Structural Estimation of Matching Markets with Transferable Utility*

Alfred Galichon[†] Bernard Salanié[‡]

July 16, 2021

In matching models with transferable utility, the partners in a match agree to transact in exchange of a transfer of numéraire (utility or money) from one side of the match to the other. While transfers may be non zero-sum (if for instance there is diminishing marginal utility, frictions, or other costs) or constrained, we focus in this chapter on the simplest case of *perfectly transferable utility*, in which the transfers are unlimited and zero-sum: the transfer agreed to by one partner is fully appropriated by the other side[1]. 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.

With perfectly transferable utility, the main object of interest is the *joint surplus function*. It maps the characteristics of a man and a woman into the surplus utility created by their match, relative to the sum of the utilities they would achieve by staying single. Knowing the joint surplus function is informative about the preferences of the partners, and about their interaction within the match. It also opens the door to counterfactual analysis, for instance of the impact of policy changes.

We assume that the analyst observes a discrete set of characteristics for each individual: their education, their age, their income category, etc. Each combination of the values of these characteristics defines a *type*. In any real-world application, men and women of a given observed type will also vary in their preferences and more generally in their ability to create joint surplus in any match. We will assume that all market participants observe this additional variation, so that it contributes in determining the observed matching. On the other hand, by definition it constitutes *unobserved heterogeneity* for the analyst. The main challenge in this field is to recover the parameters of the joint surplus function without restricting too much this two-sided unobserved heterogeneity.

*We thank Nikhil Agarwal and Paulo Somaini for their comments and Gabriele Buontempo for superb research assistance.

†New York University and Sciences Po. Support from ERC grant EQUIPRICE No. 866274 is acknowledged.

‡Columbia University.

[1]See Chapter **??** in the present volume.

Matching with transferable utility solves a linear programming problem. In recent years it has been analyzed with the methods of *optimal transport*. Under an additional "*separability*" assumption, most functions of interest are convex; then *convex duality* gives a simple and transparent path to identificationidentification of the parameters of these models[2]. The empirical implementation is especially straightforward when the unobserved heterogeneity has a *logit* form and the joint surplus is linear in the parameters. Then the parameters can be estimated by minimizing a globally convex objective function.

Section 1 introduces separable matching models. Section 2 presents assumptions under which data on "who matches whom" (the *matching patterns*) identifies the parameters of the joint surplus function, and possibly also of the distributions of unobserved heterogeneity. We will also show how these parameters can be estimated (Section 3), and how to compute the stable matchings for given parameter values (Section 4).

Notation. We use bold letters for vectors and matrices. For any doubly-indexed variable $\boldsymbol{z} = (z_{ab})$, we use the notation $\boldsymbol{z}_{a\cdot}$ to denote the vector of values of z_{ab} when b varies; and we use a similar notation for $\boldsymbol{z}_{\cdot b}$.

1 Matching with unobserved heterogeneity

1.1 Population and preferences

We consider a population of men indexed by i and a population of women indexed by j. Each match must consist of one man and one woman; and individuals may remain single. If a man i and a woman j match, the assumption of *perfectly transferable utility* implies that their respective utilities can be written as

$$\alpha_{ij} + t_{ij}$$
$$\gamma_{ij} - t_{ij}$$

where t_{ij} is the (possibly negative) transfer from j to i[3]. Transfers can take all values on the real line, and are costless. We assume that each individual knows the equilibrium values of the transfers for all matches that (s)he may take part in, as well as his/her pre-transfer utility $\alpha_{i\cdot}$ or $\gamma_{\cdot j}$.

One key feature of markets with perfectly transferable utility is that matching patterns do not depend on $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ separately, but only on their sums, which we call the joint surplus[4].

[2] We collected the elements of convex analysis used in this chapter in Appendix A.

[3] If t_{ij} is negative, it should be interpreted as a transfer of $-t_{ij}$ from i to j. Also, $t_{i0} = t_{0j} = 0$

[4] Strictly speaking, it only is a "surplus" when all α_{i0} and γ_{0j} are zero. We follow common usage here.

