Structural Estimation of Matching Markets with Transferable Utility

Alfred Galichon Bernard Salanié

November 2, 2020

In matching with transferable utility, by definition the partners in a match can exchange a good that has value for each partner. This still leaves room for several subcases: the utility possibility frontier within a match may be nonlinear (if for instance that good has diminishing marginal utility); transfers may be constrained; and the transfer may occasion frictions or other costs. We focus in this chapter on the simplest case, in which there is a good that has constant marginal utility, and the value of this marginal utility is the same for all agents. This defines *perfectly transferable utility*.

For simplicity, we also limit our discussion to the one-to-one bipartite model: each match consists of two partners, drawn from two separate subpopulations. The paradigmatic example is the *heterosexual marriage market*, in which the two subpopulations are men and women. We will use these terms for concreteness.

We assume that the econometrician observes a discrete set of characteristics of men and women: their education, their age, their income category etc. Each combination of the values of these characteristics defines a *type*. Section 1.1 introduces the main tools of this chapter when types contain all payoff-relevant information. In any real-world application, men and women of a given type will also vary in their preferences and more generally in their ability to create joint surplus in any match. We add this unobserved heterogeneity in Section 2.

Matching with transferable utility solves a linear programming problem. In recent years it has been analyzed with the methods of optimal transportation. Under an additional "separability" assumption, most functions of interest are convex; then convex duality gives a simple and transparent path to identification of the parameters of these models[1]. The empirical implementation is especially straightforward when the unobserved heterogeneity has a multinomial logit form and the joint surplus is linear in the parameters. Then the parameters can be estimated by minimizing a globally convex objective function. This is the approach presented in Galichon and Salanié (2020), on which much of this chapter is based.

[1] We collected the elements of convex analysis used in this chapter in an Appendix.

1 Matching without unobserved heterogeneity

1.1 Population and preferences

For the sake of exposition, we start by excluding any unobserved heterogeneity. The data of the problem is as follows. We are given sets of types $\mathcal{X}$ for men, and $\mathcal{Y}$ for women. There are n_x men of type x and m_y women of type y. The set of marital options that are offered to men and women is the set of types of partners on the other side of the market, plus singlehood. We introduce the notation 0 for singlehood and we define $\mathcal{X}_0 = \mathcal{X} \cup \{0\}$ and $\mathcal{Y}_0 = \mathcal{Y} \cup \{0\}$ as the set of marital options that are available to respectively women and men.

We denote by μ_{xy} the number of matches between men of type x and women of type y, which is determined at equilibrium. All men of type x, and all women of type y, must be single or matched. This generates the scarcity constraints:

$$
\begin{aligned}
N_x(\mu) &:= \sum_{y \in \mathcal{Y}} \mu_{xy} + \mu_{x0} = n_x \ \ \forall x \in \mathcal{X} \\
M_y(\mu) &:= \sum_{x \in \mathcal{X}} \mu_{xy} + \mu_{0y} = m_y \ \ \forall y \in \mathcal{Y}.
\end{aligned}
$$

If a man of type x and a woman of type y match, the assumption of perfectly transferable utility implies that their respective utilities can be written as

$$
\begin{aligned}
&\alpha_{xy} + t_{xy} \\
&\gamma_{xy} - t_{xy}
\end{aligned}
$$

where t_{xy} is the transfer from y to x. We assume that if an individual remains single, (s)he obtains utility zero—a harmless normalization. We assume that each individual knows the equilibrium values of the transfers for all matches that (s)he may take part in[2].

Notations. We will denote the vector of transfers by $t = (t_{xy})$ and matching patterns by $\mu = (\mu_{xy})$. For any doubly-indexed variable z_{xy}, we use the notation $z_{x\cdot}$ to denote the vector of values of z_{xy}; and we use a similar notation for $z_{\cdot y}$.

1.2 Marital demand

Consider a man of type x. He faces a simple discrete choice problem, which consists of choosing the marital option $y \in \mathcal{Y}_0$ that maximizes his payoff $\alpha_{xy} + t_{xy}$. Denoting $G_x\left(\alpha_{x\cdot} + t_{x\cdot}\right)$ the value of this program, we get

$$
G_x\left(\alpha_{x\cdot} + t_{x\cdot}\right) := \max_{y \in \mathcal{Y}_0} \left(\alpha_{xy} + t_{xy}\right). \tag{1}
$$

We formulate this problem as a linear programming problem by introducing the conditional probability $\mu^M_{y|x}$ that a man of type x matches with a woman of type y (for single men, $y = 0$.) Note that given that there are n_x men of type x, the total number of such matches will be $\mu^M_{xy} := n_x \mu^M_{y|x}$.

[2] We also adopt the notational convention $\alpha_{x0} = \gamma_{0y} = t_{x0} = t_{0y} = 0$.

For each $x \in \mathcal{X}$, the vector $\mu^M_{\cdot|x}$ must solve the linear program

$$G_x\left(\alpha_{x\cdot} + t_{x\cdot}\right) = \max_{\mu_{\cdot|x} \geq 0} \sum_{y \in \mathcal{Y}} \mu_{y|x}\left(\alpha_{xy} + t_{xy}\right) : \sum_{y \in \mathcal{Y}} \mu_{y|x} + \mu_{0|x} = 1. \tag{2}$$

As the maximum of a family of linear functions of its arguments over a convex domain, G_x is a convex function. Moreover, since (2) is a linear program, it has a dual whose value is also $G_x(\alpha_{x\cdot} + t_{x\cdot})$. Denote u_x the multiplier of the adding-up constraint in the primal program (2). Then u_x solves the dual program

$$G_x\left(\alpha_{x\cdot} + t_{x\cdot}\right) = \min_{u_x \geq 0} u_x : \; u_x \geq \alpha_{xy} + t_{xy} \;\; \forall y \in \mathcal{Y} \tag{3}$$

While this program is trivial here, it will become more interesting when we introduce unobserved heterogeneity. Note that both (2) and (3) incorporate our normalization that singles get zero utility.

Proceeding similarly with women, we have the equivalent formulations

$$\begin{aligned} H_y\left(\gamma_{\cdot y} - t_{\cdot y}\right) &:= \max_{x \in \mathcal{X}} (\gamma_{xy} - t_{xy}) && (4) \\ &= \max_{\mu_{\cdot|y} \geq 0} \sum_{x \in \mathcal{X}} \mu_{x|y}\left(\gamma_{xy} - t_{xy}\right) : \sum_{x \in \mathcal{X}} \mu_{y|x} + \mu_{0|y} = 1 && (5) \\ &= \min_{v_y \geq 0} v_y : \; v_y \geq \gamma_{xy} - t_{xy} \;\; \forall x \in \mathcal{X}. && (6) \end{aligned}$$

with solutions $\mu^W_{\cdot|y}$ and v_y. Now since G_x and H_y are convex functions, they have subgradients ∂G_x and ∂H_y; and by the envelope theorem, $\mu^M_{\cdot|x} \in \partial G_x\left(\alpha_{x\cdot} + t_{x\cdot}\right)$ and $\mu^W_{\cdot|y} \in \partial H_y\left(\gamma_{\cdot y} - t_{\cdot y}\right)$. Introduce G and H as the sum of the value functions of men and women, respectively, that is

$$G(\alpha + t) := \sum_{x \in \mathcal{X}} n_x G_x(\alpha_{x\cdot} + t_{x\cdot}) \text{ and } H\left(\gamma - t\right) := \sum_{y \in \mathcal{Y}} m_y H_y(\gamma_{\cdot y} - t_{\cdot y}), \tag{7}$$

we have, recalling that $\mu^M_{xy} = n_x \mu^M_{y|x}$ and defining $\mu^W_{xy} := m_y \mu^W_{x|y}$ that

$$\mu^M \in \partial G\left(\alpha + t\right) \text{ and } \mu^W \in \partial H\left(\gamma - t\right).$$

