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

Alfred Galichon (New York University)

Tuesday: "Optimal transport I"
Block 5. One-dimensional matching

- Copulas and comonotonicity
- Positive Assortative Matching
- The Wage Equation

- [OTME], Ch. 4 and App. C
- Complements:
    - Becker (1973). A Theory of Marriage: Part I. *JPE*.
    - Sattinger (1993). Assignment Models of the Distribution of Earnings. *JEL*.
    - Gabaix and Landier (2008). Why has CEO Pay Increased So Much? *QJE*.
   - Terviö (2008). The Difference That CEOs Make: An Assignment Model Approach. *AER*.

Section 1

MOTIVATION

MOTIVATION

▶ Today we shall study one-dimensional models of matching. The application we shall envision is to the literature on CEO compensation: talented CEOs arguably generate more output if they run large firms than if they run a small business. Hence in these models, the CEO's talent is leveraged by the firm size.

▶ The output of a match is CEO's talent multiplied by a term capturing the firm's size. This surplus function belongs to a class which leads to positive assortative matching: the optimal matching will consists of assigning the most talented CEO with the biggest firm, etc.

▶ The model we shall follows a literature in family economics, see Becker (1973), and labor economics: Sattinger (1993), Terviö (2008), Gabaix and Landier (2008). Our application will be based on Gabaix and Landier's CEO compensation data. The difference with the empirical literature on marriage is that one observes typically less relevant characteristics on CEOs, but we observe the transfers, which are typically not observed whithin the couple.

- For each firm, we observe the size of the firm (market cap) and the compensation
- Distribution of CEO talent (to be defined later) is calibrated in their paper and we will take it as given; so will be the production function
- Our exercise will consist of predicting the CEO compensation. (Actually GL do the opposite: they use the distribution of wages to calibrate the production function).

Executive Compensation and Market Cap of Top 500 Firms
normalized to 1 in 1980
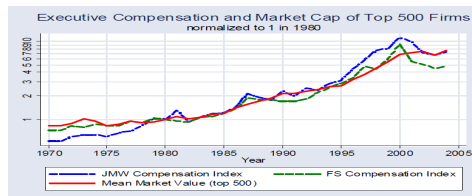
Section 2

POSITIVE ASSORTATIVE MATCHING

- Assume $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. This implies that both workers and firms are characterized by scalar characteristics x and y. Restrictive, but most of the applied literature in Economics to this day has focused on this case, which already generates interesting economic insights.

- Assume that there is the same number of workers and firms, and this number is normalized to one. Let P be the probability distribution of the workers types x, and Q the probability distribution of the firms types y.

- E.g. the literature on CEO compensation studies the matching problem of firms and managers: each manager has a measure of talent $x \in \mathbb{R}_+$ (extra return generated), and each firm has market capitalization $y \in \mathbb{R}_+$. Then, the economic value generated by a manager of talent x running a firm of size y is $\Phi(x, y) = xy$. Assume manager of talent x is assigned to firm of size $y = T(x)$. The constraint on the assignment map T is that $T \# P = Q$, which means that each firm is run by a manager and each manager runs a firm. Then the total value created is

$$\mathbb{E}_P\left[\Phi\left(X, T(X)\right)\right] = \mathbb{E}\left[XT(X)\right].$$

- Under a natural assuption on Φ, called *supermodularity*, we will see that the optimal coupling is such that $Y = T(X)$, where T is nondecreasing. T will be given an explicit characterization.

▶ Recall that given a probability distribution P on the real line, one can define the *quantile* of that distribution, denoted F_P^{-1}, which is a map from $[0, 1]$ to R, nondecreasing and continuous from the right, and such that if $U \sim \mathcal{U} := \mathcal{U}([0, 1])$, then

$$F_P^{-1}(U) \sim P. \qquad (1)$$

F_P^{-1} is the generalized inverse of the c.d.f. of P, F_P, the proper inverse when P has a positive density, in which case $U = F_P(X)$. Note also that if $X \sim P$, then $F_P(X)$ has a uniform distribution if and only if P has no mass point.

▶ Representation (1) extends to the case of bivariate distributions: for $\pi \in \mathcal{M}(P, Q)$, there exists a pair (U, V) of uniform random variables such that

$$\left(F_P^{-1}(U), F_Q^{-1}(V)\right) \sim \pi, \qquad (2)$$

and the c.d.f. associated with the distribution of (U, V) is called the *copula* associated with distribution π.

