# MATCHING MODELS WITH GENERAL TRANSFERS

Alfred Galichon (NYU Econ+Courant)

Summer school "Optimal transport meets economics," July 2018
Lecture 2, July 23. Separable matching models

# Learning objectives: lecture 2

- Matching with unobserved heterogeneities
- Microfoundations of regularized optimal transport
- Estimation of matching models

- [CS] Choo and Siow (2006). "Who Marries Whom and Why," *Journal of Political Economy*.
- [GS] Galichon and Salanié (2012-2018). "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," Preprint.
- Decker, Lieb, McCann, and Stephens (2012). "Unique Equilibria and Substitution Effects in a Stochastic Model of the Marriage Market," *Journal of Economic Theory*.
- [DG] Dupuy and Galichon (2014). "Personality traits and the marriage market," *Journal of Political Economy*.
- [COQ2] Chiappori, Oreffice and Quintana-Domeque (2016). "Bidimensional Matching with Heterogeneous Preferences in the Marriage Market," Preprint.
- [CSW] Chiappori, Salanié, and Weiss (2017). "Partner Choice and the Marital College Premium," AER.
- [L] Low (2016). "Pricing the Biological Clock: Reproductive Capital on the US Marriage Market," Preprint.

Section 1

# INTRODUCTION: THE ECONOMICS OF MARRIAGE

- In the footsteps of Becker, empirical studies on the marriage market had long been focused on one-dimensional models.
- However, it is desirable to move beyond one-dimensional models. There are often many observed characteristics, and it is not always the case that the sorting can be captured by a single-dimensional model. One-dimensional models typically predict positive assortative matching, a theoretical prediction stemming from assumptions of supermodularity of the surplus function, which does not necessarly hold in the data. Fortunately, optimal transport provide tools to study multidimensional models.
- Yet any model of matching based on optimal transport will not be exploitable because it will generate far too strong predictions, namely that some matchings will never hold. This is rather counterfactual: in the data, one observes virtually any combination of type.
- Hence, need to regularize the matching model, and we shall do so by introducing unobserved heterogeneity. The model so obtained will be exploitable for estimation and identification purposes. The first such model (with transfers) is the model by Choo and Siow (2006). We shall see a generalization of this model by G and Salanié (2012).

Section 2

# Finite population model

- The reference for this lecture is [CS] and [GS]. Consider a heterosexual marriage matching market. The set of types (observable characteristics) is $\mathcal{X}$ for men, and $\mathcal{Y}$ for women. There are $n_x$ men of type $x$, and $m_y$ women of type $y$.
- Assume that if a man $i \in \mathcal{I}$ of type $x_i$ and a woman $j \in \mathcal{J}$ of type $y_j$ match, they get respective utilities

$$\alpha_{x_i y_j} + w_{ij} + \varepsilon_{iy_j}$$
$$\gamma_{x_i y_j} - w_{ij} + \eta_{x_i j}$$

  where $w_{ij}$ is the transfer from $i$ to $j$. If they remain single $i$ and $j$ get respectively $\varepsilon_{i0}$ and $\eta_{0j}$.
- The random utility vectors $(\varepsilon_y)$ and $(\eta_x)$ are drawn from probability distributions $\mathbf{P}_x$ and $\mathbf{Q}_y$, respectively. In the sequel we shall work with a finite number of agents of each type, and then we'll investigate the limit of these results.

- The matching surplus between $i$ and $j$ is therefore

$$\tilde{\Phi}_{ij} = \Phi_{x_i y_j} + \varepsilon_{iy_j} + \eta_{x_i j}$$

where $\Phi_{xy} = \alpha_{xy} + \gamma_{xy}$. The value of optimal matching is thus, under its dual form,

$$\begin{aligned} &\min_{u_i, v_j} \sum_{i \in \mathcal{I}} u_i + \sum_{j \in \mathcal{J}} v_j \\ &s.t.\ u_i + v_j \geq \Phi_{x_i y_j} + \varepsilon_{iy_j} + \eta_{x_i j} \\ &\quad\quad u_i \geq \varepsilon_{i0} \\ &\quad\quad v_j \geq \eta_{j0} \end{aligned}$$