1.3 Equilibrium

In equilibrium, the n_x men of type x seek to match with $\mu^M_{xy} = n_x \mu^M_{y|x}$ women of type y, while the m_y women of type y seek $\mu^W_{xy} = m_y \mu^W_{x|y}$ men of type x. Therefore the transfers t must take values such that

$$\mu^M_{xy} = \mu^W_{xy} = \mu_{xy} \tag{8}$$

for all pairs of types (x, y). Let us now sum the welfare of all participants in the market to compute the total joint surplus[3]:

$$\mathcal{W}(t) := G(\alpha+t)+H(\gamma-t) = \sum_{x\in\mathcal{X}} n_x \max_{y\in\mathcal{Y}} \{\alpha_{xy} + t_{xy}, 0\}+\sum_{y\in\mathcal{Y}} m_y \max_{x\in\mathcal{X}} \{\gamma_{xy} - t_{xy}, 0\} \tag{9}$$

which has a linear programming formulation as the sum of the dual problems (3) and (6) above

$$\begin{aligned} \mathcal{W}(t) = \min_{\substack{u_x\geq 0\\ v_y\geq 0}} \quad & \left(\sum_{x\in\mathcal{X}} n_x u_x + \sum_{y\in\mathcal{Y}} m_y v_y\right) \\ s.t. \quad & u_x \geq \alpha_{xy} + t_{xy} \;\; \forall y \\ & v_y \geq \gamma_{xy} - t_{xy} \;\; \forall x. \end{aligned}$$

The function $\mathcal{W}(t)$ is obviously convex in t, with a subdifferential $\partial G - \partial H$ wich is the set of vectors $\left(\mu_{xy}^M - \mu_{xy}^W\right)_{xy}$, where $\mu^M \in \partial G(\alpha + t)$ and $\mu^W \in \partial H(\gamma - t)$. Hence, in view of relation (8), at equilibrium the vector $0 = \mu^M - \mu^W$ should be in the subdifferential of $\mathcal{W}$. Therefore the equilibrium transfers $t = (t_{xy})$ must minimize $\mathcal{W}(t)$, and solve

$$\begin{aligned} \min_{(t_{xy})} \mathcal{W}(t) = \min_{\substack{u_x\geq 0\\ v_y\geq 0}} \quad & \left(\sum_{x\in\mathcal{X}} n_x u_x + \sum_{y\in\mathcal{Y}} m_y v_y\right) \\ s.t. \quad & u_x + v_y \geq \Phi_{xy} \end{aligned} \tag{10}$$

where the matrix $\Phi \equiv \alpha_{xy} + \gamma_{xy}$ is the *joint surplus* of a match between types x and y. As this formulation makes clear, the solutions u_x and v_y only depend on Φ and on the *margins* n and m. Once u_x and v_y are known, the transfers t_{xy} can be reconstructed as any solution to the set of inequalities

$$\gamma_{xy} - v_y \leq t_{xy} \leq u_x - \alpha_{xy} \;\; \forall(x, y).$$

By duality, the problem above can be interpreted as dual to the *optimal assignment problem*, and we have

$$\begin{aligned} \min_{(t_{xy})} \mathcal{W}(t) = \max_{\mu\geq 0} & \sum_{\substack{x\in\mathcal{X}\\ yin\mathcal{Y}}} \mu_{xy}(\alpha_{xy} + \gamma_{xy}) \\ & \text{s.t.} \sum_{y\in\mathcal{Y}} \mu_{xy} \leq n_x \;\; \forall x \in \mathcal{X} \text{ and } \sum_{x\in\mathcal{X}} \mu_{xy} \leq m_y \;\; \forall y \in \mathcal{Y}. \end{aligned}$$

which provides the equilibirum matching $\left(\mu_{xy}\right)$ in (8).

[3]Note that $\mathcal{W}$ also depends on α, γ, n, and m. We deleted them from the notation for readability.

1.4 Identification

Now suppose that the analyst observes the margin vectors (n_x) and (m_y) and the equilibrium matching patterns (μ_{xy}). Let us denote G^* the Legendre-Fenchel transform of the convex function G:

$$G^*(\mu) = \sup_{a \in \mathbb{R}^{\mathcal{Y} \times \mathcal{Y}}} \left(\sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} a_{xy} - G(a) \right).$$

It is another convex function; and by the theory of *convex duality* we know that since

$$\mu^M \in \partial G(\alpha + t),$$

we also have

$$\alpha + t \in \partial G^*(\mu^M). \tag{11}$$

Since the function G is easy to compute, so is the function G^*. Observing the matching patterns and the margin vectors therefore identifies all values of $(\alpha_{xy} + t_{xy})$ and $(\gamma_{xy} - t_{xy})$. By simple addition, it gives all values Φ_{xy} of the joint surplus. If the analyst observes the transfers, it also identifies α and γ.

2 Matching with Unobserved Heterogeneity

A proper econometric setting requires that we allow for unobserved heterogeneity, and that we spell out our assumptions on its distribution. Most crucially, the analyst cannot observe all the determinants of the joint surplus $\tilde{\Phi}_{mw}$ generated by a hypothetical match between a man m and a woman w. A priori, it could depend on interactions between types, between types and unobserved characteristics, and between the unobserved characteristics of both partners.

2.1 Separability

Much of the literature has settled on excluding interactions between unobserved characteristics, and this is the path we take here. We impose:

Assumption 1 (Separability)**.** *The joint surplus generated by a match between man m with type x and woman w with type y is*

$$\tilde{\Phi}_{mw} = \Phi_{xy} + \varepsilon_{my} + \eta_{wx}. \tag{12}$$

The utility of man m and woman w if unmatched are ε_{m0} and η_{w0} respectively.

In the language of analysis of variance models, the separability assumption rules out two-way interactions between unobserved characteristics, conditional on observed types. While this is restrictive, it still allows for rich patterns of matching in equilibrium. For instance, all women may like educated men, but those women who value education more are more likely (everything equal) to marry a more educated man, provided that they in turn have observed or unobserved characteristics that more educated men value more.

