and $\mathcal{J}_0$ are both discrete, and when the set of possible transfers (δ) is also discrete. In this case, on can apply the algorithm by Kelso and Crawford (1981), which has been shown to be an isotone fixed point algorithm by Halfield and Milgrom (2005).
- When the set of possible transfers δ is continuous, G, Kominers and Weber (2016) have proposed an iterative fitting procedure, which compute an approximate solution to the problem.

- Outside of ARUMs, BLP's contraction mapping is no longer necessarily a contraction mapping, and may not converge in some number of instances.
- Consider the NARUM problem

$$u_i = \max_{j \in \mathcal{J}_0} \mathcal{U}_{\varepsilon_i j}\left(\delta_j\right)$$

by taking η_{ij} iid logit independent of ε and $T > 0$ and replace the problem as in BLP, by

$$u_i = \max_{j \in \mathcal{J}_0} \mathcal{U}_{\varepsilon_i j}\left(\delta_j\right) + T\eta_{ij}$$

- Therefore the expected indirect utility conditional on ε becomes the smooth-max (aka McFadden's smooth accept-reject simulator)

$$u_i = T \log \sum_{j \in \mathcal{J}_0} \exp\left(\mathcal{U}_{\varepsilon_i j}\left(\delta_j\right) / T\right)$$

- One seeks δ so that the induced choice probabilities are s, that is

$$s_j = \sum_{i=1}^{N} \frac{1}{N} \frac{\exp\left(\mathcal{U}_{\varepsilon_i j}\left(\delta_j\right) / T\right)}{\exp\left(u_i / T\right)}.$$

- Letting

$$\pi_{ij} = \frac{1}{N} \exp\left(\frac{\mathcal{U}_{\varepsilon_i j}\left(\delta_j\right) - u_i}{T}\right),$$

one has

$$\begin{cases} \sum_{j \in \mathcal{J}_0} \pi_{ij} = \frac{1}{N} \\ \sum_{1 \leq i \leq N} \pi_{ij} = s_j \end{cases}$$

- The generalized IPFP algorithm proposed in G, Kominers and Weber (2016) allows for fast and scalable computation of this problem by iteratively fitting these equations. Namely, given any starting point δ^0, solve

$$\begin{cases} u_i^{k+1} = T \log \sum_{j \in \mathcal{J}_0} \exp\left(\mathcal{U}_{\varepsilon_i j}\left(\delta_j^0\right) / T\right) \\ \delta_j^{k+1} \text{ such that } \frac{1}{N} \sum_{1 \leq i \leq N} \exp\left(\frac{\mathcal{U}_{\varepsilon_i j}\left(\delta_j^{k+1}\right) - u_i}{T}\right) = s_j \end{cases}$$

- Can be implemented in parallel. The algorithm can be proven to converge.

- Other tested algorithms include:
  - Kelso-Crawford: for exposition purposes; not competitive
  - Non-additive version of Bertsekas' auction algorithm: good empirical performances

Section 4

STRUCTURE OF THE IDENTIFIED SET

We have already proven the following theorem:

THEOREM

The set-valued map $\tilde{\sigma}^{-1}(s)$ defined for $s \in \mathbb{R}^{\mathcal{J}}$ is isotone in Veinott's strong set order. That is, if $s \leq s'$, $\delta \in \tilde{\sigma}^{-1}(s)$ and $\delta' \in \tilde{\sigma}^{-1}(s')$, then

$$\delta \wedge \delta' \in \tilde{\sigma}^{-1}(s) \text{ and } \delta \vee \delta' \in \tilde{\sigma}^{-1}(s') .$$

- Theorem connects a number of existing results:
  - Berry, Gandhi and Haile (2013): point-valued case (under "connected strong substitutes"), NARUM
  - Topkis' theorem: set-valued-case, ARUM
  - Demange and Gale (1985): lattice structure

- As a result of the isotonicity theorem, we have (applying the theorem with $s = s'$) that if $\tilde{\sigma}^{-1}(s)$ is nonempty, then it is a sublattice of $\mathbb{R}^{\mathcal{J}}$.
- In that case, there are minimal and maximal elements $\tilde{\delta}^{\min}(s)$ and $\tilde{\delta}^{\max}(s)$ of $\tilde{\sigma}^{-1}(s)$

$$\tilde{\delta}^{\min}(s) = \min \tilde{\sigma}^{-1}(s) \text{ and } \tilde{\delta}^{\max}(s) = \max \tilde{\sigma}^{-1}(s)$$

  and $\tilde{\delta}^{\min}(s)$ and $\tilde{\delta}^{\max}(s)$ are isotone w.r.t. s.
- One can also show that $\tilde{\sigma}^{-1}(s)$ is connected. Lattice structure and connectedness are well-known properties in matching theory.