- Written like this, the lp has $|\mathcal{I}| + |\mathcal{J}|$ variables and $|\mathcal{I}| \times |\mathcal{J}| + |\mathcal{I}| + |\mathcal{J}|$ constraints. Assuming that there are $K$ individuals per type for each type, this is $K\left(|\mathcal{X}| + |\mathcal{Y}|\right)$ variables and $K^2\left(|\mathcal{X}| \times |\mathcal{Y}|\right) + K\left(|\mathcal{X}| + |\mathcal{Y}|\right)$ constraints.
- The number of constraints is **quadratic** with respect to $K$. Fortunately, economic reasoning will help us reduce this complexity.

## LEMMA

*Consider the set $\mathcal{I}_{xy}$ of men of type $x$ matched to women of type $y$ at equilibrium. If $\mathcal{I}_{xy}$ is nonempty, then $u_i - \varepsilon_{iy}$ is a constant across $\mathcal{I}_{xy}$.*

## PROOF.

For $i \in \mathcal{I}$ such that $x_i = x$,

$$\begin{aligned} u_i &= \max_{j \in \mathcal{J}} \left\{ \tilde{\Phi}_{ij} - v_j, \varepsilon_{i0} \right\} \\ &= \max_{y \in \mathcal{Y}} \left\{ U_{xy} + \varepsilon_{iy}, \varepsilon_{i0} \right\} \end{aligned}$$

where $U_{xy} = \max_{j : y_j = y} \left\{ \Phi_{xy} + \eta_{x_i j} - v_j \right\}$, thus $u_i \geq U_{xy} + \varepsilon_{iy}$ with equality on $\mathcal{I}_{xy}$. With similar notations, $v_j \geq V_{xy} + \eta_{xj}$ with equality on $\mathcal{J}_{xy}$. As a result, if $\mathcal{I}_{xy}$ is nonempty, then $U_{xy} + V_{xy} = \Phi_{xy}$ and $\forall i \in \mathcal{I}_{xy}$, $u_i = U_{xy} + \varepsilon_{iy}$. □

- In the sequel, we shall see that *adding* an auxiliary variable to the previous lp will lead to *decreasing* the computational complexity of the problem.
- Observe that the first set of constraints is reexpressed by saying that, for every $x \in \mathcal{X}$, $y \in \mathcal{Y}$,

$$\min_{i:x_i=x} \{u_i - \varepsilon_{iy}\} + \min_{j:y_j=y} \{v_j - \eta_{xj}\} \geq \Phi_{xy}.$$

- Hence, letting $U_{xy} = \min_{i:x_i=x} \{u_i - \varepsilon_{iy}\}$ and $V_{xy} = \min_{j:y_j=y} \{v_j - \eta_{xj}\}$, a solution of the previous lp should satisfy

$$u_i = \max_{y \in \mathcal{Y}} \{U_{xy} + \varepsilon_{iy}, \varepsilon_{i0}\} \text{ and } v_j = \max_{x \in \mathcal{X}} \{V_{xy} + \varepsilon_{xj}, \varepsilon_{0j}\}.$$

# A simplification (ctd)

- The problem rewrites as

$$\min_{u_i, v_j, U_{xy}, V_{xy}} \sum_{i \in \mathcal{I}} u_i + \sum_{j \in \mathcal{J}} v_j \tag{1}$$

$$\begin{aligned} s.t.\ & U_{xy} + V_{xy} \geq \Phi_{xy} \;\; [\mu_{xy} \geq 0] \\ & u_i \geq U_{x_i y} + \varepsilon_{i y_j} \;\; [\mu_{iy}] \\ & v_j \geq V_{x y_j} + \eta_{x_i j} \;\; [\mu_{xi}] \\ & u_i \geq \varepsilon_{i0} \;\; [\mu_{i0}] \\ & v_j \geq \eta_{j0} \;\; [\mu_{0x}] \end{aligned}$$

- This problem has $K\left(|\mathcal{X}| + |\mathcal{Y}|\right) + |\mathcal{X}| \times |\mathcal{Y}|$ variables and $\left(|\mathcal{X}| \times |\mathcal{Y}|\right) + K\left(2\,|\mathcal{X}| \times |\mathcal{Y}| + |\mathcal{X}| + |\mathcal{Y}|\right)$ constraints.
- The number of constraint is now **linear** with respect to $K$.