2.2 Equilibrium

We continue to assume that the econometrician only has data on the numbers of potential partners of each type (the margins n and m) and on "who matches with whom": the number of matches μ_{xy} between men of type x and women of type w. Convex duality will remain the key to our approach to identification. First, separability will allow us to extend the definitions of G_x and G_x^*. We start by rewriting the dual characterization of equilibrium in (10) as

$$\begin{aligned} \min_{\substack{u_m \geq \varepsilon_{m0} \\ v_w \geq \eta_{w0}}} \quad & \left(\sum_m u_m + \sum_w v_w \right) \\ s.t. \quad & u_m + v_w \geq \tilde{\Phi}_{mw} \;\; \forall m, w. \end{aligned} \tag{13}$$

Given Assumption 1, the constraint in (13) can be rewritten as

$$(u_m - \varepsilon_{my}) + (v_w - \eta_{wx}) \geq \Phi_{x_m y_w} \;\; \forall m, w. \tag{14}$$

Now define $U_{xy} = \min_{m: x_m = x} \{u_m - \varepsilon_{my}\}$ and $V_{xy} = \min_{w: y_w = y} \{v_w - \eta_{wx}\}$. The constraint becomes

$$U_{xy} + V_{xy} \geq \Phi_{xy} \;\; \forall x, y.$$

Moreover, by definition[4] $u_m = \max_{y \in \mathcal{Y}_0} (U_{x_m y} + \varepsilon_{my})$ and $v_w = \max_{x \in \mathcal{X}_0} (V_{x y_w} + \eta_{wx})$, so that we can rewrite the dual program as

$$\begin{aligned} \min_{U,V} \quad & \left(\sum_m \max_{y \in \mathcal{Y}_0} (U_{x_m y} + \varepsilon_{my}) + \sum_w \max_{x \in \mathcal{X}_0} (V_{x y_w} + \eta_{wx}) \right) \\ s.t. \quad & U_{xy} + V_{xy} \geq \Phi_{xy} \;\; \forall x, y. \end{aligned}$$

The Lagrange multiplier associated with the constraint $U_{xy} + V_{xy} \geq \Phi_{xy}$ is $\mu_{xy} \geq 0$, the number of matches between types x and y. Often no "matching cell" is empty in the data, so that $\mu_{xy} > 0$ for all (x, y). Then we can replace V_{xy} with $(\Phi_{xy} - U_{xy})$ to obtain a simple unconstrained program:

$$\min_U \left(\sum_m \max_{y \in \mathcal{Y}_0} (U_{x_m y} + \varepsilon_{my}) + \sum_w \max_{x \in \mathcal{X}_0} (\Phi_{x y_w} - U_{x y_w} + \eta_{wx}) \right).$$

We just reduced the dimensionality of the problem from the number of individuals in the market to the product of the numbers of their observed types. Since the latter is typically orders of magnitude smaller than the former, this is a drastic simplification. Assumption 1 was the key ingredient: without it, we would have a term ξ_{mw} interacting the unobservables in the joint surplus $\tilde{\Phi}_{mw}$ and (14) would lose its nice separable structure.

[4] Setting $U_{x0} = V_{0y} = 0$.

2.3 Identification

Let us now assume that we observe a market characterized by a matrix (Φ_{xy}) and distributions of the unobserved terms ε and η. For any man m of type x, the random vector $\varepsilon_m = (\varepsilon_{my})_{y \in \mathcal{Y}_0}$ is distributed according to $\mathbb{P}_x$. Similarly, for any woman w of type y, the random vector $\eta_w = (\eta_{wx})_{x \in \mathcal{X}_0}$ is distributed according to $\mathbb{Q}_y$. This allows us to define the convex functions G_x and G in a way that naturally extends expressions (3) and (7) by

$$G_x(U_{x\cdot}) := E_{\mathbb{P}_x} \max_{y \in \mathcal{Y}_0} (U_{xy} + \varepsilon_{my}) \text{ and } G(U) = \sum_{x \in \mathcal{X}} n_x G_x(U_{x\cdot})$$

Suppose that the market is large, so that averages can be approximated by expectations. Defining $H_y(V_{\cdot y})$ and $H(V)$ as with $G_x(U_{x\cdot})$ and $G(U)$, we see that the equilibrium U minimizes $G(U) + H(V)$ subject to constraints $U_{xy} + V_{xy} = \Phi_{xy}$. Now recall that μ_{xy} was the multiplier of the latter constraint. Simple algebra shows that

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

Taking Legendre–Fenchel transforms as in (11) gives

$$U_{xy} = \frac{\partial G^*}{\partial \mu_{xy}}(\mu)$$
$$V_{xy} = \frac{\partial H^*}{\partial \mu_{xy}}(\mu).$$

Since U and V must add to Φ, we obtain a system of $|\mathcal{X}| \times |\mathcal{Y}|$ equations

$$\Phi_{xy} = \frac{\partial G^*}{\partial \mu_{xy}}(\mu) + \frac{\partial H^*}{\partial \mu_{xy}}(\mu). \tag{15}$$

This identifies the Φ matrix in the joint surplus as a function of the observed matching patterns (μ_{xy}) and the shape of the functions G^* and H^*. The latter in turn depend on the distributions $\mathbb{P}_x$ and $\mathbb{Q}_y$.

2.4 The Total Joint Surplus

Just as in Section (1.1), the equilibrium matching maximizes the total joint surplus. The corresponding primal program is

$$\mathcal{W}(\Phi, n, m) = \max_{\mu \geq 0} \left(\sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy} - \mathcal{E}(\mu; n, m) \right) \tag{16}$$

where

$$\mathcal{E}(\mu; n, m) = G^*(\mu; n) + H^*(\mu; m)$$

is the *generalized entropy* of the matching μ. In this formula, G^* and H^* are the Legendre-Fenchel transforms of G and H. It is easy to check that the first-order conditions in (16) (which is globally concave) coincide with the identification formula (15).

The two parts of the objective function in (16) have a natural interpretation. The sum $\sum_{x,y} \mu_{xy} \Phi_{xy}$ reflects the value of matching on observed types only. The generalized entropy term $-\mathcal{E}(\mu; n, m)$ is the sum of the values that are generated by matching unobserved heterogeneities with observed types: e.g. men of type x with a high value of ε_{my} being more likely to match with women of type y.

2.5 Extending the Generalized Entropy

We skipped over an important technical issue: the Legendre-Fenchel transform of G_x is equal to $+\infty$ unless $\sum_{y \in \mathcal{Y}} \mu_{xy} = N_x(\mu) - \mu_{x0} \leq n_x$. Therefore the objective function in (16) is minus infinity when any of these scarcity constraints is violated.

There are two approaches to making the problem well-behaved. We can simply add the constraints to the program. As it turns out, extending the generalized entropy beyond its domain is sometimes a much better approach. To do this, we replace n_x with $N_x(\mu)$ in the definition of the generalized entropy, and we add terms that make the function finite for all μ while preserving its convexity:

$$E(\mu) = \mathcal{E}(\mu; N(\mu), M(\mu)) + f(N(\mu)) + g(M(\mu))$$

where f and g are strictly convex functions of (respectively) the vectors $N(\mu) = (N_x(\mu))_{x \in \mathcal{X}}$ and $M(\mu) = (M_y(\mu))_{y \in \mathcal{Y}}$, which are themselves linear functions of μ.

Any pair of functions f and g that satisfies these conditions will make the function E well-defined and strictly convex over all of $\mathbb{R}^{|\mathcal{X}|\,|\mathcal{Y}|}$; as we will see in Section 2.7, clever choices may lead to a further drastic simplification in the estimation procedure.

