A model of decentralized (aggregate) matching without transfers

Alfred Galichon (NYU Econ+Courant)   Yu-Wei Hsieh (USC Econ)

Conference on market design,
Columbia University, April 13, 2018

- This talk is about aggregate equilibria in **decentralized matching market without prices**, aka decentralized non-transferable utility (NTU) markets.
- Decentralized NTU markets arises in a number of situations (taxis, health care, rent-controlled housing...)
- Need for analytical tools for regulatory purposes (fix prices to optimize efficiency/fairness tradeoff).
- Also, operators doing dynamic pricing need consider the market as NTU over short time scales. Eg. Uber sets fixed prices for some time span–during that time span, the market is NTU.

- This talk is about aggregate equilibria in **decentralized matching market without prices**, aka decentralized non-transferable utility (NTU) markets.
- Decentralized NTU markets arises in a number of situations (taxis, health care, rent-controlled housing...)
- Need for analytical tools for regulatory purposes (fix prices to optimize efficiency/fairness tradeoff).
- Also, operators doing dynamic pricing need consider the market as NTU over short time scales. Eg. Uber sets fixed prices for some time span–during that time span, the market is NTU.

- This talk is about aggregate equilibria in **decentralized matching market without prices**, aka decentralized non-transferable utility (NTU) markets.
- Decentralized NTU markets arises in a number of situations (taxis, health care, rent-controlled housing...)
- Need for analytical tools for regulatory purposes (fix prices to optimize efficiency/fairness tradeoff).
- Also, operators doing dynamic pricing need consider the market as NTU over short time scales. Eg. Uber sets fixed prices for some time span–during that time span, the market is NTU.

- This talk is about aggregate equilibria in **decentralized matching market without prices**, aka decentralized non-transferable utility (NTU) markets.
- Decentralized NTU markets arises in a number of situations (taxis, health care, rent-controlled housing...)
- Need for analytical tools for regulatory purposes (fix prices to optimize efficiency/fairness tradeoff).
- Also, operators doing dynamic pricing need consider the market as NTU over short time scales. Eg. Uber sets fixed prices for some time span–during that time span, the market is NTU.

Section 1

Individual cardinal stable matchings

- Consider "men" $i \in \mathcal{I}$ and "women" $j \in \mathcal{J}$. One of each type. If i and j match, then i gets α_{ij} and j gets γ_{ij}. Unmatched agent's utility normalized to zero. Let μ_{ij} be such that

$$\mu_{ij} \in \{0,1\}, \ \sum_j \mu_{ij} \leq 1 \text{ and } \sum_i \mu_{ij} \leq 1$$

- Classical notion=**Ordinal NTU matching**. μ is an ordinally stable matching (OSM) if

$$\forall i,j : \max\left\{ U_i^{\mu} - \alpha_{ij}, V_j^{\mu} - \gamma_{ij} \right\} \geq 0, \ U_i^{\mu} \geq 0, \ V_j^{\mu} \geq 0$$

where $U_i^{\mu} := \sum_{j'} \mu_{ij'} \alpha_{ij'}$ and $V_j^{\mu} := \sum_{i'} \mu_{i'j} \gamma_{i'j}$.

- Proposed notion=**Cardinal NTU matching**. (μ, u, v) is a cardinally stable matching (CSM) if

$$\begin{cases} \forall i,j : \max\{u_i - \alpha_{ij}, v_j - \gamma_{ij}\} \geq 0, \ u_i \geq 0, \ v_j \geq 0 \\ \mu_{ij} > 0 \Longrightarrow \max\{u_i - \alpha_{ij}, v_j - \gamma_{ij}\} = 0 \\ \sum_j \mu_{ij} = 0 \Longrightarrow u_i = 0, \ \sum_i \mu_{ij} = 0 \Longrightarrow v_j = 0 \end{cases}$$

- Consider "men" $i \in \mathcal{I}$ and "women" $j \in \mathcal{J}$. One of each type. If i and j match, then i gets α_{ij} and j gets γ_{ij}. Unmatched agent's utility normalized to zero. Let μ_{ij} be such that

$$\mu_{ij} \in \{0, 1\}, \ \sum_j \mu_{ij} \leq 1 \text{ and } \sum_i \mu_{ij} \leq 1$$

- Classical notion=**Ordinal NTU matching**. μ is an ordinally stable matching (OSM) if

$$\forall i, j : \max \left\{ U_i^\mu - \alpha_{ij}, V_j^\mu - \gamma_{ij} \right\} \geq 0, \ U_i^\mu \geq 0, \ V_j^\mu \geq 0$$

where $U_i^\mu := \sum_{j'} \mu_{ij'} \alpha_{ij'}$ and $V_j^\mu := \sum_{i'} \mu_{i'j} \gamma_{i'j}$.