- Lagrange multipliers:
  - The Lagrange multiplier $\mu_{xy}$ is interpreted as the number of matchings between types $x$ and $y$.
  - The Lagrange multiplier $\mu_{iy}$ ($y \in \mathcal{Y}_0$) is interpreted as a 0-1 indicator that man $i$ chooses a type $y$
  - The Lagrange multiplier $\mu_{xj}$ ($x \in \mathcal{X}_0$) is interpreted as a 0-1 indicator that woman $j$ chooses a type $x$
- Utilities:
  - $i$ solves a discrete choice problem $u_i = \max_{y \in \mathcal{Y}} \left\{ U_{xy} + \varepsilon_{iy}, \varepsilon_{i0} \right\}$
  - $j$ solve a discrete choice problem $v_j = \max_{x \in \mathcal{X}} \left\{ V_{xy} + \eta_{xj}, \eta_{0j} \right\}$.
- $U_{xy}$ and $V_{xy}$ are related by $U_{xy} + V_{xy} \geq \Phi_{xy}$ with equality if $\mu_{xy} > 0$.

Section 3

# LARGE POPULATION MODEL

- [GS] look at the limit of previous markets when the number of market participants gets large, holding fixed the frequency of each types.
- In the large population limit $n_x$ and $m_y$ are now interpreted as the mass distribution of respective types $x$ and $y$.
- We shall from now on assume that $\mathbf{P}_x$ and $\mathbf{Q}_y$, the distributions of random utility vectors $(\varepsilon_y)$ and $(\eta_x)$, have a density with full support. This will ensure that $G$, $G^*$, $H$ and $H^*$ are continuously differentiable.

- Under these assumptions, Problem (1) becomes

$$\min_{U,V} G(U) + H(V)$$
$$s.t.\ U_{xy} + V_{xy} \geq \Phi_{xy} \ \ [\mu_{xy}]$$

where

$$G(U) = \sum_{x\in\mathcal{X}} n_x \mathbb{E}_{\mathbf{P}}\left[\max_{y\in\mathcal{Y}}\left\{U_{xy} + \varepsilon_{iy}, \varepsilon_{i0}\right\}\right]$$
$$H(V) = \sum_{y\in\mathcal{Y}} m_y \mathbb{E}_{\mathbf{Q}}\left[\max_{x\in\mathcal{X}}\left\{V_{xy} + \eta_{xj}, \eta_{0j}\right\}\right]$$

- By first order conditions,

$$\frac{\partial G(U)}{\partial U_{xy}} = \mu_{xy} = \frac{\partial H(V)}{\partial V_{xy}}.$$

and $\mu_{xy} > 0$ for every $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

- The primal problem corresponding to (1) is

$$\max_{\mu_{xy}\geq 0} \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \mu_{xy}\Phi_{xy} - \mathcal{E}\left(\mu\right)$$

where

$$\mathcal{E}\left(\mu\right) = G^{*}\left(\mu\right) + H^{*}\left(\mu\right)$$

- Recall $G^{*}\left(\mu\right) = \max\left\{\sum_{xy}\mu_{xy}U_{xy} - G\left(U\right)\right\}$ is the Legendre transform of $G$, and similarly for $H^{*}$.
- By first order conditions, we get the identifcation formula of $\Phi$

$$\Phi_{xy} = \frac{\partial G^{*}\left(\mu\right)}{\partial\mu_{xy}} + \frac{\partial H^{*}\left(\mu\right)}{\partial\mu_{xy}}.$$