Armed with the extended generalized entropy E, we can replace $\mathcal{E}(\mu; n, m)$ with $E(\mu) - f(n) - g(m)$ and rewrite (16) as

$$\begin{aligned} W(\Phi; n, m) = \max_{\mu \geq 0} \quad & \left(\sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy} - E(\mu) + f(n) + g(m) \right) \\ s.t. \quad & N(\mu) = n \\ & M(\mu) = m. \end{aligned}$$

As a convex program, it has a dual formulation that can be written in terms of the Legendre-Fenchel transform E^* of E:

$$E^*\left((z_{xy}), (z_{x0}), (z_{0y})\right) = \max_{\mu \geq 0} \left(\sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} z_{xy} \mu_{xy} + \sum_{x \in \mathcal{X}} z_{x0} \mu_{x0} + \sum_{y \in \mathcal{Y}} z_{0y} \mu_{0y} - E(\mu) \right).$$

Letting (u_x) and (v_y) denote the multipliers of the scarcity constraints and $\Phi - u - v = (\Phi_{xy} - u_x - v_y)_{x,y}$, simple calculations show that the dual is:

$$W(\Phi; n, m) = \min_{u,v \geq 0} \left(\langle n, u \rangle + \langle m, v \rangle + E^* (\Phi - u - v, -u, -v) \right) + f(n) + g(m)$$

Taking first-order conditions in this C^1, strictly convex program gives

$$\begin{cases} n_x = \frac{\partial E^*}{\partial z_{xy}} + \frac{\partial E^*}{\partial z_{x0}} \\ m_y = \frac{\partial E^*}{\partial z_{xy}} + \frac{\partial E^*}{\partial z_{0y}} \end{cases}. \tag{17}$$

Now by the envelope theorem,

$$\begin{cases} \mu_{xy} = \frac{\partial E^*}{\partial z_{xy}} (\Phi - u - v, -u, -v) \\ \mu_{x0} = \frac{\partial E^*}{\partial z_{x0}} (\Phi - u - v, -u - v) \\ \mu_{0y} = \frac{\partial E^*}{\partial z_{0y}} (\Phi - u - v, -u, -v) \end{cases} \tag{18}$$

so that the conditions in (17) are simply $N(\mu) = n$ and $M(\mu) = m$.

These systems of equations will serve as the basis for a computationally attractive estimation procedure in Section 2.7.

2.6 The Multinomial Logit Model

Following a long tradition in discrete choice models, much of the literature has focused on the case when the distributions $\mathbb{P}_x$ and $\mathbb{Q}_y$ are standard type I extreme value (Gumbel). Under this distributional assumption, the G_x functions take a very simple and familiar form:

$$G_x(U_{x\cdot}) = \frac{\exp(U_{xy})}{1 + \sum_{t \in \mathcal{Y}} \exp(U_{xt})};$$

and the generalized entropy function $\mathcal{E}$ is just the usual entropy:

$$\mathcal{E}(\mu; n, m) = 2 \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \log \mu_{xy} + \sum_{x \in \mathcal{X}} \mu_{x0} \log \mu_{x0} + \sum_{y \in \mathcal{Y}} \mu_{0y} \log \mu_{0y}. \tag{19}$$

Equation (15) can be rewritten to yield the following *matching function*, which links the numbers of singles, the joint surplus, and the numbers of matches:

$$\mu_{xy} = \sqrt{\mu_{x0} \mu_{0y}} \exp\left(\frac{\Phi_{xy}}{2} \right). \tag{20}$$

To construct the extended entropy function E, we rely on the primitive of the logarithm:

$$f(N) = \sum_{x \in \mathcal{X}} (N_x \log N_x - N_x) \quad \text{and} \quad g(M) = \sum_{y \in \mathcal{Y}} (M_x \log M_y - M_y).$$

The reason for this a priori non-obvious choice of strictly convex functions is that many of the terms in the derivatives of the resulting extended entropy will cancel out. In fact, simple calculations give

$$E^{*}\left(z\right)=2\sum_{\substack{x\in\mathcal{X}\\y\in\mathcal{Y}}}\exp\left(\frac{z_{xy}}{2}\right)+\sum_{x\in\mathcal{X}}\exp\left(z_{x0}\right)+\sum_{y\in\mathcal{Y}}\exp\left(z_{0y}\right). \tag{21}$$

The equations in (18) become

$$\begin{cases}\mu_{xy}=\exp\left(\frac{\Phi_{xy}-u_x-v_y}{2}\right)\\ \mu_{x0}=\exp\left(-u_x\right)\\ \mu_{0y}=\exp\left(-v_y\right)\end{cases} \tag{22}$$

In the multinomial logit model, the distributions $\mathbb{P}_x$ and $\mathbb{Q}_y$ have no free parameter: the only unknown parameters in the model are those that determine the joint surplus matrix Φ. Using (20) (or eliminating u and v from (22)) gives *Choo and Siow's formula*:

$$\Phi_{xy}=\log\frac{\mu_{xy}^{2}}{\mu_{x0}\mu_{0y}} \tag{23}$$

Plugging in estimates $\hat{\mu}$ of the matching patterns in this formula gives a closed-form estimator $\hat{\Phi}$ of the joint surplus matrix. On the other hand, determining the equilibrium matching patterns μ for given primitive parameters Φ, n, m is more involved; and it is necessary in order to evaluate counterfactuals that modify these primitives of the model. We will show how to do it in Section 3.2 below. In addition, the econometrician may want to assume that the joint surplus matrix Φ belongs in a parametric family Φ^{λ}. While this could be done by finding the value of λ that minimize the distance between Φ^{λ} and the $\hat{\Phi}$ obtained from (23), there are better ways as we will now see.

2.7 Parametric estimation

Let us return to the general (separable) model. Assume that the analyst observes the numbers of individuals in each type $\hat{n}, \hat{m}$ and of the numbers of matches in each pair of types $\hat{\mu}$. Depending on the context, the econometrician may choose to allocate more parameters to the matrix Φ or to the distributions $\mathbb{P}_x$ and $\mathbb{Q}_y$. Let λ parameterize the joint surplus Φ, and β the distributions.

A natural choice of parameterization for Φ_{xy}^{λ} is the linear expansion

$$\Phi_{xy}^{\lambda}=\sum_{k=1}^{K}\lambda_k\phi_{xy}^{k}$$

where the basis functions ϕ_{xy}^{k} are given and the λ_k coefficients are to be estimated. The generalized entropy is a function of the unknown parameters β via the distributions $\mathbb{P}_x$ and $\mathbb{Q}_y$. Given any choice of strictly convex functions

f and g, the construction in Section 2.5 gives an extended entropy function E that also depends on β.