- Proposed notion=**Cardinal NTU matching**. (μ, u, v) is a cardinally stable matching (CSM) if

$$\begin{cases} \forall i, j : \max \{u_i - \alpha_{ij}, v_j - \gamma_{ij}\} \geq 0, \ u_i \geq 0, \ v_j \geq 0 \\ \mu_{ij} > 0 \Longrightarrow \max \{u_i - \alpha_{ij}, v_j - \gamma_{ij}\} = 0 \\ \sum_j \mu_{ij} = 0 \Longrightarrow u_i = 0, \ \sum_i \mu_{ij} = 0 \Longrightarrow v_j = 0 \end{cases}$$

- Consider "men" $i \in \mathcal{I}$ and "women" $j \in \mathcal{J}$. One of each type. If i and j match, then i gets α_{ij} and j gets γ_{ij}. Unmatched agent's utility normalized to zero. Let μ_{ij} be such that

$$\mu_{ij} \in \{0,1\}, \ \sum_j \mu_{ij} \leq 1 \text{ and } \sum_i \mu_{ij} \leq 1$$

- Classical notion=**Ordinal NTU matching**. μ is an ordinally stable matching (OSM) if

$$\forall i,j : \max\left\{U_i^\mu - \alpha_{ij}, V_j^\mu - \gamma_{ij}\right\} \geq 0, \ U_i^\mu \geq 0, \ V_j^\mu \geq 0$$

where $U_i^\mu := \sum_{j'} \mu_{ij'}\alpha_{ij'}$ and $V_j^\mu := \sum_{i'} \mu_{i'j}\gamma_{i'j}$.

- Proposed notion=**Cardinal NTU matching**. (μ, u, v) is a cardinally stable matching (CSM) if

$$\begin{cases} \forall i,j : \max\{u_i - \alpha_{ij}, v_j - \gamma_{ij}\} \geq 0, \ u_i \geq 0, \ v_j \geq 0 \\ \mu_{ij} > 0 \Longrightarrow \max\{u_i - \alpha_{ij}, v_j - \gamma_{ij}\} = 0 \\ \sum_j \mu_{ij} = 0 \Longrightarrow u_i = 0, \ \sum_i \mu_{ij} = 0 \Longrightarrow v_j = 0 \end{cases}$$

Proposition. In the setting above:
(1) If μ is a OSM, then (μ, U^{μ}, V^{μ}) is an CSM. Convertly, if (μ, u, v) is a CSM, then μ is a OSM.
(2) Let μ be a OSM. Then set of $(u, -v)$ such that (μ, u, v) is a CSM is a lattice.
(3) The set of $(u, -v)$ such that there exists a μ such that (μ, u, v) is a CSM is a lattice.

The direct implication of (1) is obvious. Let us show the converse of (1). Consider (μ, u, v) a CSM, and assume μ is not an OSM. Then there is a blocking pair, or a blocking individual. In the first case one has

$$\max\left\{U_i^\mu - \alpha_{ij}, V_j^\mu - \gamma_{ij}\right\} < 0$$

Assume $\sum_j \mu_{ij} = 1$. Then let j' be such that $\mu_{ij'} = 1$; we have $U_i^\mu = \alpha_{ij'}$ and $\max\left\{u_i - \alpha_{ij'}, v_{j'} - \gamma_{ij'}\right\} = 0$, hence $u_i \leq \alpha_{ij'} = U_i^\mu$. If on the contrary $\sum_j \mu_{ij} = 0$, then we have $U_i^\mu = 0 = u_i$. Similarly one can show that $v_j \leq V_j^\mu$. Therefore, we have

$$\max\left\{u_i - \alpha_{ij}, v_j - \gamma_{ij}\right\} \leq \max\left\{U_i^\mu - \alpha_{ij}, V_j^\mu - \gamma_{ij}\right\} < 0$$

so the existence of a blocking pair leads to a contradiction. If there is a blocking individual $U_i^\mu < 0$, but in that case a similar logic implies that $u_i \leq U_i^\mu < 0$, a contradiction as well.

- Why bother introducing CSMs if they are essentially equivalent to the classical OSMs?
- The reason is that CSMs allow for a natural notion of *aggregate* decentralized matching, which OSMs don't.
- If there are multiple indistinguishable agents, a natural requirement of decentralized equilibrium is to satisfy equal treatment – i.e. that identical individuals should get the same payoffs at equilibrium.