Section 4

# Choo and Siow's logit model

- In Choo and Siow's model [CS], the heterogeneities in tastes are Gubmel, this

$$\mathcal{E}\left(\mu\right) = 2\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \mu_{xy}\log\mu_{xy} + \sum_{x\in\mathcal{X}} \mu_{x0}\log\mu_{x0} + \sum_{y\in\mathcal{Y}} \mu_{0y}\log\mu_{0y}.$$

  Note that $\mathcal{E}\left(\mu\right) < \infty$ if and only if $\mu \in \mathcal{M}\left(n, m\right)$.

- We call this model the *TU-logit* model. Tomorrow, we shall see how to generalize this to *ITU-logit* models, including *NTU-logit*.
- By first order conditions above, Choo-Siow's TU-logit model implies the following matching function:

$$\mu_{xy} = M_{xy}\left(\mu_{x0}, \mu_{0y}\right) := \sqrt{\mu_{x0}}\sqrt{\mu_{0y}}\exp\left(\frac{\Phi_{xy}}{2}\right).$$

- As a result, $\partial \mathcal{E}(\mu) / \partial \mu_{xy} = 2 \log \mu_{xy} - \log \mu_{x0} - \log \mu_{0y}$, which implies that $\Phi_{xy}$ is estimated by *Choo and Siow's identification formula*

$$\hat{\Phi}_{xy} = \log \frac{\hat{\mu}_{xy}^2}{\hat{\mu}_{x0} \hat{\mu}_{0y}}.$$

- Consistent with our previous remarks, only the joint surplus $\Phi_{xy} = \alpha_{xy} + \gamma_{xy}$ is identified. However, if the transfers $\hat{w}_{xy}$ are observed too (e.g. wages in labour market), then $U_{xy} = \alpha_{xy} + w_{xy}$ and $V_{xy} = \gamma_{xy} - w_{xy}$, so that $\alpha$ and $\gamma$ are separately identified by

$$\begin{cases} \hat{\alpha}_{xy} = \log \frac{\hat{\mu}_{xy}}{\hat{\mu}_{x0}} - \hat{w}_{xy} \\ \hat{\gamma}_{xy} = \log \frac{\hat{\mu}_{xy}}{\hat{\mu}_{0y}} + \hat{w}_{xy} \end{cases}$$

- Write down the equilibrium equations in the TU-logit model:

$$\begin{cases} \sum_{y \in \mathcal{Y}} \sqrt{\mu_{x0}} \sqrt{\mu_{0y}} \exp\left(\frac{\Phi_{xy}}{2}\right) + \mu_{x0} = n_x \\ \sum_{x \in \mathcal{X}} \sqrt{\mu_{x0}} \sqrt{\mu_{0y}} \exp\left(\frac{\Phi_{xy}}{2}\right) + \mu_{0y} = m_y \end{cases}$$

- Setting $a_x = \sqrt{\mu_{x0}}$, $b_y = \sqrt{\mu_{0y}}$, and $K_{xy} = \exp\left(\Phi_{xy}/2\right)$, this rewrites as

$$\begin{cases} \sum_{y \in \mathcal{Y}} K_{xy} a_x b_y + a_x^2 = n_x \\ \sum_{x \in \mathcal{X}} K_{xy} a_x b_y + b_y^2 = m_y \end{cases}$$

- Hence, the IPFP consists in iteratively solving quadratic equations, i.e.

$$\begin{cases} a_x^{2t+1} = \sqrt{n_x + \left(\sum_{y\in\mathcal{Y}} b_y^{2t} K_{xy}/2\right)^2} - \sum_{y\in\mathcal{Y}} b_y^{2t} K_{xy}/2 \\ b_y^{2t+2} = \sqrt{m_y + \left(\sum_{x\in\mathcal{X}} a_x^{2t+1} K_{xy}/2\right)^2} - \sum_{x\in\mathcal{X}} a_x^{2t+1} K_{xy}/2 \end{cases}$$