- A pair of random variables (X, Y) is *comonotone* if there is $U \sim \mathcal{U}$ such that $X = F_P^{-1}(U)$ and $Y = F_Q^{-1}(U)$. Equivalently, X and Y are said to exhibit *Positive Assortative Matching (PAM)*.
- The copula associated with a pair of comonotone random variables is the c.d.f. associated with (U, U), which is $F(u, v) = \min(u, v)$. This copula is called the *upper Fréchet-Hoeffding copula*.
- Note that when the cdf of X is continuous, there is a much simpler equivalent statement of comonotonicity:

LEMMA

If the distribution of X has no mass points, then X and Y are comonotone if and only if there exists a nondecreasing map T such that $Y = T(X)$. Moreover, one can choose $T(x) = F_Q^{-1}(F_P(x))$.

PROOF.

Consider $U \sim \mathcal{U}$ such that $X = F_P^{-1}(U)$ and $Y = F_Q^{-1}(U)$. If the distribution of X has no mass point, then $U = F_P(X)$. Hence, $Y = F_Q^{-1}(F_P(X))$. $\qquad\square$

- Assume Φ is *supermodular*, that is, for every scalars x, x', y and y',

$$\Phi\left(x \vee x', y \vee y'\right) + \Phi\left(x \wedge x', y \wedge y'\right) \geq \Phi\left(x, y\right) + \Phi\left(x', y'\right), \quad (3)$$

where $x \vee x'$ and $x \wedge x'$ denote respectively the maximum and the minimum between scalars x and x'. When Φ is twice continuously differentiable (which we will assume from now on), this is equivalent to

$$\frac{\partial^2 \Phi\left(x, y\right)}{\partial x \partial y} \geq 0. \quad (4)$$

- Assume that there are two types of workers $\mathcal{X} = \{\underline{x}, \bar{x}\}$ and firms $\mathcal{Y} = \{\underline{y}, \bar{y}\}$. An equivalent restatement of Condition (3) is then

$$\bar{x} \geq \underline{x} \text{ and } \bar{y} \geq \underline{y} \text{ implies } \Phi\left(\bar{x}, \bar{y}\right) + \Phi\left(\underline{x}, \underline{y}\right) \geq \Phi\left(\bar{x}, \underline{y}\right) + \Phi\left(\underline{x}, \bar{y}\right) \quad (5)$$

which asserts that the total output created is higher if the high types match together and the low types match together (assortative matching) rather than if mixed high/low pairs are formed.

EXAMPLE

The following examples of surplus functions are supermodular:

(i) Cobb-Douglas function: $\Phi(x, y) = x^a y^b$ $(x, y \geq 0)$, with $a, b \geq 0$,

(ii) General multiplicative form: $\Phi(x, y) = \zeta(x) \xi(y)$ with ζ and ξ nondecreasing,

(iii) Leontieff: $\Phi(x, y) = \min(x, y)$,

(iv) C.E.S. function: $\Phi(x, y) = (x^{-\rho} + y^{-\rho})^{-1/\rho}$, $\rho \geq 0$,

(v) $\Phi(x, y) = \phi(x - y)$ where ϕ is concave; in particular, $\Phi(x, y) = -|x - y|^p$, $p \geq 1$ or $\Phi(x, y) = -(x - y - k)^+$,

(vi) $\Phi(x, y) = \phi(x + y)$, where ϕ convex.

Section 3

THE REARRANGEMENT THEOREM

THEOREM

(i) Assume that Φ is supermodular. Then the primal of the Monge-Kantorovich problem

$$\sup_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_\pi \left[\Phi \left(X, Y \right) \right] \tag{6}$$

has a comonotone solution.

(ii) Conversely, if Problem (6) has a comonotone solution for any choice of probability distributions P and Q on the real line, then Φ is supermodular.

(iii) If, in addition, P has no mass points, then there is an optimal assignment which is is pure and satisfies $Y = T \left(X \right)$ where

$$T \left(x \right) = F_Q^{-1} \circ F_P \left(x \right). \tag{7}$$

The proof of part (i) is based on the following lemma.

LEMMA

Let Z_1 and Z_2 be two Bernoulli random variables of respective success probability p_1 and p_2. Then $\mathbb{E}[Z_1 Z_2] \leq \min(p_1, p_2)$.

PROOF.

As $Z_2 \leq 1$, $\mathbb{E}[Z_1 Z_2] \leq \mathbb{E}[Z_1] = p_1$. Similarly $\mathbb{E}[Z_1 Z_2] \leq \mathbb{E}[Z_2] = p_2$. Thus, $\mathbb{E}[Z_1 Z_2] \leq \min(p_1, p_2)$. □

We are now ready to sketch the proof of Theorem 3.

SKETCH OF PROOF OF THEOREM 3.

(i) Take $U \sim \mathcal{U}$, and $X = F_P^{-1}(U)$ and $Y = F_Q^{-1}(U)$. By (1), $X \sim P$ and $Y \sim Q$ and (X, Y) is comonotone by definition. The proof is in three steps.