- Why bother introducing CSMs if they are essentially equivalent to the classical OSMs?
- The reason is that CSMs allow for a natural notion of *aggregate* decentralized matching, which OSMs don't.
- If there are multiple indistinguishable agents, a natural requirement of decentralized equilibrium is to satisfy equal treatment – i.e. that identical individuals should get the same payoffs at equilibrium.

- Why bother introducing CSMs if they are essentially equivalent to the classical OSMs?
- The reason is that CSMs allow for a natural notion of *aggregate* decentralized matching, which OSMs don't.
- If there are multiple indistinguishable agents, a natural requirement of decentralized equilibrium is to satisfy equal treatment – i.e. that identical individuals should get the same payoffs at equilibrium.

Section 2

Aggregate cardinal stable matchings

- Assume that there are n_x men's types, $x \in \mathcal{X}$ and m_y women's types, $y \in \mathcal{Y}$. If x and y match, then x gets α_{xy} and y gets γ_{xy}. Unmatched agent's utility normalized to zero. Let μ_{xy} be such that

$$\mu_{xy} \in \mathbb{N}, \sum_{y \in \mathcal{Y}} \mu_{xy} \leq n_x \text{ and } \sum_{x \in \mathcal{X}} \mu_{xy} \leq m_y$$

- (μ, u, v) is an aggregate CSM if

$$\begin{cases} \forall x, y : \max\{u_x - \alpha_{xy}, v_y - \gamma_{xy}\} \geq 0, \; u_x \geq 0, \; v_y \geq 0 \\ \mu_{xy} > 0 \Longrightarrow \max\{u_x - \alpha_{xy}, v_y - \gamma_{xy}\} = 0 \\ \sum_{y \in \mathcal{Y}} \mu_{xy} = 0 \Longrightarrow u_x = 0, \; \sum_{x \in \mathcal{X}} \mu_{xy} = 0 \Longrightarrow v_y = 0 \end{cases}$$

- Assume that there are n_x men's types, $x \in \mathcal{X}$ and m_y women's types, $y \in \mathcal{Y}$. If x and y match, then x gets α_{xy} and y gets γ_{xy}. Unmatched agent's utility normalized to zero. Let μ_{xy} be such that

$$\mu_{xy} \in \mathbb{N}, \ \sum_{y \in \mathcal{Y}} \mu_{xy} \leq n_x \text{ and } \sum_{x \in \mathcal{X}} \mu_{xy} \leq m_y$$

- (μ, u, v) is an aggregate CSM if

$$\begin{cases} \forall x, y : \max\{u_x - \alpha_{xy}, v_y - \gamma_{xy}\} \geq 0, \ u_x \geq 0, \ v_y \geq 0 \\ \mu_{xy} > 0 \Longrightarrow \max\{u_x - \alpha_{xy}, v_y - \gamma_{xy}\} = 0 \\ \sum_{y \in \mathcal{Y}} \mu_{xy} = 0 \Longrightarrow u_x = 0, \ \sum_{x \in \mathcal{X}} \mu_{xy} = 0 \Longrightarrow v_y = 0 \end{cases}$$

- Assume that there are 2 identical passengers and 1 driver. The value of being unmatched (for the passengers and the driver alike) is 0. The value of being matched is 1, both for the passengers and driver.
- In a model with prices (Uber model–transferable utility), the price of the ride will be 1, so that the driver's payoff is 2, and both passengers' payoffs is zero. Thus, passengers are indifferent between being matched and unmatched.
- In a classical model without transfers (taxi model–nontransferable utility), there are two stable matchings in each of which the matched passenger gets one, while the unmatched gets zero. Thus in this Gale-Shapley solution, one passenger is happier than the other one.

- Assume that there are 2 identical passengers and 1 driver. The value of being unmatched (for the passengers and the driver alike) is 0. The value of being matched is 1, both for the passengers and driver.
- In a model with prices (Uber model–transferable utility), the price of the ride will be 1, so that the driver's payoff is 2, and both passengers' payoffs is zero. Thus, passengers are indifferent between being matched and unmatched.
- In a classical model without transfers (taxi model–nontransferable utility), there are two stable matchings in each of which the matched passenger gets one, while the unmatched gets zero. Thus in this Gale-Shapley solution, one passenger is happier than the other one.

- Assume that there are 2 identical passengers and 1 driver. The value of being unmatched (for the passengers and the driver alike) is 0. The value of being matched is 1, both for the passengers and driver.
- In a model with prices (Uber model–transferable utility), the price of the ride will be 1, so that the driver's payoff is 2, and both passengers' payoffs is zero. Thus, passengers are indifferent between being matched and unmatched.
- In a classical model without transfers (taxi model–nontransferable utility), there are two stable matchings in each of which the matched passenger gets one, while the unmatched gets zero. Thus in this Gale-Shapley solution, one passenger is happier than the other one.