- This is implemented by:

```
Kb_x = c(K_xy %*% b_y)
a_x = sqrt(n_x + Kb_x * Kb_x /4 ) - Kb_x / 2
Ka_y = c(t(a_x) %*% K_xy )
b_y = sqrt(m_y + Ka_y * Ka_y /4 ) - Ka_y / 2
```

- Cf [GS]. The dual problem is given by

$$\min_{u,v}\left\{\begin{array}{c}\sum_x n_x u_x + \sum_y m_y v_y \\ +2\sum_{xy}\sqrt{n_x m_y}\exp\left(\frac{\Phi_{xy}-u_x-v_y}{2}\right) \\ +\sum_x n_x \exp\left(-u_x\right)+\sum_y m_y \exp\left(-v_y\right)\end{array}\right\}$$

- Remarks:
  - This problem is an unconstrained convex optimization problem, so this formulation will be quite useful.
  - If $(u, v)$ is solution, $u_x = -\log \mu_{0|x} = -\log\left(\mu_{x0}/n_x\right)$ and $v_y = -\log\left(\mu_{0|y}\right)$.
  - Note that the IPFP algorithm just seen interprets as (blockwise) **coordinate descent** in the dual problem.

- Now assume that the scaling factor (=standard deviation) of the Gumbel is $\sigma$. Instead of $\varepsilon_y$ and $\eta_x$, these heterogeneities are $\sigma\varepsilon_y$ and $\sigma\eta_x$. As a result, the entropy is now $\sigma\mathcal{E}(\mu)$ instead of $\mathcal{E}(\mu)$.
- The expression of the social welfare is now

$$\max_{\mu\in\mathcal{M}(n,m)} \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \mu_{xy}\Phi_{xy} - \sigma\mathcal{E}(\mu)$$

which tends to the usual optimal assignment problem when $\sigma \to 0$ (no heterogeneity). Tends to random matching solution when $\sigma \to +\infty$.

- Assume $\Phi_{xy}$ is parameterized as

$$\Phi_{xy}^{\lambda} = \sum_k \lambda_k \phi_{xy}^k$$

so that the welfare function can be written

$$W(\lambda) = \max_{\mu \in \mathcal{M}(n,m)} \{ \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy}^{\lambda} - \sigma \mathcal{E}(\mu) \}$$

$$= \min_{u,v} \left\{ \begin{array}{c} \sum_x n_x u_x + \sum_y m_y v_y + 2 \sum_{xy} \sqrt{n_x m_y} \exp\left( \frac{\Phi_{xy}^{\lambda} - u_x - v_y}{2} \right) \\ + \sum_x n_x \exp(-u_x) + \sum_y m_y \exp(-v_y) \end{array} \right\}$$

- As a result,

$$\frac{\partial W}{\partial \lambda_k} = \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy}^{\lambda} \phi_{xy}^k$$

where $\mu^{\lambda}$ is the optimal $\mu$ given $\lambda$.

- We will pick the parameter $\lambda$ such that $\sum_{xy} \mu^{\lambda}_{xy} \phi^k_{xy} = \sum_{xy} \hat{\mu}_{xy} \phi^k_{xy}$ for each $k$.
- For this, do

$$\min_{\lambda} \left\{ W(\lambda) - \sum_{xy} \hat{\mu}_{xy} \Phi^{\lambda}_{xy} \right\}$$

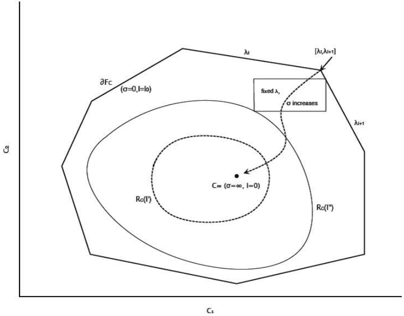
which reexpresses using the dual expression as

$$\min_{u,v,\lambda} \left\{ \begin{array}{c} \sum_x n_x u_x + \sum_y m_y v_y + 2 \sum_{xy} \sqrt{n_x m_y} \exp\left( \frac{\Phi^{\lambda}_{xy} - u_x - v_y}{2} \right) \\ + \sum_x n_x \exp(-u_x) + \sum_y m_y \exp(-v_y) - \sum_{xy} \hat{\mu}_{xy} \Phi^{\lambda}_{xy} \end{array} \right\}.$$