Given these modelling choices, we define the following function:

$$F\left(u, v, \lambda; \beta\right) \equiv \langle \hat{n}, u\rangle + \langle \hat{m}, v\rangle - \sum_{xy} \hat{\mu}_{xy}\Phi^{\lambda}_{xy} + E^{*}_{\beta}\left(\Phi^{\lambda} - u - v, -u, -v\right). \quad (24)$$

Note that since Φ^{λ} is linear in λ and the extended entropy is strictly convex, F is strictly convex in (u, v, λ) for any β. Its derivatives are

$$\begin{cases} \frac{\partial F}{\partial u_x} = n_x - \sum_{y\in\mathcal{Y}} \frac{\partial E^{*}_{\beta}}{\partial z_{xy}} - \frac{\partial E^{*}_{\beta}}{\partial z_{x0}} \\ \frac{\partial F}{\partial v_y} = m_y - \sum_{x\in\mathcal{X}} \frac{\partial E^{*}_{\beta}}{\partial z_{xy}} - \frac{\partial E^{*}_{\beta}}{\partial z_{0y}} \\ \frac{\partial F}{\partial \lambda_k} = \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \left(\frac{\partial E^{*}_{\beta}}{\partial z_{xy}} - \hat{\mu}_{xy}\right)\phi^{k}_{xy} \end{cases}$$

From the system (18) we know that the derivatives of E^{*}_{β} above are just the matching patterns $\mu(u, v, \lambda; \beta)$ implied by u and v when the joint surplus matrix is Φ^{λ} and the distributions $\mathbb{P}_x$ and $\mathbb{Q}_y$ correspond to the parameter value β. As a consequence, the derivatives of F can be written as

$$\frac{\partial F}{\partial u_x} = n_x - N_x(\mu(u, v, \lambda; \beta)) \quad (25)$$

$$\frac{\partial F}{\partial v_y} = m_y - M_y(\mu(u, v, \lambda; \beta)) \quad (26)$$

$$\frac{\partial F}{\partial \lambda_k} = \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \mu_{xy}(u, v, \lambda; \beta)\phi^{k}_{xy} - \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \hat{\mu}_{xy}\phi^{k}_{xy}. \quad (27)$$

For fixed β, minimizing F over (u, v, λ) ensures that the predicted matching has the same margins $\hat{n}$ and $\hat{m}$ as the observed matching $\hat{\mu}$; and that the predicted expectations of the basis functions ϕ^{k} coincide with their expectations in the data. Since F is strictly convex in in all of its $(|\mathcal{X}| + |\mathcal{Y}| + K)$ arguments (u, v, λ), minimizing it is a simple task. In addition to estimating the parameters of the joint surplus, it directly yields estimates of the expected utilities of each type.

For the multinomial logit model of Section 2.6, there are no β parameters and the function F becomes

$$\begin{aligned} F(u, v, \lambda) = {} & 2\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \exp\left(\frac{\Phi^{\lambda}_{xy} - u_x - v_y}{2}\right) + \sum_{x\in\mathcal{X}} \exp\left(-u_x\right) + \sum_{y\in\mathcal{Y}} \exp\left(-v_y\right) \\ & - \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \hat{\mu}_{xy}\left(\Phi^{\lambda}_{xy} - u_x - v_y\right) + \sum_{x\in\mathcal{X}} \hat{\mu}_{x0}u_x + \sum_{xx\in\mathcal{Y}} \hat{\mu}_{0y}v_y. \end{aligned} \quad (28)$$

In less constrained models, the parameters β of the distributions $\mathbb{P}_x$ and $\mathbb{Q}_y$ must also be estimated. This can be done by minimizing the distance between the observed matching patterns $\hat{\mu}$ and the μ^β that result from plugging the minimizers of $F(\cdot,\cdot,\cdot;\beta)$ into (18).

3 Computation

In many models (and certainly in the multinomial logit model with a linear joint surplus), minimizing the function F is the most appealing way to estimate the parameters. In all of this section, we assume that the joint surplus is indeed linear in the parameters λ; we consider any distributional parameter β as fixed and we omit it from the notation.

3.1 Gradient descent

The simplest approach to minimizing F is through gradient descent. Denoting $\theta = (u, v, \lambda)$, we start from a reasonable $\theta^{(0)}$ and we iterate:

$$\theta^{(t+1)} = \theta^{(t)} - \epsilon^{(t)} \nabla F \left(\theta^{(t)}\right)$$

where $\epsilon^{(t)} > 0$ is a small enough parameter. Given (25)–(27), we get

$$\begin{aligned}
u_x^{(t+1)} &= u_x^{(t)} + \epsilon^{(t)} \left(n_x - N_x(\mu^{(t)})\right) \\
v_y^{(t+1)} &= v_y^{(t)} + \epsilon^{(t)} \left(m_y - M_y(\mu^{(t)})\right) \\
\lambda_k^{(t+1)} &= \lambda_k^{(t)} + \epsilon^{(t)} \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \left(\mu_{xy}^{(t)} - \hat{\mu}_{xy}\right) \phi_{xy}^k,
\end{aligned}$$

denoting $\mu^{(t)}$ the result of plugging $(u^{(t)}, v^{(t)}, \lambda^{(t)})$ into (18).

This algorithm has a simple intuition: we adjust u_x in proportion of the excess of x types, v_y in proportion of the excess of y types, and λ in proportion of the mismatch between the k-th moment predicted by θ and the observed k-th moment.

In the multinomial logit model, $u_x^{(0)} = -\log(\hat{\mu}_{x0}/\hat{n}_x)$ and $v_y^{(0)} = -\log(\hat{\mu}_{0y}/\hat{m}_y)$ are excellent choices of initial values.

3.2 Coordinate descent

An even faster way to proceed is sometimes available. Fixing λ, let us us solve for the equilibrium matching:

$$\min_{\substack{u,v\\ \lambda\geq 0}} F\left(u, v, \lambda\right). \tag{29}$$

Coordinate descent consists of minimizing F iteratively with respect to its two argument vectors: with respect to u keeping v fixed, then with respect to v keeping u fixed at its new value, etc.

Let $v^{(t)}$ be the current value of v. Minimizing F with respect to u for $v = v^{(t)}$ yields a set of $|\mathcal{X}|$ equations in $|\mathcal{X}|$ unknowns: $u_x^{(t+1)}$ is the value of u_x that solves

$$\begin{aligned}\hat{n}_x = &\sum_{y\in\mathcal{Y}} \frac{\partial E^*}{\partial z_{xy}} \left(\Phi^\lambda - u - v^{(t)}, -u, -v^{(t)}\right) \\ &+ \frac{\partial E^*}{\partial z_{x0}} \left(\Phi^\lambda - u - v^{(t)}, -u, -v^{(t)}\right).\end{aligned}$$

These equations can in turned be solved coordinate by coordinate: we start with $x = 1$ and solve the $x = 1$ equation for $u_1^{(t+1)}$ fixing $(u_2, \ldots, u_{|\mathcal{X}|}) = (u_2^{(t)}, \ldots, u_{|\mathcal{X}|}^{(t)})$; then we solve the $x = 2$ equation for $u_2^{(t+1)}$ fixing $(u_1, u_3, \ldots, u_{|\mathcal{X}|}) = (u_1^{(t+1)}, u_3^{(t)} \ldots, u_{|\mathcal{X}|}^{(t)})$, etc. The convexity of the function E^* implies that the right-hand side of each equation is strictly decreasing in its scalar unknown, which makes it easy to solve.

The multinomial logit model constitutes an important special case in which these equations can be solved with elementary calculations, for any joint surplus matrix Φ. Define $S_{xy} := \exp(\Phi_{xy}/2)$; $a_x := \exp(-u_x)$; and $b_y := \exp(-v_y)$. It is easy to see that the system of equations that determines $u^{(t+1)}$ becomes

$$a_x^2 + a_x \sum_{y\in\mathcal{Y}} b_y^{(t)} S_{xy} = n_x \ \ \forall x \in \mathcal{X}.$$

These are $|\mathcal{X}|$ functionally independent quadratic equations, which can be solved in closed-form and in parallel. Once this is done, a similar system of independent quadratic equations gives $b^{(t+1)}$ from $a^{(t+1)}$. This procedure, introduced in Galichon and Salanié (2020), generalizes the Iterative Proportional Fitting Procedure (IPFP), also known as Sinkhorn's algorithm. It converges globally and very fast. Once the solutions (a_x) and (b_y) are obtained, the equilibrium matching patterns for this Φ are given by $\mu_{x0} = a_x^2$, $\mu_{0y} = b_y^2$ and $\mu_{xy} = a_x b_y S_{xy}$. For initial values, $a_x^{(0)} = \sqrt{\hat{\mu}_{x0}}$ and $b_y^{(0)} = \sqrt{\hat{\mu}_{0y}}$ are obviously good choices.