- However, people don't like to be unhappier than their peers!
  - For example, passengers will fight for the only available taxi...
  - ... or they will wait in line, and the length of the line will make each passenger indifferent between waiting in line and opting out.
  - In both cases, the driver is not better off, but both passengers have destroyed utility so that they are indifferent between being matched or unmatched, and both passengers have the same payoff (i.e., zero) at equilibrium.
- If, on the contrary, there are two drivers and one passengers, the story is reversed: drivers will fight / wait in line, and destroy utility so that both drivers get zero payoff; in this case, the passenger gets surplus one.

Section 3

Aggregate deferred acceptance

- Algorithm.
  - Let $\mu_{xy}^{A,0} = n_x$.
  - At step t, pick

$$\begin{cases} \mu_{xy}^{P,t} \in \arg\max_{\mu \in \mathbb{N}^{\mathcal{X} \times \mathcal{Y}}} \left\{ \sum_{xy} \mu_{xy} \alpha_{xy} : \mu_{xy} \leq \mu_{xy}^{A,t-1}, \sum_{y \in \mathcal{Y}} \mu_{xy} \leq n_x \;\; \left[u_x^t\right] \right\} \\ \mu_{xy}^{D,t} \in \arg\max_{\mu \in \mathbb{N}^{\mathcal{X} \times \mathcal{Y}}} \left\{ \sum_{xy} \mu_{xy} \gamma_{xy} : \mu_{xy} \leq \mu_{xy}^{P,t}, \sum_{x \in \mathcal{X}} \mu_{xy} \leq m_y \;\; \left[v_y^t\right] \right\} \end{cases}$$

    and update the available offers

$$\mu_{xy}^{A,t} = \mu_{xy}^{A,t-1} - \left(\mu_{xy}^{P,t} - \mu_{xy}^{D,t} \right)$$

  - When $\mu_{xy}^{D,t} = \mu_{xy}^{P,t}$, stop.
- Note that when $n_x = 1$ for all x and $m_y = 1$ for all y, this is *exactly* Gale and Shapley.
- **Theorem**. The algorithm converges in a finite number T of steps and $\left(\mu_{xy}^{D,T}, u_x^T, v_y^T \right)$ is an aggregate CSM.

- Assume $\alpha_{ij} = \alpha_{x_i y_j} + \varepsilon_{i y_j}$ and $\gamma_{ij} = \gamma_{x_i y_j} + \eta_{x_i j}$, ε and η iid Gumbel, as Choo and Siow (2006) and G. and Salanié (2016), but here in the NTU case.
- As in the TU case, one can show that the equilibrium waiting times τ_{xy}^S and τ_{xy}^D only depend on the observable characteristics x and y.
- Letting $U_{xy} = \alpha_{xy} - \tau_{xy}^S$ and $V_{xy} = \gamma_{xy} - \tau_{xy}^D$, we have

$$u_i = \max_{y \in \mathcal{Y}} \left\{ U_{xy} + \varepsilon_{iy}, \varepsilon_{i0} \right\}$$

$$v_j = \max_{x \in \mathcal{X}} \left\{ V_{xy} + \eta_{xj}, \eta_{0j} \right\}$$

so the equilibrium relates U_{xy} to the supply-side conditional choice probabilities μ_{xy} / n_x, V_{xy} to the demand-side ccp μ_{xy} / m_y, and U_{xy} to V_{xy} by

$$\max \left\{ U_{xy} - \alpha_{xy}, V_{xy} - \gamma_{xy} \right\} = 0.$$

- Assume $\alpha_{ij} = \alpha_{x_i y_j} + \varepsilon_{i y_j}$ and $\gamma_{ij} = \gamma_{x_i y_j} + \eta_{x_i j}$, ε and η iid Gumbel, as Choo and Siow (2006) and G. and Salanié (2016), but here in the NTU case.
- As in the TU case, one can show that the equilibrium waiting times τ^S_{xy} and τ^D_{xy} only depend on the observable characteristics x and y.
- Letting $U_{xy} = \alpha_{xy} - \tau^S_{xy}$ and $V_{xy} = \gamma_{xy} - \tau^D_{xy}$, we have

$$u_i = \max_{y \in \mathcal{Y}} \left\{ U_{xy} + \varepsilon_{iy}, \varepsilon_{i0} \right\}$$

$$v_j = \max_{x \in \mathcal{X}} \left\{ V_{xy} + \eta_{xj}, \eta_{0j} \right\}$$