Step 1. For $a, b \in \mathbb{R}$, consider surplus function
$\phi_{ab}(x, y) := 1\{x \geq a\} 1\{y \geq b\}$, and let $Z_1 = 1\{X \geq a\}$ and
$Z_2 = 1\{Y \geq b\}$. Z_1 and Z_2 are two Bernoulli random variables of
respective success probability $p_1 = 1 - F_P(a)$ and $p_2 = 1 - F_Q(b)$, thus
$\mathbb{E}[Z_1 Z_2] \leq \min(p_1, p_2)$, but a straightforward calculation shows that the
inequality actually holds as an equality. Hence (X, Y), which is comonotone,
is optimal for each surplus function ϕ_{ab}.

Step 2. Assume $\mathcal{X} = [\underline{x}, \bar{x}]$ and $\mathcal{Y} = [\underline{y}, \bar{y}]$ are compact intervals. Then

$$F(x, y) = \frac{\Phi(x, y) - \Phi(\underline{x}, y) - \Phi(x, \underline{y}) + \Phi(\underline{x}, \underline{y})}{\Phi(\bar{x}, \bar{y}) - \Phi(\underline{x}, \bar{y}) - \Phi(\bar{x}, \underline{y}) + \Phi(\underline{x}, \underline{y})}$$

is a c.d.f. associated to a probability measure ζ, and hence
$$F(x, y) = \iint \phi_{ab}(x, y) \, d\zeta(a, b). \qquad \square$$

PROOF.

As a result, if $\pi \in \mathcal{M}(p, q)$ is the distribution of (X, Y) where X and Y are comonotone, then

$$\int F(x, y) \, d\pi(x, y) \geq \int F(x, y) \, d\tilde{\pi}(x, y)$$

for every $\tilde{\pi} \in \mathcal{M}(p, q)$. But as F is of the form
$F(x, y) = K\Phi(x, y) + f(x) + g(y) + c$ with $K > 0$, and because
$\int \{f(x) + g(y) + c\} \, d\pi(x, y) = \int \{f(x) + g(y) + c\} \, d\tilde{\pi}(x, y)$ for every
$\tilde{\pi} \in \mathcal{M}(p, q)$, it results that

$$\int \Phi(x, y) \, d\pi(x, y) \geq \int \Phi(x, y) \, d\tilde{\pi}(x, y) \ \forall \tilde{\pi} \in \mathcal{M}(p, q)$$

which completes step 2. $\qquad \square$

PROOF.

Step 3. When $\mathcal{X}$ and $\mathcal{Y}$ are the real line, the result still holds by approximation.

(ii) The converse follows by taking for P the discrete probability with two mass points $\underline{x}$ and $\bar{x}$ with probability $1/2$ each, and Q the discrete probability with two mass points $\underline{y}$ and $\bar{y}$ also each with probability $1/2$. Then if (6) has a solution such that $F_P^{-1}(U)$ and $Y = F_Q^{-1}(U)$, for $U \sim \mathcal{U}([0,1])$, it follows that condition (5) holds.

(iii) follows from (i) and Lemma 1. $\qquad\square$

Note that the assumptions made in Theorem 3 do not guarantee that all the optimal assignments are comonotone. Indeed, the trivial example where $\Phi(x, y) = 0$ for every x and y provides an example of supermodular surplus function, for which any assignment is optimal. For this reason, we provide a strengthening of the previous result, which ensures uniqueness. We will assume Φ is strictly supermodular, that is if both $\bar{x} > \underline{x}$ and $\bar{y} > \underline{y}$ hold, then
$$\Phi(\bar{x}, \bar{y}) + \Phi(\underline{x}, \underline{y}) > \Phi(\bar{x}, \underline{y}) + \Phi(\underline{x}, \bar{y}).$$

THEOREM

Assume that Φ is strictly supermodular, and P has no mass point. Then the primal Monge-Kantorovich problem (6) has a unique optimal assignment, and this assignment is characterized by $Y = T(X)$ where T is given by (7).

THE WAGE EQUATION

▶ Assume (u, v) is a solution to the dual of the Monge-Kantorovich problem

$$\inf \mathbb{E}_P \left[u \left(X \right) \right] + \mathbb{E}_Q \left[v \left(Y \right) \right] \tag{8}$$
$$s.t. \ u \left(x \right) + v \left(y \right) \geq \Phi \left(x, y \right)$$

▶ Then $v \left(y \right)$ is intepreted as the value of the problem of a firm of type y, choosing the optimal worker x. Then the firm's program is

$$v \left(y \right) = \max_x \left\{ \Phi \left(x, y \right) - u \left(x \right) \right\}$$

thus by first order conditions, one is led to the *wage equation*

$$u' \left(x \right) = \frac{\partial \Phi}{\partial x} \left(x, T \left(x \right) \right), \tag{9}$$

where T is given by (7).