3.3 Hybrid Algorithms

In order to minimize the function F, one can also alternate between coordinate descent steps on u and v and gradient descent steps on λ, as suggested by Carlier et al. (2020). In the multinomial logit model, this would combine the updates

$$\begin{cases} \left(a_x^{(t+1)}\right)^2 + a_x^{(t+1)} \sum_{y\in\mathcal{Y}} b_y^{(t)} S_{xy}^{(t)} = n_x \\ \left(b_y^{(t+1)}\right)^2 + b_y^{(t+1)} \sum_{x\in\mathcal{X}} a_y^{(t+1)} S_{xy}^{(t)} = m_y \\ \lambda_k^{(t+1)} = \lambda_k^{(t)} + \epsilon^{(t)} \sum_{\substack{x\in\mathcal{X} \\ y\in\mathcal{Y}}} \left(a_x^{(t+1)} b_y^{(t+1)} S_{xy}^{(t)} - \hat{\mu}_{xy}\right) \phi_{xy}^k \end{cases}$$

where $S_{xy}^{(t)} = \exp(\sum_{k=1}^{K} \phi_{xy}^{k} \lambda_k^{(t)}/2)$.

A proof of convergence of hybrid algorithms is given in Carlier et al. (2020), in a more general setting that allows for model selection based on penalty functions.

4 Inference

4.1 The Sampling Level

In matching markets, the sample may be drawn from the population at the individual level or at the match level. Take the marriage market as an example. With individual sampling, each man or woman in the population would be a sampling unit. In fact, household-based sampling is more common in population surveys: when a household is sampled, data is collected on all of its members. Some of these households consist of a single man or woman, and others consist of a married couple. We assume here that sampling is at the household level.

4.2 The Asymptotic Distribution of the Estimator

Recall that $\hat{\mu}_{xy}$, $\hat{\mu}_{x0}$ and $\hat{\mu}_{0y}$ are the number of matches of type xy, $x0$ and $0y$, respectively in our sample. Denote

$$N_h = \sum_{x \in \mathcal{X}} \hat{\mu}_{x0} + \sum_{y \in \mathcal{Y}} \hat{\mu}_{0y} + \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \hat{\mu}_{xy}$$

the number of households in our sample, and let

$$\hat{\pi}_{xy} = \frac{\hat{\mu}_{xy}}{N_h}, \hat{\pi}_{x0} = \frac{\hat{\mu}_{x0}}{N_h} \text{ and } \hat{\pi}_{0y} = \frac{\hat{\mu}_{0y}}{N_h}$$

the empirical sample frequencies of matches of type xy, $x0$ and $0y$, respectively. Let π be the population analogs of $\hat{\pi}$. In the multinomial logit case, recall that our preferred method estimates $\theta = (\lambda, u, v)$ by

$$\min_{\theta} F(\theta, \hat{\pi})$$

where

$$\begin{aligned} F(\theta, \hat{\pi}) \quad = \quad & \sum_{x \in \mathcal{X}} \exp(-u_x) + \sum_{y \in \mathcal{Y}} \exp(-v_y) + 2 \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \exp\left(\frac{\Phi_{xy}^{\lambda} - u_x - v_y}{2}\right) \\ & - \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \hat{\pi}_{xy} \left(\Phi_{xy}^{\lambda} - u_x - v_y\right) + \sum_{x \in \mathcal{X}} \hat{\pi}_{x0} u_x + \sum_{y \in \mathcal{Y}} \hat{\pi}_{0y} v_y. \end{aligned}$$

The estimators of the matching probabilities have an asymptotic distribution

$$\hat{\pi}_h \sim \mathcal{N}\left(0, \frac{V_\pi}{N_h}\right) \text{ where } V_\pi = diag(\pi) - \pi\pi^\top.$$

where $h \in \mathcal{X} \times \mathcal{Y} \cup \mathcal{X} \times \{0\} \cup \{0\} \times \mathcal{Y}$. Totally differentiating the first order conditions $F_\theta(\hat{\theta}, \hat{\pi}) = 0$ gives

$$\theta \sim \mathcal{N}\left(0, \frac{V_\theta}{N_h}\right)$$

where

$$V_\theta = \left(F_{\theta\theta}\right)^{-1} F_{\theta\pi} V_\pi F_{\theta\pi}^\top \left(F_{\theta\theta}\right)^{-1}$$

and the F_{ab} represent the blocks of the Hessian of F at $(\hat{\theta}, \hat{\pi})$. Easy calculations show that $F_{\theta\theta}$ in turn decomposes into

$$\begin{pmatrix} F_{uu} = ... & F_{uv} = \left(\frac{\pi^\lambda_{xy}}{2}\right)_{xy} & F_{u\lambda} = -\frac{1}{2}\left(\sum_y \pi^\lambda_{xy}\phi^k_{xy}\right)_{xk} \\ . & F_{vv} = ... & F_{v\lambda} = -\frac{1}{2}\left(\sum_x \pi^\lambda_{xy}\phi^k_{xy}\right)_{yk} \\ . & . & F_{\lambda\lambda} = \frac{1}{2}\left(\sum^\lambda_{x,y} \hat{\pi}_{xy}\phi^k_{xy}\phi^l_{xy}\right)_{kl} \end{pmatrix}$$

where

$$F_{uu} = \operatorname{diag}\left(\left(\frac{1}{2}\sum_y \pi^\lambda_{xy} + \pi^\lambda_{x0}\right)_x\right) \text{ and } F_{vv} = \operatorname{diag}\left(\left(\frac{1}{2}\sum_x \pi^\lambda_{xy} + \pi^\lambda_{0y}\right)_y\right)$$

and

$$F_{\theta\pi} = \begin{pmatrix} \left(1^\top_{\mathcal{Y}} \otimes I_{\mathcal{X}}\right) & I_X & 0 \\ \left(I_{\mathcal{Y}} \otimes 1^\top_{\mathcal{X}}\right) & 0 & I_Y \\ \left(-\phi^k_{xy}\right)_{k,xy} & 0 & 0 \end{pmatrix}.$$

5 Other Implementation Issues

Let us now very briefly discuss three issues that often crop up in applications.