so the equilibrium relates U_{xy} to the supply-side conditional choice probabilities μ_{xy} / n_x, V_{xy} to the demand-side ccp μ_{xy} / m_y, and U_{xy} to V_{xy} by

$$\max \left\{ U_{xy} - \alpha_{xy}, V_{xy} - \gamma_{xy} \right\} = 0.$$

- Assume $\alpha_{ij} = \alpha_{x_i y_j} + \varepsilon_{i y_j}$ and $\gamma_{ij} = \gamma_{x_i y_j} + \eta_{x_i j}$, ε and η iid Gumbel, as Choo and Siow (2006) and G. and Salanié (2016), but here in the NTU case.
- As in the TU case, one can show that the equilibrium waiting times τ^S_{xy} and τ^D_{xy} only depend on the observable characteristics x and y.
- Letting $U_{xy} = \alpha_{xy} - \tau^S_{xy}$ and $V_{xy} = \gamma_{xy} - \tau^D_{xy}$, we have

$$u_i = \max_{y \in \mathcal{Y}} \left\{ U_{xy} + \varepsilon_{iy}, \varepsilon_{i0} \right\}$$

$$v_j = \max_{x \in \mathcal{X}} \left\{ V_{xy} + \eta_{xj}, \eta_{0j} \right\}$$

so the equilibrium relates U_{xy} to the supply-side conditional choice probabilities μ_{xy} / n_x, V_{xy} to the demand-side ccp μ_{xy} / m_y, and U_{xy} to V_{xy} by

$$\max \left\{ U_{xy} - \alpha_{xy}, V_{xy} - \gamma_{xy} \right\} = 0.$$

- Galichon-Salanié showed that if one defines

$$\begin{cases} G(U) = \mathbb{E}\left[\sum_{x\in\mathcal{X}} n_x \max_{y\in\mathcal{Y}} \{U_{xy} + \varepsilon_y, \varepsilon_0\}\right] \\ H(V) = \mathbb{E}\left[\sum_{y\in\mathcal{Y}} m_y \max_{x\in\mathcal{X}} \{V_{xy} + \eta_x, \eta_0\}\right] \end{cases}$$

then at equilibrium, one gets

$$U = \nabla G^*(\mu) \text{ and } V = \nabla H^*(\mu)$$

where $G^*(\mu) = \max_U \left\{\sum_{xy} \mu_{xy} U_{xy} - G(U)\right\}$ and $H^*(\mu) = \max_V \left\{\sum_{xy} \mu_{xy} V_{xy} - H(V)\right\}$ are the Legendre-Fenchel transforms of G and H.

- In the present context, the equilibirum equations boils down to

$$\max\left\{\frac{\partial G^*}{\partial \mu_{xy}}(\mu) - \alpha_{xy}, \frac{\partial H^*}{\partial \mu_{xy}}(\mu) - \gamma_{xy}\right\} = 0.$$

- Galichon-Salanié showed that if one defines

$$\begin{cases} G(U) = \mathbb{E}\left[\sum_{x\in\mathcal{X}} n_x \max_{y\in\mathcal{Y}}\left\{U_{xy}+\varepsilon_y, \varepsilon_0\right\}\right] \\ H(V) = \mathbb{E}\left[\sum_{y\in\mathcal{Y}} m_y \max_{x\in\mathcal{X}}\left\{V_{xy}+\eta_x, \eta_0\right\}\right] \end{cases}$$

then at equilibrium, one gets

$$U = \nabla G^*(\mu) \text{ and } V = \nabla H^*(\mu)$$

where $G^*(\mu) = \max_U \left\{\sum_{xy} \mu_{xy} U_{xy} - G(U)\right\}$ and $H^*(\mu) = \max_V \left\{\sum_{xy} \mu_{xy} V_{xy} - H(V)\right\}$ are the Legendre-Fenchel transforms of G and H.

- In the present context, the equilibirum equations boils down to

$$\max\left\{\frac{\partial G^*}{\partial \mu_{xy}}(\mu) - \alpha_{xy}, \frac{\partial H^*}{\partial \mu_{xy}}(\mu) - \gamma_{xy}\right\} = 0.$$