Definition 1 (Joint Surplus)**.** *The joint surplus of a match is the sum of (pre- or post-transfers) utilities:*

$$\tilde{\Phi}_{ij} = (\alpha_{ij} + t_{ij}) + (\gamma_{ij} - t_{ij}) = \alpha_{ij} + \gamma_{ij}.$$

We extend the definition to singles with $\tilde{\Phi}_{i0} = \alpha_{i0}$ and $\tilde{\Phi}_{0j} = \gamma_{0j}$.

To see this, note that any change

$$(\alpha_{ij}, \gamma_{ij}) \to (\alpha_{ij} - \delta, \gamma_{ij} + \delta)$$

can be neutralized by adding δ to the transfer t_{ij}. This combined change leaves post-transfer utilities unchanged; therefore it does not affect the decisions of the market participants.

A *matching* is simply a set $\boldsymbol{d}$ of 0–1 variables (d_{ij}) such that $d_{ij} = 1$ if and only if i and j are matched, along with 0–1 variables d_{i0} (resp. d_{0j}) that equal 1 if and only if man i (resp. woman j) is unmatched (single). It is *feasible* if no partner is matched more than once:

for all i, $\sum_j d_{ij} + d_{i0} = 1$; and for all j, $\sum_i d_{ij} + d_{0j} = 1$.

1.2 Stability

Our notion of equilibrium is *stability*. Its definition in the context of models with perfectly transferable utility is as follows[5].

Definition 2 (Stability—primal definition)**.** *A feasible matching is stable if and only if*

- *no match has a partner who would rather be single*
- *no pair of currently unmatched partners would rather be matched.*

The first requirement translates into $\alpha_{ij} - \alpha_{i0} \leq t_{ij} \leq \gamma_{ij} - \gamma_{0j}$ for all matched (i, j), that is if $d_{ij} = 1$. The second one is easier to spell out if we define u_i (resp. v_j) to be the post-transfer utility of man i (resp. woman j) at the stable matching. Then we require that if $d_{ij} = 0$, we cannot find a value of the transfer t_{ij} that satisfies both $\alpha_{ij} + t_{ij} > u_i$ and $\gamma_{ij} - t_{ij} > v_j$. Obviously, this is equivalent to requiring that $\tilde{\Phi}_{ij} \leq u_i + v_j$. Note that if $d_{ij} = 1$, then this inequality is binding since the joint surplus must be the sum of the post-transfer utilities. Moreover, the first requirement can be rewritten as $u_i \geq \alpha_{i0}$ for all men and $v_j \geq \gamma_{0j}$, with equality if man i or woman j is single.

We summarize this in an equivalent definition of *stability*.

Definition 3 (Stability—dual definition)**.** *A feasible matching $\boldsymbol{d}$ is stable if and only if the post-transfer utilities u_i and v_j satisfy*

[5] It can be seen as a special case of the more general definition of stability in Chapter **??**. See Definition **??**.

- *for all* i*,* $u_i \geq \tilde{\Phi}_{i0}$*, with equality if* i *is unmatched; and for all* j*,* $v_j \geq \tilde{\Phi}_{0j}$*, with equality if* j *is unmatched*
- *for all* i *and* j*,* $u_i + v_j \geq \tilde{\Phi}_{ij}$*, with equality if* i *and* j *are matched.*

The conditions in Definition 3 are exactly the *Karush-Kuhn-Tucker optimality conditions* of the following maximization program:

$$\begin{aligned}
\max_{\boldsymbol{d} \geq 0} \quad & \sum_{i,j} d_{ij} \tilde{\Phi}_{ij} + \sum_i d_{i0} \tilde{\Phi}_{i0} + \sum_j d_{0j} \tilde{\Phi}_{0j} \\
s.t. \quad & \sum_j d_{ij} + d_{i0} = 1 \ \ \forall i && (1) \\
& \sum_i d_{ij} + d_{0j} = 1 \ \ \forall j && (2)
\end{aligned}$$

