Matching with Trade-offs: Revealed Preferences over Competing Characteristics

Alfred Galichon Bernard Salanié

Maison des Sciences Economiques

16 October 2009

Main idea: Matching involve trade-offs

E.g in the marriage market partners vary across several dimensions, and their preferences also vary.
How efficiently does the marriage market manage all these trade-offs?
what do observed patterns reveal to us about preferences for assortativeness on education, age, wealth, beauty...?
and how do the underlying preferences translate into endogamy with respect to these characteristics?

A Theorist's Answer

RTFM (Roth-Sotomayor, *Cambridge University Press*, 1989.)
Objection: too general, does not generate any testable restriction.

An Empiricist's Reply

RTFM (Becker, *Journal of Political Economy*, 1973-74.)

Objection: one-dimensional types, no heterogeneity, stark but too strong predictions.

A step forward: Choo-Siow, *JPE* 2006—inspired some of what follows.

Agenda

- Impose structural assumptions on the surplus generated by any potential match
- Characterize the properties of the set of feasible matchings, and of socially optimal matchings;
- Estimation: use data on covariation of the types of the partners in observed matches
- in order to:
    - characterize social preferences, and
    - given a class of surplus functions, test that the observed matches are socially optimal for a member of this class.

Maintained Assumptions

- We take the populations to be matched (non-singles) as given—ignores some information that may be useful.
- Matches are bipartite—we talk about men and women, but that is irrelevant (and can be endogenized!)
- (For now) observable types are discrete.

Types

- Two N-size populations M and W must be matched; each man $\tilde{x} \in M$ must be matched with one and only one woman $\tilde{y} \in W$;
- Each individual has a full type=an observable type + a type that is observed to all agents *but not to econometrician*. A man has $\tilde{x} = (x, \varepsilon)$, with an r-dimensional observable type x. A woman has $\tilde{y} = (y, \eta)$, with an s-dimensional observable type y.
- Let $\tilde{p}$ (resp. $\tilde{q}$) be the distributions over full types of men resp. women, and p (resp. q) the induced distributions over observable types.

Matching

- A *matching* $\tilde{\pi}(\tilde{x}, \tilde{y})$ is the specification of "who marries whom", ie. the proportion of couples with men with full type $\tilde{x}$ and women with full type $\tilde{y}$. We have

$$\sum_{\tilde{x}} \tilde{\pi}(\tilde{x}, \tilde{y}) = \tilde{q}(\tilde{y}) \text{ and } \sum_{\tilde{y}} \tilde{\pi}(\tilde{x}, \tilde{y}) = \tilde{p}(\tilde{x})$$

  which we express as $\tilde{\pi} \in \mathcal{M}(\tilde{p}, \tilde{q})$.
- Similar notations for observable types: $\pi(x, y) \in \mathcal{M}(p, q)$.

Surpluses and separability

- Matching a man with full type $\tilde{x}$ and women with full type $\tilde{y}$ produces a *joint surplus* $\tilde{\Phi}\left(\tilde{x}, \tilde{y}\right)$, known to all participants.
- we impose
  Assumption S. There exist deterministic functions Φ, ξ and χ such that

$$\tilde{\Phi}(\tilde{x}, \tilde{y}) = \Phi(x, y) + \chi\left(\tilde{x}, y\right) + \xi\left(\tilde{y}, x\right).$$

- we normalize ξ and χ to have zero mean.

Chiappori-Salanié-Tillman-Weiss: if utility is transferable, this implies that there exist two functions

$$U(x, y) + V(x, y) = \Phi(x, y)$$

such that if $\tilde{x} = (x, \varepsilon)$ marries y (of whichever η) he gets utility

$$U(x, y) + \chi\left(\tilde{x}, y\right).$$

Revealed Preferences

Our goal is to estimate $\Phi(x, y)$, the surplus on observable types given that we observe:

- the distributions of observable types p and q
- and the proportions $\pi(x, y)$ of matches on observable types.

Optimal matching

- An *optimal matching* maximizes social surplus

$$\mathcal{W} = \sup_{\tilde{\pi} \in \mathcal{M}(\tilde{p}, \tilde{q})} E_{\tilde{\pi}} \left[\tilde{\Phi} \left(\tilde{X}, \tilde{Y} \right) \right]. \tag{1}$$

  (classical assignment problem).
- [**Remark on equilibrium:** Although it is not on today's agenda, welfare theorems say that optimal matchings coincide with outcome of decentralized equilibria: "Cupido's invisible hand".]