- Assume ε and η iid Gumbel, as Choo and Siow, but here in the NTU case. In this case, the ccp inversion is explicit, as $U_{xy} = \ln \mu_{xy} / \mu_{x0}$ and $V_{xy} = \ln \mu_{xy} / \mu_{0y}$.
- Thus one has existence and uniqueness of an equilibrium, and

$$\mu_{xy} = \min \left(\mu_{x0} e^{\alpha_{xy}}, \mu_{0y} e^{\gamma_{xy}} \right). \quad (1)$$

where

$$\begin{cases} \mu_{x0} + \sum_{y \in \mathcal{Y}} \min \left(\mu_{x0} e^{\alpha_{xy}}, \mu_{0y} e^{\gamma_{xy}} \right) = n_x \\ \mu_{0y} + \sum_{x \in \mathcal{X}} \min \left(\mu_{x0} e^{\alpha_{xy}}, \mu_{0y} e^{\gamma_{xy}} \right) = m_y \end{cases}$$

and this system can be efficiently solved with a nonlinear version of the IPFP (a.k.a. RAS/Sinkhorn/matrix scaling) algorithm.
- Computationally, scales extremely well – easily parallelizable.

- Assume ε and η iid Gumbel, as Choo and Siow, but here in the NTU case. In this case, the ccp inversion is explicit, as $U_{xy} = \ln \mu_{xy} / \mu_{x0}$ and $V_{xy} = \ln \mu_{xy} / \mu_{0y}$.
- Thus one has existence and uniqueness of an equilibrium, and

$$\mu_{xy} = \min \left(\mu_{x0} e^{\alpha_{xy}}, \mu_{0y} e^{\gamma_{xy}} \right). \quad (1)$$

where

$$\begin{cases} \mu_{x0} + \sum_{y \in \mathcal{Y}} \min \left(\mu_{x0} e^{\alpha_{xy}}, \mu_{0y} e^{\gamma_{xy}} \right) = n_x \\ \mu_{0y} + \sum_{x \in \mathcal{X}} \min \left(\mu_{x0} e^{\alpha_{xy}}, \mu_{0y} e^{\gamma_{xy}} \right) = m_y \end{cases}$$

and this system can be efficiently solved with a nonlinear version of the IPFP (a.k.a. RAS/Sinkhorn/matrix scaling) algorithm.

- Computationally, scales extremely well – easily parallelizable.

- Assume ε and η iid Gumbel, as Choo and Siow, but here in the NTU case. In this case, the ccp inversion is explicit, as $U_{xy} = \ln \mu_{xy} / \mu_{x0}$ and $V_{xy} = \ln \mu_{xy} / \mu_{0y}$.
- Thus one has existence and uniqueness of an equilibrium, and

$$\mu_{xy} = \min\left(\mu_{x0} e^{\alpha_{xy}}, \mu_{0y} e^{\gamma_{xy}}\right). \tag{1}$$

where

$$\begin{cases} \mu_{x0} + \sum_{y \in \mathcal{Y}} \min\left(\mu_{x0} e^{\alpha_{xy}}, \mu_{0y} e^{\gamma_{xy}}\right) = n_x \\ \mu_{0y} + \sum_{x \in \mathcal{X}} \min\left(\mu_{x0} e^{\alpha_{xy}}, \mu_{0y} e^{\gamma_{xy}}\right) = m_y \end{cases}$$

and this system can be efficiently solved with a nonlinear version of the IPFP (a.k.a. RAS/Sinkhorn/matrix scaling) algorithm.
- Computationally, scales extremely well – easily parallelizable.

- Our MMF has

$$\mu_{xy} = \min\left(\mu_{x0} e^{\alpha_{xy}}, \mu_{0y} e^{\gamma_{xy}}\right). \quad (2)$$

- Note contrast between (1) and Dagsvik-Menzel equilibrium, which assumes $\alpha_{ij} = \alpha_{x_i y_j} + \varepsilon_{ij}$ and $\gamma_{ij} = \gamma_{x_i y_j} + \eta_{ij}$ and where

$$\mu_{xy} = \mu_{x0}\mu_{0y} e^{\alpha_{xy}} e^{\gamma_{xy}}. \quad (3)$$

- Our MMF has
$$\mu_{xy} = \min\left(\mu_{x0}e^{\alpha_{xy}}, \mu_{0y}e^{\gamma_{xy}}\right). \tag{2}$$
- Note contrast between (1) and Dagsvik-Menzel equilibrium, which assumes $\alpha_{ij} = \alpha_{x_i y_j} + \varepsilon_{ij}$ and $\gamma_{ij} = \gamma_{x_i y_j} + \eta_{ij}$ and where
$$\mu_{xy} = \mu_{x0}\mu_{0y}e^{\alpha_{xy}}e^{\gamma_{xy}}. \tag{3}$$

Section 4

Multinomial choice under rationing

- τ_{xy} can be interpeted as a shadow price of the capacity constraint. Consider the constrained maximum welfare problem

$$\bar{G}(\alpha, \bar{\mu}) = \max_{\mu \geq 0} \sum_{xy} \alpha_{xy} \mu_{xy} - G^*(\mu)$$
$$s.t.\ \mu_{xy} \leq \bar{\mu}_{xy} \ \ [\tau_{xy} \geq 0]$$