if u_i and v_j are the multipliers of the feasibility conditions. Thus the stable matchings maximize the total *joint surplus* under the feasibility constraints. Program 2 above is called the primal program. Since both the objective function and the constraints are linear, its *dual* has the same value. It minimizes the sum of the post-transfer utilities under the stability constraints

$$\begin{aligned}
\min_{(u_i),(v_j)} \quad & \sum_i u_i + \sum_j v_j \\
s.t. \quad & u_i \geq \tilde{\Phi}_{i0} \ \ \forall i \\
& v_j \geq \tilde{\Phi}_{0j} \ \ \forall j \\
& u_i + v_j \geq \tilde{\Phi}_{ij} \ \ \forall i, j, && (3)
\end{aligned}$$

and the multipliers of the constraints equal the d_{i0}, d_{0j}, d_{ij} of the associated stable matching.

From an economic point of view, the linearity of these programs implies that since the feasibility set is never empty (one can always leave all men and women unmatched), there exists a stable matching, it is generically unique, and there always exists a stable matching $\boldsymbol{d}$ whose elements are all integers (zero or one). This paints a very different picture from matching with non-transferable utility, described in Chapter **??** of the present volume (see also Section **??** of Chapter **??**).

1.3 Separability

A proper econometric setting requires that we distinguish carefully what the analyst can observe from unobserved heterogeneity, which only the market participants observe. Most crucially, the analyst cannot observe all the determinants of the pre-transfer utilities α_{ij} and γ_{ij} generated by a hypothetical match between a man i and a woman j. A priori, they could depend on interactions between characteristics the analyst observes, between these characteristics and unobserved heterogeneity, and between the unobserved heterogeneity of both partners.

We now define observed characteristics as *types* $x \in \mathcal{X}$ for men, and $y \in \mathcal{Y}$ for women. These *types* are observed by all market participants as well as the analyst. 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 continue to use 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 options that are available to respectively women and men.

Men and women of a given type also have other characteristics which are not observed by the analyst. A man i who has observed type x, or a woman j who has observed type y, may be a more or less appealing partner in any number of ways. In so far as these characteristics are payoff-relevant, they contribute to determining who matches whom. We will assume in this chapter that contrary to the analyst, all participants observe these additional characteristics. To the analyst, they constitute *unobserved heterogeneity.* It is important to note that this distinction is data-driven: richer data converts *unobserved heterogeneity* into types.

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

Assumption 4 (Separability)**.** *The joint surplus generated by a match between man i with type x and woman j with type y is*

$$\tilde{\Phi}_{ij} = \Phi_{xy} + \varepsilon_{iy} + \eta_{jx}. \tag{4}$$

The utility of man i and woman j if unmatched are ε_{i0} and η_{j0} 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 give a higher value to education 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.

Since the analyst can only observe types, we now redefine a matching as a collection $\boldsymbol{\mu}$ of non-negative numbers: μ_{xy} denotes the number of matches between men of type x and women of type y, which is determined in equilibrium and observed by the analyst. All men of type x, and all women of type y, must be single or matched. This generates the feasibility constraints:

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

In the following, we denote $x_i = x$ if man i is of type x, and $y_j = y$ if woman j is of type y.

1.4 Equilibrium

Convex duality will be the key to our approach to *identification*. We start by rewriting the dual characterization of the stable matching in (3) as

$$\begin{aligned} \min_{\substack{u_i \geq \varepsilon_{i0} \\ v_j \geq \eta_{j0}}} & \quad \left(\sum_i u_i + \sum_j v_j \right) \\ s.t. & \quad u_i + v_j \geq \tilde{\Phi}_{ij} \;\; \forall i, j. \end{aligned} \tag{5}$$

Given Assumption 4, the constraint in (5) can be rewritten as

$$(u_i - \varepsilon_{iy}) + (v_j - \eta_{jx}) \geq \Phi_{x_i y_j} \;\; \forall i, j. \tag{6}$$

Define $U_{xy} = \min_{i: x_i = x} \{u_i - \varepsilon_{iy}\}$ and $V_{xy} = \min_{j: y_j = y} \{v_j - \eta_{jx}\}$ for $x, y \neq 0$; and without loss of generality, set $U_{x0} = V_{0y} = 0$ for $x, y > 0$. The constraint becomes