- As before, consider the regularization of the problem

$$u^T(\varepsilon) = \mathbb{E} \max_{j \in \mathcal{J}_0} \left\{ \mathcal{U}_{\varepsilon j}(\delta_j) + T\eta_j \right\},$$

  where $T > 0$ and (η_j) is a vector of i.i.d. logit random variables.
- Then $\pi^T(j|\varepsilon) = \exp\left(\mathcal{U}_{\varepsilon j}(\delta_j) / T\right) / \left(\sum_{k \in \mathcal{J}_0} \exp\left(\mathcal{U}_{\varepsilon k}(\delta_k) / T\right)\right)$, and one can show that under mild assumptions:
  - for each $T > 0$, there is a unique vector $\left(\delta_j^T\right)$ such that $\pi^T \in \mathcal{M}(P, s)$
  - as $T \to 0$, $\left(u^T, \delta^T\right)$ tends to a stable outcome in the matching problem, which shows nonemptyness of $\tilde{\Sigma}^{-1}(s)$.
- This regularization is also useful for computational purposes.

Consistency

- In practice, s^n is observed instead of s (sampling) and P^n instead of P (simulation). Assume $s^n \to s$ and $P^n \Rightarrow P$ as $n \to +\infty$.
- By the lattice property of $\tilde{\Sigma}^{-1}(s)$, we have that for any $j \in \mathcal{J}$, $\overline{\delta}_j = \tilde{\delta}^{\max}(s)$ is given by

$$\max_{\substack{\delta_{-j} \in \mathbb{R}^{\mathcal{J} \backslash \{j\}} \\ \delta_j \in \mathbb{R}}} \left\{ \delta_j : \forall B \subseteq \mathcal{J}_0 : \sum_{j \in B} s_j \leq P \left(\max_{j \in B} \mathcal{U}_{\varepsilon j}(\delta_j) \geq \max_{j \in \mathcal{J}_0 \backslash B} \mathcal{U}_{\varepsilon j}(\delta_j) \right) \right\}$$

- Hence,

$$\overline{\delta}_j = \max_{\delta_{-j} \in \mathbb{R}^{\mathcal{J} \backslash \{j\}}} \min_{\substack{B \subseteq \mathcal{J}_0 \\ j \notin B}} F_{jB}(\delta_{-j}), \text{ where}$$

$$F_{jB}(\delta_{-j}) = \max \left\{ \delta_j : \sum_{j \in B} s_j \leq P \left(\max_{j \in B} \mathcal{U}_{\varepsilon j}(\delta_j) \geq \max_{j \in \mathcal{J}_0 \backslash B} \mathcal{U}_{\varepsilon j}(\delta_j) \right) \right\}.$$

- It is possible to show that if $s^n \to s$ and $P^n \Rightarrow P$, then $F^n_{jB}(\delta_{-j}) \to F_{jB}(\delta_{-j})$ uniformly. Thus $\overline{\delta}^n_j \to \overline{\delta}_j$: desired consistency holds.

- A necessary condition for $\delta \in \tilde{\sigma}^{-1}(s)$ is

$$\Pr\left(\mathcal{U}_{\varepsilon j}^{-1}\mathcal{U}_{\varepsilon j}(\delta_0) > \delta_j : \forall j \in \mathcal{J}\right) \leq 1 - s_0 \leq \Pr\left(\mathcal{U}_{\varepsilon j}^{-1}\mathcal{U}_{\varepsilon j}(\delta_0) \geq \delta_j : \forall j \in \mathcal{J}\right)$$

- Hence, if the random vector $\left(\mathcal{U}_{\varepsilon j}^{-1}\mathcal{U}_{\varepsilon j}(\delta_0), j \in \mathcal{J}\right)$ has a nonvanishing density, this condition is equivalent to $F_Z(-\delta) = 1 - s_0$, where $Z_j = -\mathcal{U}_{\varepsilon j}^{-1}\mathcal{U}_{\varepsilon j}(\delta_0)$. Thus there cannot be $\underline{\delta} \leq \overline{\delta}$, $\underline{\delta} \neq \overline{\delta}$, and $F(\underline{\delta}) = F(\overline{\delta})$.
- Thus if the random vector $\left(\mathcal{U}_{\varepsilon j}^{-1}\mathcal{U}_{\varepsilon j}(\delta_0), j \in \mathcal{J}\right)$ has a nonvanishing density, then $\tilde{\sigma}^{-1-1}(s)$ contains at most one point.