- The set of vectors $C^{\lambda}$ where $C_k^{\lambda} = \sum_{xy} \mu_{xy}^{\lambda} \phi_{xy}^{k}$ is a convex and compact set called *covariogram*. Its boundaries correspond to solutions of the Monge-Kantorovich problem without regularization; the point 0 corresponds to the independent coupling ($\|\Phi\| / \sigma \to 0$).
- Note that increasing $\lambda^k$ will increase the weight of $\phi^k$ in the optimization problem, and thus will increase the corresponding moment $C_k^{\lambda}$.
- We look for $\lambda$ such that the predicted covariances $C_k^{\lambda}$ equates the observed covariances $\hat{C}_k = \sum_{xy} \hat{\mu}_{xy} \phi_{xy}^{k}$, which is computed from our sample by

$$\hat{C}_k = \frac{1}{n} \sum_{i=1}^{n} \phi_{x_i y_i}^{k}.$$

# FIGURE



Section 5

# Numerical experiments

- In '`benchmark.R`', we compare a number of different methods to perform equilibrium computation in a TU-logit model.
- Take a toy example where $x$=age of man, and $y$=age of woman. There are 25 categories of age in the studied population, between 16 and 40, and the surplus is
$$\Phi\left(x,y\right)=-\left|x-y\right|$$
and the heterogeneities are iid Gumbel.
- $n_x$ and $m_y$ are taken from Choo and Siow's ACS data stored in the subdirectory '`/ChooSiowData`'.

- In this course, we have discussed a number of techniques to compute equilibrium:
  - Gradient descent
  - Newton descent
  - Jacobi iteration (IPFP)
  - Linear programming (after discretization of heterogeneity)
- We will now design them, implement them and benchmark them.

- Recall dual version

$$\min_{(U_{xy})} \left\{ G\left(U\right) + H\left(\Phi - U\right) \right\} \tag{2}$$

where

$$G\left(U\right) = \sum_{x \in \mathcal{X}} n_x \log \left( 1 + \sum_{y \in \mathcal{Y}} \exp\left(U_{xy}\right) \right)$$

$$H\left(V\right) = \sum_{y \in \mathcal{Y}} m_y \log \left( 1 + \sum_{x \in \mathcal{X}} \exp\left(V_{xy}\right) \right)$$

- This is a convex optimization problem in dimension $|\mathcal{X}| \times |\mathcal{Y}|$.

- To solve the optimization problem

$$\min W(U)$$

where $U$ is convex, gradient descent consists in

$$U_{t+1} = U_t - \epsilon_t \nabla W(U_t)$$

for $\epsilon_t > 0$ small enough.

- Intuition: with enough smoothness

$$\begin{aligned} W(U_{t+1}) &= W(U_t) + \langle \nabla W(U_t), U_{t+1} - U_t \rangle + O\left( \|U_{t+1} - U_t\|^2 \right) \\ &= W(U_t) - \epsilon_t \|\nabla W(U_t)\|^2 + O\left( \epsilon_t^2 \right). \end{aligned}$$

- More on this soon (Keith).

- If $W(U) = G(U) + H(\Phi - U)$, then

$$\nabla W(U) = \nabla G(U) - \nabla H(\Phi - U)$$

  which interprets as the *market imbalance* (supply minus demand for $xy$ matches).

- Here, with logit heterogeneities,

$$\frac{\partial W(U)}{\partial U_{xy}} = \frac{n_x \exp(U_{xy})}{1 + \sum_{y \in \mathcal{Y}} \exp(U_{xy})} - \frac{m_y \exp(\Phi_{xy} - U_{xy})}{1 + \sum_{x \in \mathcal{X}} \exp(\Phi_{xy} - U_{xy})}.$$

- See implementation in '`benchmark.R`'.