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

Moreover, by definition $u_i = \max_{y \in \mathcal{Y}_0} (U_{x_i y} + \varepsilon_{iy})$ and $v_j = \max_{x \in \mathcal{X}_0} (V_{x y_j} + \eta_{jx})$, so that we can rewrite the *dual program* as

$$\begin{aligned} \min_{\boldsymbol{U}, \boldsymbol{V}} & \quad \left(\sum_i \max_{y \in \mathcal{Y}_0} (U_{x_i y} + \varepsilon_{iy}) + \sum_j \max_{x \in \mathcal{X}_0} (V_{x y_j} + \eta_{jx}) \right) \\ s.t. & \quad U_{xy} + V_{xy} \geq \Phi_{xy} \;\; \forall x, y. \end{aligned}$$

Inspection of the objective function shows that the inequality constraint $U_{xy} + V_{xy} \geq \Phi_{xy}$ can be replaced by an equality; indeed, if it were strict, one could weakly improve the objective function while satisfying the constraint. Since this implies that $U_{xy} + V_{xy} = \Phi_{xy}$, we can replace V_{xy} with $(\Phi_{xy} - U_{xy})$ to obtain a simple formula for the total *joint surplus*:

$$\mathcal{W} = \min_{\boldsymbol{U}} \left(\sum_i \max_{y \in \mathcal{Y}_0} (U_{x_i y} + \varepsilon_{iy}) + \sum_j \max_{x \in \mathcal{X}_0} (\Phi_{x y_j} - U_{x y_j} + \eta_{jx}) \right). \tag{7}$$

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 4 was the key ingredient: without it, we would have an unobserved term ξ_{ij} interacting the unobservables in the joint surplus $\tilde{\Phi}_{ij}$ and (6) would lose its nice separable structure.

Moreover, the nested min-max in equation (7) is not as complex as it seems. Consider the expression

$$G_x(U_{x\cdot}) := \frac{1}{n_x} \sum_{x_i = x} \max_{y \in \mathcal{Y}_0} (U_{xy} + \varepsilon_{iy}).$$

When the number of individuals n_x tends to infinity, G_x converges to the Emax operator, namely

$$G_x(U_{x\cdot}) := \mathbb{E}[\max_{y\in\mathcal{Y}_0}(U_{xy} + \varepsilon_{iy})].$$

We shall assume from now on that this *large market limit* is a good approximation[6].

Since the maximum is taken over a collection of linear functions of $\boldsymbol{U}_{x\cdot}$, its value is a convex function, and so is G_x. Defining $H_y(\boldsymbol{V}_{\cdot y})$ similarly, we obtain

$$\mathcal{W} = \min_{\boldsymbol{U}} \left(G(\boldsymbol{U}) + H(\boldsymbol{\Phi} - \boldsymbol{U})\right) \tag{8}$$

where

$$\begin{aligned} G(\boldsymbol{U}) &:= \sum_{x\in\mathcal{X}} n_x G_x(\boldsymbol{U}_{x\cdot}) \\ H(\boldsymbol{V}) &:= \sum_{y\in\mathcal{Y}} m_y H_y(\boldsymbol{V}_{\cdot y}). \end{aligned}$$

These functions play a special role in our analysis. Since G is convex, it has a subgradient everywhere, which is a singleton almost everywhere. It is easy to see that the derivative of $\max_{y\in\mathcal{Y}_0}(U_{xy} + \varepsilon_{iy})$ with respect to U_{xy} equals 1 if y achieves a strict maximum, and 0 if it is not a maximum. As a consequence, the subgradient of G_x with respect to U_{xy} is[7] the proportion of men of type x whose match is of type y. We denote this proportion $\mu^M_{y|x}$. Finally, we note that the subgradient of G with respect to U_{xy} is n_x times the subgradient of G_x, that is the number μ^M_{xy} . To conclude (and using similar definitions for H):

$$\begin{aligned} \boldsymbol{\mu}^M &= \partial G(\boldsymbol{U}) \\ \boldsymbol{\mu}^W &= \partial H(\boldsymbol{V}). \end{aligned}$$

In equilibrium we must have $\mu^M_{xy} = \mu^W_{xy}$ for all x, y. This should not come as a surprise as it translates the first-order conditions in (8):

$$\partial G(\boldsymbol{U}) \cap \partial H(\boldsymbol{\Phi} - \boldsymbol{U}) \neq \emptyset.$$

2 Identification

Now let us denote G^* the *Legendre-Fenchel transform* of the convex function G:

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

[6] See also Chapter **??** for matching in large markets.