THEOREM

(i) Assume Φ is supermodular and continuously differentiable with respect to its first variable. Assume P has no mass point. Then the dual Monge-Kantorovich problem (8) has a solution (u, v). Further, u solves the wage equation (9). Hence, u is determined up to a constant c by

$$u(x) = c + \int_{x_0}^x \frac{\partial \Phi}{\partial x}(t, T(t)) \, dt. \tag{10}$$

(ii) Assume further that Q has no mass point, and that Φ is also continuously differentiable with respect to its second variable. Then v is given by

$$v(y) = c' + \int_{T(x_0)}^y \frac{\partial \Phi}{\partial y}\left(T^{-1}(z), z\right) dz, \tag{11}$$

where c and c' are related by $c + c' = \Phi(x_0, T(x_0))$.

INTERPRETATION: IMPERFECT COMPETITION AND RENTS

▶ It is important to distinguish between $\partial \Phi \left(x, T \left(x \right) \right) / \partial x$, which is the partial derivative of $\Phi \left(x, y \right)$ applied at $\left(x, y \right) = \left(x, T \left(x \right) \right)$, and $d \Phi \left(x, T \left(x \right) \right) / dx$, which is the total derivative of $\Phi \left(x, T \left(x \right) \right)$ with respect to x. One has

$$\frac{d \Phi \left(x, T \left(x \right) \right)}{dx} = \frac{\partial \Phi \left(x, T \left(x \right) \right)}{\partial x} + \frac{\partial \Phi \left(x, T \left(x \right) \right)}{\partial y} T' \left(x \right)$$

▶ This decomposition has an interesting interpretation in terms of differential rent. The total derivative $d \Phi \left(x, T \left(x \right) \right) / dx$ is the marginal increase in value between a firm run by a manager of talent x and a firm run by a manager of talent $x + dx$. This differential value is split between the manager's differential rent $\left(\partial \Phi \left(x, T \left(x \right) \right) / \partial x = u' \left(x \right) \right)$ and the firm's differential rent (
$\left(\partial \Phi \left(x, T \left(x \right) \right) / \partial y \right) T' \left(x \right) = dv \left(T \left(x \right) \right) / dx$).

▶ This discussion highlights the nature of the assignment model, which is a model of imperfect competition. In this model, managers compete against each other, and likewise on the other side of the market. Managers are imperfect substitutes for each other. This imperfect competition is the source of the rent, as made apparent by formula (9).

Section 4

APPLICATION: GABAIX AND LANDIER'S WAGE EQUATION

GABAIX-LANDIER'S MODEL OF CEO COMPENSATION

▶ Gabaix and Landier (2008) assume that

$$\Phi(x, y) = Cxy^{\gamma}$$

where C and γ are parameters.

▶ If X is the talent of the CEO, then the distribution of X can be parameterized through its quantile by

$$Q_X(t) = X^{\max} - \frac{B}{\beta N}(N(1-t))^{\beta}$$

(such a distribution is consistent with extreme value theory). N is the number of firms and β is a parameter.

▶ If Y is the size of the firm, then its distribution is assumed to be Pareto with distribution $1/\alpha$, i.e.

$$Q_Y(t) = A(N(1-t))^{-\alpha}$$

(they find that $\alpha = 1$, i.e. Zipf's law, fits the data quite well).

GABAIX-LANDIER'S WAGE EQUATION

- The wage of CEO indexed by $t \in [0,1]$ is $w(t)$; firm $y = Q_Y(t)$'s problem is

$$\max_{t \in [0,1]} \left\{ CQ_X(t) y^\gamma - w(t) \right\} = \max_{t \in [0,1]} \left\{ -C \frac{B}{\beta N} \left(N(1-t) \right)^\beta y^\gamma - w(t) \right\}$$

and by foc

$$
\begin{aligned}
w'(t) &= BCy^\gamma \left(N(1-t) \right)^{\beta-1} \\
&= A^\gamma BC \left(N(1-t) \right)^{-\alpha\gamma+\beta-1}
\end{aligned}
$$

where it is assumed $\alpha\gamma > \beta$. When N is large, this yields

$$w(t) \approx \frac{A^\gamma BC}{\alpha\gamma - \beta} \left(N(1-t) \right)^{-\alpha\gamma+\beta}$$

- "Superstar" effect (Rosen, 1981): if $\beta > 0$, then the wages are unbounded, even though talent is bounded – talent is very valuable at the top end of the distribution.

- If firm sizes scales by λ, then compensations should scale by λ^γ.
- Gabaix and Landier's calibration yields $\gamma \simeq 1$, hence constant return to scale: compensations should scale by the same amount than the firm sizes.
- They argue: "In the U.S., between 1980 and 2003, the average firm market value of the largest 500 firms (debt plus equity) has increased (in real terms) by a factor of 6 (i.e., a 500% increase) (...). Under constant returns to scale, CEO "productivity" has increased by 500%, and which made total pay increase by 500%."