'MATH+ECON+CODE' MASTERCLASS ON MATCHING MODELS, OPTIMAL TRANSPORT AND APPLICATIONS

Alfred Galichon (New York University)

Spring 2019
Day 5, January 18 2019: "Empirical matching models"
Block 15. Rank constained models

LEARNING OBJECTIVES: BLOCK 15

- affinity matrix
- index models
- rank-constraint models

- Becker (1973). A Theory of Marriage: Part I. *JPE*.
- [COQ] Chiappori, Oreffice and Quintana-Domeque (2012). "Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market," *Journal of Political Economy*.

- Chiappori, Oreffice and Quintana-Domeque [COQ] assume individuals match on a scalar "index of attractiveness" subsuming BMI, salary, education. Then the surplus function is

$$\Phi(x, y) = \left(\sum_k \zeta_k x_k\right) \left(\sum_l \nu_l y_l\right)$$

  $\zeta_k / \zeta_{k'}$ and $\nu_l / \nu_{l'}$ are marginal rates of substitution: how much richer do men/women need to be in order to compensate an increase in Body Mass Index?

- This problem can be solves by looking for the vectors of weights ζ and ν such that the rank correlation of $\zeta^\intercal x$ and $\nu^\intercal y$ is maximal.

- [COQ] look at the characteristics of married couples, in particular body mass index, wages, and education.
- According to [COQ] (*Journal of Political Economy*, 2012): "Men may compensate 1.3 additional units of BMI with a 1%-increase in wages, while women may compensate two BMI units with one year of education."

- Recall

$$\mathcal{W}(A) = \max_{\pi \in \mathcal{M}(P,Q)} \int x' A y d\pi(x,y) - \sigma \int \pi(x,y) \, d\pi(x,y).$$

and note that

$$\frac{\partial \mathcal{W}(A)}{\partial A_{ij}} = C_{ij}^{A}$$

- We are therefore looking for the estimator $\hat{A}$ of the true A such that $\partial \mathcal{W}(A) / \partial A_{ij} = \hat{C}_{ij}$.
- Thus we shall estimate A by $\hat{A}$ the solution of

$$\min_{A} \left\{ \mathcal{W}(A) - \sum_{ij} A_{ij} \hat{C}_{ij} \right\}$$

which is a nice and smooth convex minimization problem.

- Several proposal to estimate ζ and ν:
  1. Becker (1973): use Hotelling's canonical correlation analysis
  $$\max_{\zeta,\nu} \mathbb{E}\left[\zeta^{\intercal} X Y^{\intercal} \nu\right],$$
  which is unbiased if (X, Y) is Gaussian. Can be biased outside that case, cf. Dupuy-Galichon (AES, 2015).
  2. Chiappori, Oreffice and Quintana-Domeque (JPE 2013): when Y is 1-dimensional, regress Y on X.
  3. Terviö (AER 2007): maximize Spearman's rank correlation
  $$\max_{\zeta,\nu} \mathbb{E}\left[F_{\zeta^{\intercal} X}\left(\zeta^{\intercal} X\right) F_{\nu^{\intercal} Y}\left(\nu^{\intercal} Y\right)\right],$$
  where $F_{\zeta^{\intercal} X}$ and $F_{\nu^{\intercal} Y}$ are the cdfs of $\zeta^{\intercal} X$ and $\nu^{\intercal} Y$ respectively.
  4. In the spirit of Han (JE 1987), maximize
  $$\sum_{ij}\left(1\left\{\zeta^{\intercal} X_i > \zeta^{\intercal} X_j\right\} 1\left\{\nu^{\intercal} Y_i > \nu^{\intercal} Y_j\right\} + 1\left\{\zeta^{\intercal} X_i < \zeta^{\intercal} X_j\right\} 1\left\{\nu^{\intercal} Y_i < \nu^{\intercal} Y_j\right\}\right)$$
  5. Dupuy-Galichon-Sun (2017): perform rank-constrained estimation of $\Phi(x, y) = x' A y$ using nuclear norm regularization.

- Recall that any $d \times d$ matrix A has a singular value decomposition

$$A = U \Lambda V^{\intercal}$$

where U and V are orthogonal matrices, and $\Lambda = diag\left(\lambda_1, ..., \lambda_d\right)$ is diagonal with positive entries ordered in descending order, i.e. $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d \geq 0$.

- Note:
  - Λ are the eigenvalues of $AA^{\intercal}$, and also of $A^{\intercal}A$.
  - If A is symmetric positive, then Λ are the eigenvalues of A
  - The rank of A is the number of nonzero entries of λ.
- The nuclear norm of A, denoted $\left|A\right|_*$, is simply the L1 norm of λ, that is

$$\left|A\right|_* = \sum_{i=1}^{d} \lambda_i.$$

  - Controlling for nuclear norm is a good proxy for controlling for rank.
  - Further, the nuclear norm is convex.

- The nuclear norm can be expressed as

$$\left|A\right|_* = \max_{U,V \in O_d} Tr\left(U^\intercal AV\right)$$

from which its gradient may be inferred (from the envelope theorem).

- In general, one can use the nuclear norm for problems of the type

$$\min_A W\left(A\right) + \gamma \left|A\right|_*$$

which will drive low-rank solutions.