[7] Neglecting the measure zero cases where the subgradient is not a singleton.

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

$$\boldsymbol{\mu}^M = \partial G(\boldsymbol{U}),$$

we also have $\boldsymbol{U} = G^*(\boldsymbol{\mu}^M)$, that is

$$U_{xy} = \frac{\partial G^*}{\partial \mu_{xy}}(\boldsymbol{\mu}^M). \tag{9}$$

Similarly,

$$V_{xy} = \frac{\partial H^*}{\partial \mu_{xy}}(\boldsymbol{\mu}^W). \tag{10}$$

2.1 Identifying the Joint Surplus

In equilibrium, $\boldsymbol{\mu}^M = \boldsymbol{\mu}^W := \boldsymbol{\mu}$ and $\boldsymbol{U} + \boldsymbol{V} = \boldsymbol{\Phi}$; therefore we obtain

$$\Phi_{xy} = \frac{\partial G^*}{\partial \mu_{xy}}(\boldsymbol{\mu}) + \frac{\partial H^*}{\partial \mu_{xy}}(\boldsymbol{\mu}). \tag{11}$$

Observing the matching patterns thus identifies all values of U_{xy}, V_{xy}, and Φ_{xy}, provided that we have enough information to evaluate the function G. Since the shape of the function G only depends on the distribution of the *unobserved heterogeneity* terms, this is the piece of information we need.

Assumption 5 (Distribution of the unobserved heterogeneity)**.** *For any man i of type x, the random vector $\boldsymbol{\varepsilon}_{i\cdot} = (\varepsilon_{iy})_{y \in \mathcal{Y}_0}$ is distributed according to $\mathbb{P}_x$.*

Similarly, for any woman j of type y, the random vector $\boldsymbol{\eta}_{j\cdot} = (\eta_{jx})_{x \in \mathcal{X}_0}$ is distributed according to $\mathbb{Q}_y$.

Note that (11) is a system of $|\mathcal{X}| \times |\mathcal{Y}|$ equations. To repeat, it identifies the $\boldsymbol{\Phi}$ matrix in the joint surplus as a function of the observed matching patterns ($\boldsymbol{\mu}$) and the shape of the functions G^* and H^*. The latter in turn only depend on the distributions $\mathbb{P}_x$ and $\mathbb{Q}_y$. It is important to stress that the joint surplus is uniquely identified given any choice of these distributions. Identifying the distributions themselves requires more restrictions and/or more data.

2.2 Generalized Entropy

We already know from Section 1.2 that the stable matching maximizes the total joint surplus. The corresponding *primal program* is

$$\mathcal{W}(\boldsymbol{\Phi}, \boldsymbol{n}, \boldsymbol{m}) = \max_{\boldsymbol{\mu} \geq 0} \left(\sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy} - \mathcal{E}(\boldsymbol{\mu}; \boldsymbol{n}, \boldsymbol{m}) \right) \tag{12}$$

where

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

is the *generalized entropy* of the matching $\boldsymbol{\mu}$. It is easy to check that the first-order conditions in (12) (which is globally concave) coincide with the *identification* formula (11).

The two parts of the objective function in (12) 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 $-\mathcal{E}(\boldsymbol{\mu};\boldsymbol{n},\boldsymbol{m})$ 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_{iy} being more likely to match with women of type y.

We skipped over an important technical issue: the Legendre-Fenchel transform of G_x is equal to $+\infty$ unless $\sum_{y\in\mathcal{Y}}\mu_{xy} = N_x(\boldsymbol{\mu}) - \mu_{x0} \leq n_x$. Therefore the objective function in (12) is minus infinity when any of these feasibility constraints is violated. There are two approaches for 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, as we will show in Section 3.