- Then, classically

$$\bar{G}(\alpha, \bar{\mu}) = G(\alpha - \tau) + \sum_{xy} \bar{\mu}_{xy} \tau_{xy}, \text{ and}$$
$$\partial \bar{G}(\alpha, \bar{\mu}) / \partial U_{xy} = \partial G(\alpha - \tau) / \partial U_{xy}.$$

- A natural measure of the market inefficiency is the total time waited in line: $\sum_{xy} \mu_{xy} \tau_{xy}$. It is lost to the passengers, and not appropriated by the taxi drivers.

- **Theorem 1**. The shadow price τ_{xy} is an antitone function of the vector of number of available offers $\bar{\mu}$.
- **Theorem 2**. The number of nondemanded options $\bar{\mu} - \mu$ is an isotone function of the capacity vector $\bar{\mu}$.
- Comments:
  - Theorem 1 is intuitive: it says that when the constraint becomes tighter ($\bar{\mu}$ decreases), the vector of Lagrange multipliers τ increases.
  - Theorem 2 expresses gross substitutes: it says that when there are more options, the options that were dominated are still dominated.

- **Proof of Theorem 1**. $\bar{G}(\alpha, \bar{\mu}) = \min_{\tau \geq 0} \left\{ G(\alpha - \tau) + \sum_{xy} \tau_{xy} \bar{\mu}_{xy} \right\}$, hence $\tau = \arg\max_{\tau \geq 0} \left\{ -G(\alpha - \tau) + \sum_{xy} \tau_{xy} (-\bar{\mu}_{xy}) \right\}$. By Topkis' theorem, τ is an isotone function of $-\bar{\mu}$, hence an antitone function of $\bar{\mu}$.

- **Proof of Theorem 2**. $\bar{\mu} - \mu = \bar{\mu} - \partial \bar{G} / \partial \alpha$, hence $\partial (\bar{\mu} - \mu) / \partial \bar{\mu} = I - \partial^2 \bar{G} / \partial \alpha \partial \bar{\mu} = I - \partial \tau / \partial \alpha = \partial U / \partial \alpha$, where $U = \alpha - \tau$. But

$$U = \arg\max_{0 \leq U \leq \alpha} \left\{ -G(U) + \sum_{xy} (U_{xy} - \alpha_{xy}) \bar{\mu}_{xy} \right\},$$

so that by Topkis' theorem again, U is an isotone function of α. Thus, $\partial U / \partial \alpha$ is entrywise positive; hence, so is $\partial (\bar{\mu} - \mu) / \partial \bar{\mu}$. As a result, $\bar{\mu} - \mu$ is isotone in $\bar{\mu}$.

- In the logit case, we look for $\tau_{xy} \geq 0$ such that

$$\alpha_{xy} - \tau_{xy} = \log \frac{\mu_{xy}}{\mu_{x0}}$$
$$\tau_{xy} > 0 \Longrightarrow \mu_{xy} = \bar{\mu}_{xy}$$

- Thus, the demand is given by $\mu_{xy} = \min\left(\bar{\mu}_{xy}, \mu_{x0} e^{\alpha_{xy}}\right)$, where μ_{x0} solves the scalar equation

$$\mu_{x0} + \sum_{y \in \mathcal{Y}} \min\left(\bar{\mu}_{xy}, \mu_{x0} e^{\alpha_{xy}}\right) = n_x.$$

(very easy to solve for μ_{x0} numerically).

Section 5

THE AGGREGATE DEFERRED ACCEPTANCE ALGORITHM WITH RANDOM UTILITY

- Step 0. Initialize by

$$\mu_{xy}^{A,0} = n_x.$$

- Step $t \geq 1$.
  - Proposal phase: Passengers make proposals subject to availability constraint:

$$\mu^{P,t} \in \arg\max_{\mu} \left\{ \sum \mu_{xy} \alpha_{xy} - G^* \left(\mu \right) : \mu \leq \mu^{A,t-1} \ \left[\tau^{G,t} \geq 0 \right] \right\}.$$

  - Disposal phase: Taxis pick up their best offers among the proposals:

$$\mu^{D,t} \in \arg\max_{\mu} \left\{ \sum \mu_{xy} \gamma_{xy} - H^* \left(\mu \right) : \mu \leq \mu^{P,t} \ \left[\tau^{H,t} \geq 0 \right] \right\}.$$