Gumbel assumption

To make this practical we make assumptions that generate a multinomial logit structure (as in Choo-Siow 2006):
the $\chi(\tilde{x}, y)$ are independent of each other and of y, Gumbel with scale factor σ_1
the $\xi(\tilde{y}, x)$ are independent of each other and of x, Gumbel with scale factor σ_2.
Multinomial logit: man x matches with woman y with probability

$$\pi(y|x) \propto exp(U(x,y)/\sigma_1).$$

Specification of the observable surplus

With discrete types, Φ can be written as a linear combination of indicator functions:

$$\Phi(X, Y) \equiv \sum_{x,y} \Lambda_{x,y} \mathbb{1}(X = x, Y = y).$$

In many cases the analyst will want to restrict this further:

$$\Phi(X, Y) \equiv \sum_{k=1}^{K} \Lambda_k \phi^k(X, Y)$$

for unknown parameters Λ and known *basis functions* ϕ^k.

Quadratic example: $\Phi_\Lambda(x, y) = -\sum_{k=1}^{K} \Lambda_k \left(x^k - y^k\right)^2$, where Λ_k is the preference for endogamy on characteristic k.

Optimal surplus

Theorem. Under separability+Gumbel, the social surplus is

$$\mathcal{W}(\theta) = \sup_{\pi \in \mathcal{M}(p,q)} \left(E_\pi \left[\Phi_\Lambda (X, Y) \right] - \sigma I (\pi) \right) + \sigma_1 S (p) + \sigma_2 S (q) .$$

where $I(\pi) = \sum \pi(x, y) \ln \frac{\pi(x,y)}{p(x)q(y)}$ is the *mutual information* of π,

$\sigma = \sigma_1 + \sigma_2$ measures total heterogeneity,
and S is entropy: e.g. $S(p) = -\sum_x p(x) \ln p(x)$.

Intuition:

- with $\sigma = 0$ heterogeneity, the observed matching should maximize observable surplus, obviously;
- with zero information on types, $\sigma = \infty$ and the optimal matching should maximize randomness, i.e. entropy: $\pi(x, y) \equiv p(x)q(y)$.

Feasible matchings

Now assume $\Phi_\Lambda = \sum_k \Lambda_k \phi^k$.

- For any observable matching π, we denote $C^k(\pi)$ the *covariation vector*:

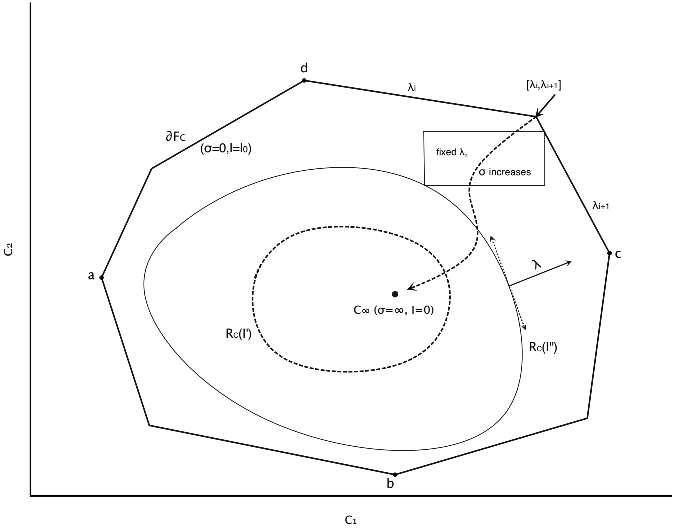
$$C^k(\pi) = E_\pi\left[\phi^k(X,Y)\right].$$

Quadratic example. $C^k(\pi) = E_\pi\left[\left(X^k - Y^k\right)^2\right]$

Nonparametric example. $C^{ij}(\pi) = \pi\left(x^i, y^j\right)$.

- The *feasible covariation set*, denoted $\mathcal{F}_C$, is the set of vectors C for which there exists a matching π such that $C = C(\pi)$. It is a convex set, and…

The feasible covariation set



Rationalizing information

Proposition.

1. Take any $C \in \mathcal{F}_c$; there exists a unique $I = I_r(C)$ such that for some Λ and σ,

$$\sup_{\pi \in \mathcal{M}(P,Q)} \left(E_\pi \left[\Phi_\Lambda (X, Y) \right] - \sigma I(\pi) \right) = C \cdot \Lambda - \sigma I$$

   Call $I_r(C)$ is the *rationalizing information* at C.
2. The function I_r is convex; it is minimized for the random matching and increases towards the boundary ∂F_c.
3. any vector Λ that rationalizes C is normal to the corresponding level set of $I_r(.)$,

$$\frac{dC^i}{dC^j} = -\frac{\Lambda_j}{\Lambda_i}.$$

Identification

Some remarks:

- Only total heterogeneity $\sigma = \sigma_1 + \sigma_2$ matters
- The model is homogeneous in (Λ, σ) – hence only $\lambda = \Lambda / \sigma$ matters.
- We shall discuss nonparametric identification (of $\Phi(x, y)$), then parametric identification (of Λ)

Nonparametric identification

Logit structure: there exists a decomposition $\Phi(x,y) = U(x,y) + V(x,y)$ such that an x-man matches with a y-woman with conditional probability

$$\pi(y|x) \propto exp(U(x,y)/\sigma_1)$$

For woman y we get

$$\pi(x|y) \propto exp(V(x,y)/\sigma_2)$$

Easy calculations give

$$\Phi(x,y) = U(x,y) + V(x,y) = (\sigma_1 + \sigma_2) \log \pi(x,y) + (...)$$

where (...) is additively separable in (x, y).
Therefore:

Φ is identified up to additively separable (and irrelevant) terms.