2.3 The 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:

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

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

$$\mathcal{E}\left(\boldsymbol{\mu};\boldsymbol{n},\boldsymbol{m}\right) = 2\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}}\mu_{xy}\log\mu_{xy} + \sum_{x\in\mathcal{X}}\mu_{x0}\log\mu_{x0} + \sum_{y\in\mathcal{Y}}\mu_{0y}\log\mu_{0y}. \tag{13}$$

Equation (11) can be rewritten to yield the following *matching function*, which links the numbers of singles, the joint surplus, and the numbers of matches:

$$\mu_{xy} = \sqrt{\mu_{x0}\mu_{0y}}\exp\left(\frac{\Phi_{xy}}{2}\right). \tag{14}$$

In the *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 (14) gives *Choo and Siow's formula*:

$$\Phi_{xy} = \log\frac{\mu_{xy}^2}{\mu_{x0}\mu_{0y}} \tag{15}$$

3 Estimation

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.

Recall that $\hat{\mu}_{xy}$, $\hat{\mu}_{x0}$ and $\hat{\mu}_{0y}$ are the number of matches of type (x, y), $(x, 0)$ and $(0, y)$, 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

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

the empirical sample frequencies of matches of type (x, y), $(x, 0)$ and $(0, y)$, respectively. Let $\boldsymbol{\pi}$ be the population analog of $\hat{\boldsymbol{\pi}}$. The estimators of the matching probabilities have an asymptotic distribution

$$\hat{\boldsymbol{\pi}} \sim \mathcal{N}\left(0, \frac{\boldsymbol{V}_{\boldsymbol{\pi}}}{N_h}\right). \tag{16}$$

We seek to estimate a *parametric model* of the matching market. This involves specifying functional form for the matrix $\boldsymbol{\Phi}$ and choosing families of distributions for the *unobserved heterogeneity* $\mathbb{P}_x$ and $\mathbb{Q}_y$. We denote $\boldsymbol{\lambda}$ the parameters of $\boldsymbol{\Phi}$, $\boldsymbol{\beta}$ the parameters of the distributions, and our aim is to estimate $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\beta})$. Depending on the context, the analyst may choose to allocate more parameters to the matrix Φ or to the distributions $\mathbb{P}_x$ and $\mathbb{Q}_y$. We assume that the model is well-specified in that the data was generated by a matching market with true parameters $\boldsymbol{\theta}_0$.

We will assume in this section that the analyst is able to compute the stable matching $\boldsymbol{\mu}^{\boldsymbol{\theta}}$ for any value of the parameters $\boldsymbol{\theta}$. We provide several ways to do so efficiently in Section 4.

3.1 The Maximum Likelihood Estimator

In this setting, the log-likelihood function of the sample is simply the sum over all households of the log-probabilities of the observed matches. Let us fix the value of the parameters of the model at $\boldsymbol{\theta}$. We denote $\boldsymbol{\mu}^{\boldsymbol{\theta}}$ the equilibrium matching patterns for these values of the parameters and the observed margins $\boldsymbol{n}$ and $\boldsymbol{m}$.

A household may consist of a match between a man of type x and a woman of type y, of a single man of type x, or of a single woman of type y. The corresponding probabilities are respectively $\mu^{\boldsymbol{\theta}}_{xy}/N^{\boldsymbol{\theta}}_h$, $\mu^{\boldsymbol{\theta}}_{x0}/N^{\boldsymbol{\theta}}_h$, and $\mu^{\boldsymbol{\theta}}_{0y}/N^{\boldsymbol{\theta}}_h$, where

$$N^{\boldsymbol{\theta}}_h := \sum_{x,y \in \mathcal{X} \times \mathcal{Y}} \mu^{\boldsymbol{\theta}}_{xy} + \sum_{x \in \mathcal{X}} \mu^{\boldsymbol{\theta}}_{x0} + \sum_{y \in \mathcal{Y}} \mu^{\boldsymbol{\theta}}_{0y}$$