5.1 Continuous Types

While we modeled types as discrete-valued in this chapter, there are applications where this is not appropriate. Dupuy and Galichon (2014) showed how to incorporate continuous types in a separable model. The idea, following Dagsvik (2000), is to model the choice of possible partners as generated by the points of a specific Poisson process. This generates a matching function that is similar to the multinomial logit model of Section 2.6. One can also mix discrete- and continuous-valued types, as in Guadalupe et al. (2020). Dupuy and Galichon (2014) show that when the surplus function is bilinear on $\mathbb{R}^{d_x} \times \mathbb{R}^{d_y}$: $\Phi(x, y) = x^\top A y$, then π solves

$$\max_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_\pi\left[X^\top A Y\right] - \mathbb{E}_\pi\left[\ln \pi(X, Y)\right].$$

The maximization is over $\mathcal{M}(P,Q)$, the set of joint distributions of (X,Y) such that X has distribution P and Y has distribution Q. At the optimum, for every $x \in \mathbb{R}^{d_x}$ for every $y \in \mathbb{R}^{d_x}$ and for every k and l such that $1 \leq k \leq d_x$ and $1 \leq l \leq d_y$, one has

$$\frac{\partial^2 \ln \pi(x,y)}{\partial x_k \partial y_l} = A_{kl}.$$

In particular, when P and Q are Gaussian distributions, Bojilov and Galichon (2016) show that the optimal matching (X,Y) is a Gaussian vector whose distribution can be obtained in closed form. Suppose for instance that $d_x = d_y = 1$; $P = \mathcal{N}\left(0, \sigma_x^2\right)$; $Q = \mathcal{N}\left(0, \sigma_y^2\right)$; and $\Phi(x,y) = axy$, Then at the optimum $VX = \sigma_x^2$, $VY = \sigma_y^2$, and $\operatorname{corr}(X,Y) = \rho$ where ρ is related to a by

$$a\sigma_x\sigma_y = \frac{\rho}{1-\rho^2}.$$

5.2 Using Several Markets

We have focused on the case when the analyst has data on one market. If data on several markets is available; matches do not cross market boundaries; and some of the primitives of the model coincide across markets, then this can be used to relax the conditions necessary for identification.

As an example, Chiappori et al. (2017) pooled Census data on thirty cohorts in the US in order to study the changes in the marriage returns to education. To do this, they assumed that the supermodularity module of the function Φ changed at a constant rate over the period.

Fox et al. (2018) show how given enough markets, one can identify the distribution of the unobserved heterogeneity if it is constant across markets.

5.3 Using Additional Data

In applications to the labor market for instance, the analyst often has some information on transfers—wages in this case. This information can be used in estimating the underlying matching model. It is especially useful if it is available at the level of each individual match. Aggregate data on transfers has more limited value (Salanié, 2015).

6 Semiparametric Approaches

This chapter has emphasized maximum-likelihood estimation and moment-matching in separable models. Let us now describe two alternative approaches that do not require that the analyst fully specify the distribution of unobserved heterogeneity.

6.1 Exploiting Monotonicity

In most one-sided random utility models of discrete choice, the probability that a given alternative is chosen increases with its mean utility. Manski (1975) used this monotonicity property to construct a maximum score estimator of the parameters of such discrete choice models. Assume that alternative j has utility $U(x_{ij}, \theta_0) + u_{ij}$ for individual i. Let $J(i)$ be the choice of individual i and for any given θ, denote

$$R_i(\theta) \equiv \sum_{j \neq J(i)} \mathbb{1}\left(U(x_{i,J(i)}, \theta) > U(x_{ij}, \theta)\right)$$

the rank (from the bottom) of the chosen alternative $J(i)$ among the mean utilities. Choose any increasing function F. If (for simplicity) the u_{ij} are iid across i and j, maximizing the score function

$$\sum_i F\left(R_i(\theta)\right)$$

over θ yields a consistent estimator of θ_0. The underlying intuition is simply that the probability that j is chosen is an increasing function of the differences of mean utilities $U(x_{ij}, \theta) - U(x_{ik}, \theta)$ for all $k \neq j$.

It seems natural to ask whether a similar property also holds in two-sided matching with transferable utility: is there a sense in which (under appropriate assumptions) the probability of a match increases with the surplus it generates? This line of research was started by Fox (2010).

If transfers are observed, then each individual's choices is just a one-sided choice model and Manski (1975)'s result can be used essentially as is. Without data on transfers, the answer is not straightforward. In a two-sided model, the very choice of a single ranking is not self-evident. In so far as the optimal matching is partly driven by unobservables, it is generally not true that the optimal matching maximizes the joint total non-stochastic surplus for instance.

One can give a positive answer in one of the models we have already discussed: the multinomial logit specification of Choo and Siow (2006). The formula

$$\frac{\mu(x,y)^2}{\mu(x,\emptyset)\mu(\emptyset,y)} = \exp\left(\Phi(x,y)\right)$$

implies that for any (x, x', y, y'), the double log-odds ratio

$$2 \log \frac{\mu(x,y)\mu(x',y')}{\mu(x,y')\mu(x',y)}$$

equals the double difference

$$D_\Phi(x,x',y,y') \equiv \Phi(x,y) + \Phi(x',y') - \Phi(x',y) - \Phi(x,y').$$

This direct link between the observed matching patterns and the unknown surplus function justifies a maximum-score estimator

$$\max_\Phi \sum_{(x,x',y,y') \in C} \mathbb{1}\left(D_\Phi(x,x',y,y') > 0\right)$$

where C is a subset of the pairs that can be formed from the data. Bajari and Fox (2013) used this estimator to study the FCC spectrum auctions.

More generally, Graham (2011, 2014) proved that if the surplus is separable and the distribution of the unobservable heterogeneity vectors is independently and identically distributed, then the log-odds ratio and the double difference $D_\Phi(x, x', y, y')$ defined above have the same sign. While this is clearly a weaker result than in the multinomial logit model, it is enough to apply the same maximum-score estimator.

Fox (2018) extended Graham's result to exchangeable distributions on a many-to-many matching market. He applied the maximum score estimator to data on trades of car parts between suppliers and assemblers. His application shows one of the main advantages of the maximum-score method: it is easier to extend to more complex matching markets[5]. It also allows the analyst to select the tuples of trades in C to emphasize those that are more relevant in a given application. The price to pay is double. First, the maximum-score estimator maximizes a discontinuous function and has slow asymptotics[6]. Second, the underlying monotonicity property only holds for distributions of unobserved heterogeneity that exclude nested logit models and random coefficients for instance.

6.2 Dimension Reduction

As we have seen, the joint surplus function is a very high-dimensional object: it is a function of the observable characteristics of both partners and of the stochastic terms. In the original Becker (1973) model of the marriage market, the joint surplus had no stochastic element; and it only depended on one quantitative trait of the two partners:

$$\Phi_{mw} = \Phi(x_m, x_w)$$

for some scalar trait x. Slightly more generally, one could have $\Phi_{mw} = \Phi(I_m, J_w)$ with indices $I_m = I(x_m)$ and $J_w = J(x_w)$. Just as in Becker (1973), if Φ is supermodular then the optimal matching is positive assortative: men with a higher I_m match with women with a higher J_w.

Chiappori et al. (2012) build on this idea to propose a model that reduces the dimensionality of the parameter space when the observable characteristics x_m and y_w have continuous distributions. For empirical work we need to take into account unobserved heterogeneity, and to restrict its variations. First, we will assume a form of nonlinear separability: all unobserved heterogeneity reduces to an ε_m that only depends on the man, by an η_w that only depends on the woman. Second, we will restrict how the distribution of ε_m (resp. η_w) depends on the observed characteristics x_m (resp y_w).

More precisely, we assume that there exist two vectors of indices $I_m = I(x_m)$ and $J_w = J(x_w)$ such that

$$\Phi_{mw} = \Phi(I_m, J_w, \varepsilon_m, \eta_w).$$

[5]As always, in the absence of large complementarities.