- Newton descent consists in doing

$$U_{t+1} = U_t - \epsilon_t \left( D^2 W \left( U_t \right) \right)^{-1} \nabla W \left( U_t \right)$$

for $\epsilon_t > 0$ small enough.

- Intuition: when $\epsilon_t \to 0$, $(U_t)$ tends to the solution of ODE

$$\frac{dU_t}{dt} = -\epsilon_t \left( D^2 W \left( U_t \right) \right)^{-1} \nabla W \left( U_t \right)$$

that is

$$\frac{d}{dt} \nabla W \left( U_t \right) = -\epsilon_t \nabla W \left( U_t \right)$$

which has solution

$$\nabla W \left( U_t \right) = \nabla W \left( U_0 \right) \exp \left( - \int_0^t \epsilon_s ds \right),$$

hence $\left\| \nabla W \left( U_t \right) \right\| \to 0$ as $t \to +\infty$.

- When $W(U) = G(U) + H(\Phi - U)$, then

$$D^2 W(U) = D^2 G(U) + D^2 H(\Phi - U)$$

which is called the *market curvature* in the hedonic equilibrium literature.

- Here, with logit heterogeneities,

$$\frac{\partial^2 G(U)}{\partial U_{xy} \partial U_{xy}} = \mu_{xy} - \frac{\mu_{xy}^2}{n_x}$$
$$\frac{\partial^2 G(U)}{\partial U_{xy} \partial U_{xy'}} = -\frac{\mu_{xy}\mu_{xy'}}{n_x} \text{ for } y \neq y'$$

where $\mu_{xy} = \partial G(U) / \partial U_{xy}$; while $\partial^2 G(U) / \partial U_{xy} \partial U_{x'y'} = 0$ for $x \neq x'$.

- Similar formulas hold for $D^2 H$.

- In Choo-Siow's TU-logit model, one has

$$\mu_{xy} = \sqrt{\mu_{x0}\mu_{0y}} \exp\left(\frac{\Phi_{xy}}{2}\right)$$

  which allowed us to reformulate equilibrium as a problem over variables $\mu_{x0}$ and $\mu_{0y}$.
- Set $a_x = -\log \mu_{x0}$ and $b_y = -\log \mu_{0y}$. Then one can view the problem as a *nodal problem*

$$\min \sum_{x\in\mathcal{X}} n_x a_x + \sum_{y\in\mathcal{Y}} m_y b_y + E\left(a, b\right) \tag{3}$$

  where $E\left(a, b\right) = 2\sum_{xy} e^{\frac{\Phi_{xy} - a_x - b_y}{2}} + \sum_x e^{-a_x} + \sum_y e^{-b_y}$.
- The problem has become an optimization problem in dimension $|\mathcal{X}| + |\mathcal{Y}|$.
  - Huge dimensionality reduction: from $|\mathcal{X}| \times |\mathcal{Y}|$ to $|\mathcal{X}| + |\mathcal{Y}|$, BUT
  - Nodal methods apply only in the logit case, while the previous methods (*edge methods*) work for any heterogeneity.

- We will solve for the equilibrium $(\mu_{x0}, \mu_{0y})$ using:
  - Gradient descent on problem (3)
  - Newton descent on problem (3)
  - IPFP

- If $F(a,b) = \sum_{x\in\mathcal{X}} n_x a_x + \sum_{y\in\mathcal{Y}} m_y b_y + E(a,b)$, then

$$\frac{\partial F}{\partial a_x} = n_x - \sum_{y\in\mathcal{Y}} e^{\frac{\Phi_{xy}-a_x-b_y}{2}} - e^{-a_x}$$

$$\frac{\partial F}{\partial b_y} = m_y - \sum_{x\in\mathcal{X}} e^{\frac{\Phi_{xy}-a_x-b_y}{2}} - e^{-b_y}$$

  which interprets as another market imbalance measure.
- In gradient descent, $a_x$ and $b_y$ adjust proportionally to market imbalance.