Parametric identification

Theorem. If $\sigma > 0$, the parameter $\lambda = \Lambda/\sigma$ is identified. It coincides with the gradient of the rationalizing information I_r:

$$\lambda_k = \frac{\partial I_r}{\partial C^k}(C).$$

Idea of proof, normalizing $\sigma = 1$: by the envelope theorem, one can show that $I_r(C)$ is the Legendre transform of the social welfare $\mathcal{W}(\lambda, 1)$

$$I_r(C) = \sup_{\lambda} \left(\sum_k \lambda_k C^k - \mathcal{W}(\lambda, 1) \right).$$

Remark: computing the function $\mathcal{W}(\lambda, 1)$ sounds daunting, but we shall show that it can be done very efficiently.

The Moment Matching estimator

We observe “who marries whom”: $\hat{\pi}(x, y)$, thus we can form the empirical covariations

$$\hat{C}^k = \sum_{x,y} \phi^k(x, y)\, \hat{\pi}(x, y)$$

Our *Moment Matching estimator* is defined as

$$\hat{\lambda}_k^{MM} = \frac{\partial I_r}{\partial C^k}\left(\hat{C}\right).$$

It has the property that the covariations predicted by the optimal matching model with parameter $\hat{\lambda}$ coincides with the empirical covariations $\hat{C}$.

Asymptotic behaviour

- Sampling error generates a Gaussian $O_P(1/\sqrt{N})$ error in estimating C;
- Then the error in estimating λ is given by

$$\hat{\lambda}_k^{MM} = \frac{\partial I_r}{\partial C^k}\left(\hat{C}\right) = \frac{\partial I_r}{\partial C^k}\left(C + \frac{z}{\sqrt{N}}\right) = \lambda_k + \frac{N(0, H_r)}{\sqrt{N}}$$

where H_r is the Hessian of the rationalizable information I_r.

Efficiency

- It can be shown that

$$\left(H_r\right)_{kl} = \frac{\partial^2 l_r}{\partial C^k \partial C^l} = \left(\mathcal{I}^{-1}\right)_{kl}$$

where

$$\mathcal{I}_{kl} = E\left[\frac{\partial \log \pi\left(X,Y\right)}{\partial \lambda_k}\frac{\partial \log \pi\left(X,Y\right)}{\partial \lambda_l}\right]$$

is the Fisher information matrix.

- **Hence the Moment Matching estimator is asymptotically efficient.**

The basic computational problem

Computing the Moment Matching estimator (or the equivalent maximum-likelihood estimator) requires solving many times

$$\mathcal{W}(\theta) = \sup_{\pi \in \mathcal{M}(P,Q)} \left(E_\pi \left[\Phi_\Lambda (X, Y) \right] - \sigma I (\pi) \right)$$

When $\sigma = 0$, classical assignment problem—"easy" but still costly.
When $\sigma > 0$, we know that $\pi(x, y)$ is of the form

$$\pi(x, y) = p(x) q(y) \exp\left(\frac{\Phi(x, y) - u(x) - v(y) - c}{\sigma} \right)$$

and this formula, along with the condition that $\pi \in \mathcal{M}(p, q)$, determines u and v.

The RAS algorithm

We use an Iterative Projection Fitting Procedure (IPFP), a.k.a. RAS algorithm, to determine u, v and therefore π.

IPFP/RAS minimizes the Kullback-Leibler distance of a given distribution to the set ofg joint distributions compatible with given marginals

Our problem can be restated to fit snugly within the IPFP setup.

Iterating

The algorithm iterates over values (u^k, v^k).
Define

$$z(x,y) = \frac{p(x)q(y)\exp(\Phi(x,y)/\sigma)}{\sum_{x,y} p(x)q(y)\exp(\Phi(x,y)/\sigma)}$$

1. Start from $u^0 \equiv -\sigma \log p$ and $v^0 \equiv 0$.
2. At step $(k+1)$ we compute iteratively

$$\exp(-v^{k+1}(y)/\sigma) = \frac{q(y)}{\sum_x z(x,y)\exp(-u^k(x)/\sigma)}$$

and

$$\exp(-u^{k+1}(x)/\sigma) = \frac{p(x)}{\sum_y z(x,y)\exp(-v^{k+1}(y)/\sigma)}.$$

The algorithm converges very fast to the solutions u and v.

Extensions

- our analysis is conditional on who stays single—loses some information but still valid.
- Strong connection with Afriat and Varian's generalized theory of Revealed preferences (GARP).
- and also with the screening problem in contract theory.