  - Update phase: The number of available offers is decreased according to the number of rejected ones

$$\mu^{A,t} = \mu^{A,t-1} - \left(\mu^{P,t} - \mu^{D,t} \right).$$

- We show convergence by showing a series of facts.
  - Fact 1: Tentatively accepted offers remain in place at the next period: $\mu^{D,t} \leq \mu^{P,t+1}$.
  - Fact 2: As t grows, $\tau^{G,t}$ weakly increases and $\tau^{H,t}$ weakly decreases.
  - Fact 3: At every step t, $\min\left(\tau_{xy}^{G,t}, \tau_{xy}^{H,t}\right) = 0$.
  - Fact 4: As $t \to \infty$, $\lim \nabla G\left(\alpha - \tau^{G,t}\right) = \lim \nabla H\left(\gamma - \tau^{H,t}\right) =: \mu$.
- As a result, $\left(\mu_{xy}, \tau_{xy}^{G,t}, \tau_{xy}^{H,t}\right)$ is an equilibrium with non-price rationing.

- Tentatively accepted offers remain proposed at the next period: $\mu^{D,t} \leq \mu^{P,t+1}$.
- **Proof**: By theorem 2, $\mu^{A,t} \leq \mu^{A,t-1}$ implies $\mu^{A,t} - \mu^{P,t+1} \leq \mu^{A,t-1} - \mu^{P,t}$, thus $\mu^{A,t} - \mu^{A,t-1} + \mu^{P,t} \leq \mu^{P,t+1}$. Thus, $\mu^{D,t} \leq \mu^{P,t+1}$.

- As t grows, $\tau^{G,t}$ weakly increases and $\tau^{H,t}$ weakly decreases.
- **Proof**:
  - One has $\mu_{xy}^{A,t-1} \leq \mu_{xy}^{A,t}$, thus as ∇G^* is isotone, $\nabla G^* \left(\mu^{A,t-1} \right) \leq \nabla G^* \left(\mu^{A,t} \right)$, hence $\alpha_{xy} - \tau_{xy}^{G,t-1} \leq \alpha_{xy} - \tau_{xy}^{G,t}$.
  - To see that $\tau^{H,t} \geq \tau^{H,t-1}$, note that

$$\tau_{xy}^{H,t} = \partial H \left(\gamma, \mu^{D,t} \right) / \partial \bar{\mu}_{xy}$$

$$\tau_{xy}^{H,t+1} = \partial H \left(\gamma, \mu^{P,t+1} \right) / \partial \bar{\mu}_{xy}$$

and $\mu^{D,t} \leq \mu^{P,t+1}$ along with the fact that $\partial H \left(\gamma, \bar{\mu} \right) / \partial \bar{\mu}_{xy}$ is antitone in $\bar{\mu}$ (Theorem 1) allows to conclude.

- At every step t, $\min\left(\tau_{xy}^{G,t}, \tau_{xy}^{H,t}\right) = 0$.
- **Proof**: $\tau_{xy}^{H,t} > 0$ implies $\tau_{xy}^{H,s} > 0$ for $s \in \{1, ..., t\}$; hence $\mu_{xy}^{P,s} = \mu_{xy}^{D,s}$, hence $\mu_{xy}^{A,t-1} = \mu_{xy}^{A,0} = n_x$. Assume $\tau_{xy}^{G,t} > 0$. Then it means that the corresponding constraint is saturated, which means $\mu_{xy}^{P,t} = \mu_{xy}^{A,t-1} = n_x$, a contradiction.

- As $t \to \infty$, $\lim \nabla G\left(\alpha - \tau^{G,t}\right) = \lim \nabla H\left(\gamma - \tau^{H,t}\right) =: \mu$.
- **Proof**: One has $\mu^{A,t-1} - \mu^{A,t} = \mu^{P,t} - \mu^{D,t} = \nabla G\left(\alpha - \tau^{G,t}\right) - \nabla H\left(\gamma - \tau^{H,t}\right)$, but as $\mu^{A,t}$ is nonincreasing and bounded, this quantity tends to zero. Further, $\tau^{G,t}$ and $\tau^{H,t}$ converge monotonically, which shows that $\lim_t \nabla G\left(\alpha - \tau^{G,t}\right) = \lim_t \nabla H\left(\gamma - \tau^{H,t}\right)$.

- A way to reconcile Gale-Shapley and competitive equilibrium.
- Large-scale computation of the logit model is easy.
- Empirical framwork allows to do econometrics.
- Comparative statics are easy to derive.
- More on this in my 'math+econ+code' week-long masterclass on competitive equilibrium, NYU, May 21-26, 2018

http://alfredgalichon.com/mec_equil