- If $F(a,b) = \sum_{x\in\mathcal{X}} n_x a_x + \sum_{y\in\mathcal{Y}} m_y b_y + E(a,b)$, then

$$\frac{\partial^2 F}{\partial a_x \partial a_{x'}} = 1_{\{x=x'\}} \left\{ \frac{1}{2} \sum_{y\in\mathcal{Y}} e^{\frac{\Phi_{xy}-a_x-b_y}{2}} + e^{-a_x} \right\}$$

$$\frac{\partial^2 F}{\partial b_y \partial b_{y'}} = 1_{\{y=y'\}} \left\{ \frac{1}{2} \sum_{x\in\mathcal{X}} e^{\frac{\Phi_{xy}-a_x-b_y}{2}} + e^{-b_y} \right\}$$

$$\frac{\partial^2 F}{\partial a_x \partial b_y} = \frac{1}{2} e^{\frac{\Phi_{xy}-a_x-b_y}{2}}$$

- Recall that setting $a_x = \sqrt{\mu_{x0}}$ and $b_y = \sqrt{\mu_{0y}}$ one can employ the following iterative scheme

$$\begin{cases} a_x^{2t+1} = \sqrt{n_x + \left(\sum_{y\in\mathcal{Y}} b_y^{2t} K_{xy}/2\right)^2} - \sum_{y\in\mathcal{Y}} b_y^{2t} K_{xy}/2 \\ b_y^{2t+2} = \sqrt{m_y + \left(\sum_{x\in\mathcal{X}} a_x^{2t+1} K_{xy}/2\right)^2} - \sum_{x\in\mathcal{X}} a_x^{2t+1} K_{xy}/2 \end{cases}$$

- The algorithm converges for any starting point $a^0$.

- Recall that the matching surplus between $i$ and $j$ is therefore

$$\tilde{\Phi}_{ij} = \Phi_{x_i y_j} + \varepsilon_{iy_j} + \eta_{x_i j}$$

where $\Phi_{xy} = \alpha_{xy} + \gamma_{xy}$. The value of optimal matching is thus, under its dual form,

$$\begin{aligned} \min_{u_i, v_j} \sum_{i \in \mathcal{I}} u_i + \sum_{j \in \mathcal{J}} v_j \\ s.t.\ u_i + v_j \geq \Phi_{x_i y_j} + \varepsilon_{iy_j} + \eta_{x_i j} \\ u_i \geq \varepsilon_{i0} \\ v_j \geq \eta_{j0} \end{aligned}$$

- Here, $\varepsilon_{iy_j}$ and $\eta_{x_i j}$ are i.i.d. draws from standard Gumbel distributions.
- This way of doing things is not competitive, but interesting to keep in mind.

- In '`benchmark.R`', e compute equilibrium in the toy model, for five of the algorithms we described:
  - Edge gradient descent
  - Edge Newton descent
  - Nodal gradient descent
  - Nodal Newton descent
  - IPFP
  - Linear programming
- The algorithms are implemented in R. Gradient descent is done via the package `nloptr`; Newton descent via `nleqslv`; linear programming via Gurobi. Benchmark is done via the `microbenchmark` package.

| **Algorithm** | **Time taken (microsec.)** | **Nb of iterations** |
|---|---|---|
| edge-gradient | 38,457 | 71 |
| edge-newton | 1,184,230 | 28 |
| nodal-gradient | 5,837 | 39 |
| nodal-newton | 5,194 | 52 |
| ipfp | 837 | 41 |

## BENCHMARK: COMMENTS

- The IPFP is a clear winner. In spite of the fact that R is not good with loops, it beats the next most efficient procedure by a factor 5.
- edge-newton is penalized by large matrix inversions. However, it converges in a remarkably low number of iterations. Could be sped up if matrix inversion is done efficiently.