[6]The maximum-score estimator converges at a cubic-root rate.

Moreover, we assume that ε_m is independent if x_m conditional on I_m, and that η_w is independent if y_w conditional on J_w; and that Φ is increasing in ε_m and in η_w.

For simplicity, we now assume that the indices are one-dimensional. The function I can be interpreted as an "attractiveness index" by which all women agree to rank men. Women may differ in the intensity of their preference for more "attractive" men; but they all agree—this is the crucial ingredient—in how they aggregate the various observable characteristics x_m into a one-dimensional index. The conditional independence property imposed on ε_m ensures that this reduction of the dimension of the problem survives the introduction of stochastic terms.

Under these conditions, Chiappori et al. (2012) show that at any optimal matching, all men with the same value of the index I_m are matched with women who share the same value of J_w. More precisely, the probability density function of the matches in (x_m, y_w) space is a function of $I(x_m)$ and $J(y_w)$:

$$\mu(x_m, y_w) \equiv \zeta\left(I(x_m), J(y_w)\right)$$

for some function ζ defined on $\mathbb{R}^2$.

References

Bajari, P., and Fox, J. 2013. Measuring the Efficiency of an FCC Spectrum Auction. *American Economic Journal: Microeconomics*, **5**, 100–146.

Becker, G. 1973. A theory of marriage, part I. *Journal of Political Economy*, **81**, 813–846.

Bojilov, R., and Galichon, A. 2016. Matching in Closed-Form: Equilibrium, identification, and comparative statics. *Economic Theory*, **61**, 587–=609.

Carlier, Guillaume, Dupuy, Arnaud, Galichon, Alfred, and Sun, Yifei. 2020. *SISTA: learning optimal transport costs under sparsity constraints.* arXiv 2009.08564.

Chiappori, P.-A., Oreffice, S., and Quintana-Domeque, C. 2012. Fatter Attraction: Anthropometric and Socioeconomic Characteristics in the Marriage Market. *Journal of Political Economy*, **120**, 659–695.

Chiappori, P.-A., Salanié, B., and Weiss, Y. 2017. Partner Choice, Investment in Children, and the Marital College Premium. *American Economic Review*, **107**, 2109–67.

Choo, E., and Siow, A. 2006. Who Marries Whom and Why. *Journal of Political Economy*, **114**, 175–201.

Dagsvik, J. 2000. Aggregation in Matching Markets. *International Economic Review*, **41**, 27–58.

Dupuy, A., and Galichon, A. 2014. Personality traits and the marriage market. *Journal of Political Economy*, **122**, 1271–1319.

Fox, J. 2010. Identification in Matching Games. *Quantitative Economics*, **1**, 203–254.

Fox, J. 2018. Estimating Matching Games with Transfers. *Quantitative Economics*, **8**, 1–38.

Fox, J., Yang, C., and Hsu, D. 2018. Unobserved Heterogeneity in Matching Games with an Appplication to Venture Capital. *Journal of Political Economy*, **126**, 1339–1373.

Galichon, A. 2016. *Optimal Transport Methods in Economics*. Princeton University Press.

Galichon, A., and Salanié, B. 2020. *Cupid's Invisible Hand: Social Surplus and Identification in Matching Models*. Columbia University mimeo.

Graham, B. 2011. Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers. In: Benhabib, J., Bisin, A., and Jackson, M. (eds), *Handbook of Social Economics*. Elsevier.

Graham, B. 2014. *Errata on "Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers"*. mimeo Berkeley.

Guadalupe, M., Rappoport, V., Salanié, B., and Thomas, C. 2020. *The Perfect Match*. Columbia University mimeo.

Manski, C. F. 1975. Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, **3**, 205—228.

Salanié, Bernard. 2015. *Identification in Separable Matching with Observed Transfers*. Columbia University mimeo.

7 Appendix: reminders on convex analysis

This brief appendix on convex analysis, contains only the bare minimum needed for this chapter. For an economic interpretation in terms of matching, see Galichon (2016), chapter 6.

In what follows, we consider a convex function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ which is not identically $+\infty$. If φ is differentiable at x, we denote its *gradient* at x as the vector of partial derivatives, that is $\nabla\varphi(x) = (\partial\varphi(x)/\partial x_1, \ldots, \partial\varphi(x)/\partial x_n)$. In that case, one has for all x and $\tilde{x}$ in $\mathbb{R}^n$

$$\varphi(\tilde{x}) \geq \varphi(x) + \nabla\varphi(x)^\top (\tilde{x} - x).$$

Note that if $\nabla\varphi(x)$ exists, then it is the only vector $y \in \mathbb{R}^n$ such that

$$\varphi(\tilde{x}) \geq \varphi(x) + y^\top(\tilde{x} - x) \quad \forall \tilde{x} \in \mathbb{R}^n, \tag{30}$$

indeed, setting $\tilde{x} = x + te_i$ where e_i is the ith vector of the canonical basis of $\mathbb{R}^n$, and letting $t \to 0^+$ yields $y_i \leq \partial\varphi(x)/\partial x_i$, while letting $t \to 0^-$ yields $y_i \geq \partial\varphi(x)/\partial x_i$. This motivates the definition of the *subdifferential* $\partial\varphi(x)$ of φ at x as the set of vectors $y \in \mathbb{R}^n$ such that relation (30) holds. Equivalently, $y \in \partial\varphi(x)$ holds if and only if

$$y^\top x - \varphi(x) \geq \max_{\tilde{x}} \left\{ y^\top \tilde{x} - \varphi(\tilde{x}) \right\}$$

that is, if and only if

$$y^\top x - \varphi(x) = \max_{\tilde{x}} \left\{ y^\top \tilde{x} - \varphi(\tilde{x}) \right\}.$$

The above development highlights a special role for the function φ^* appearing in the expression above

$$\varphi^*(y) = \max_{\tilde{x}} \left\{ y^\top \tilde{x} - \varphi(\tilde{x}) \right\}$$

which is called the *Legendre-Fenchel transform* of φ. By construction,

$$\varphi(x) + \varphi^*(y) \geq y^\top x.$$

This is called Fenchel's inequality; as we just saw, it is an equality if and only if $y \in \partial\varphi(x)$. In fact, the subdifferential can also be defined as

$$\partial\varphi(x) = \arg\max_{y} \left\{ y^\top x - \varphi^*(y) \right\}.$$

Finally, the double Legendre-Fenchel transform of a convex function φ (the transform of the transform) is simply φ itself. As a consequence, the subgradients of φ and φ^* are inverses of each other. In particular, if φ and φ^* are both differentiable then

$$(\nabla\varphi)^{-1} = \nabla\varphi^*.$$

To see this, remember that $y \in \partial\varphi(x)$ if and only if $\varphi(x) + \varphi^*(y) = y^\top x$; but since $\varphi^{**} = \varphi$, this is equivalent to $\varphi^{**}(x) + \varphi^*(y) = y^\top x$, and hence to $x \in \partial\varphi^*(y)$. As a result, the following statements are equivalent:

(i) $\varphi(x) + \varphi^*(y) = x^\top y$;

(ii) $y \in \partial\varphi(x)$;

(iii) $x \in \partial\varphi^*(y)$.