is the number of households in the stable matching for $\boldsymbol{\theta}$, which in general differs from N_h. The log-likelihood becomes

$$\log L(\boldsymbol{\theta}) := \sum_{x,y \in \mathcal{X}\times\mathcal{Y}} \hat{\mu}_{xy} \log \frac{\mu^{\boldsymbol{\theta}}_{xy}}{N^{\boldsymbol{\theta}}_h} + \sum_{x\in\mathcal{X}} \hat{\mu}_{x0} \log \frac{\mu^{\boldsymbol{\theta}}_{x0}}{N^{\boldsymbol{\theta}}_h} + \sum_{y\in\mathcal{Y}} \hat{\mu}_{0y} \log \frac{\mu^{\boldsymbol{\theta}}_{0y}}{N^{\boldsymbol{\theta}}_h}.$$

Maximizing this expression gives a *maximum likelihood estimator* that has the usual asymptotic properties: it is consistent, asymptotically normal, and asymptotically efficient. The maximization process may not be easy, however. In particular, the function $\log L$ is unlikely to be globally concave, and it may have several local extrema. This may make other approaches more attractive.

3.2 The Moment Matching Estimator

A natural choice of parameterization for $\boldsymbol{\Phi}^{\boldsymbol{\lambda}}$ is the linear expansion

$$\Phi^{\boldsymbol{\lambda}}_{xy} = \sum_{k=1}^{K} \lambda_k \phi^k_{xy}$$

where the basis functions ϕ^k_{xy} are given and the λ_k coefficients are to be estimated.

The *moment matching estimator* uses the K equalities

$$\sum_{x,y} \mu^{\boldsymbol{\theta}}_{xy} \phi^k_{xy} = \sum_{x,y} \hat{\mu}_{xy} \phi^k_{xy}$$

as its estimating equations. Both sides of these equalities can be interpreted as expected values of the basis function ϕ^k; in this sense, the estimator matches the observed and simulated (first) moments of the basis functions. By construction, it can only identify K parameters. We assume from now on that the values of the parameters of the distribution are fixed at $\boldsymbol{\beta}$, and we seek to estimate $\boldsymbol{\lambda}$.

Applying the envelope theorem to equation (12) shows that the derivative of the total joint surplus with respect to Φ_{xy} is the value of μ_{xy} for the corresponding stable matching. Using the chain rule, we obtain

$$\frac{\partial \mathcal{W}^{\boldsymbol{\beta}}}{\partial \lambda_k}(\boldsymbol{\mu}^{\boldsymbol{\theta}}, \hat{\boldsymbol{n}}, \hat{\boldsymbol{m}}) = \sum_{x,y} \mu^{\boldsymbol{\theta}}_{xy} \phi^k_{xy};$$

this allows us to rewrite the moment matching estimating equations as the first order conditions of

$$\max_{\boldsymbol{\lambda}} \left(\sum_{x,y} \hat{\mu}_{xy} \Phi^{\boldsymbol{\lambda}}_{xy} - \mathcal{W}^{\boldsymbol{\beta}}(\boldsymbol{\mu}^{\boldsymbol{\theta}}, \hat{\boldsymbol{n}}, \hat{\boldsymbol{m}}) \right). \tag{17}$$

Note that the function $\mathcal{W}$ is convex in $\boldsymbol{\Phi}$. Since $\boldsymbol{\Phi}^{\boldsymbol{\lambda}}$ is linear in $\boldsymbol{\lambda}$, the objective function of (17) is globally convex. This is of course a very appealing property in a maximization problem.

We still have to evaluate $\mathcal{W}^{\boldsymbol{\beta}}(\boldsymbol{\mu}^{\boldsymbol{\theta}}, \hat{\boldsymbol{n}}, \hat{\boldsymbol{m}}) = \sum_{x,y} \mu^{\boldsymbol{\theta}}_{xy} \Phi^{\boldsymbol{\lambda}}_{xy} - \mathcal{E}^{\boldsymbol{\beta}}(\boldsymbol{\mu}^{\boldsymbol{\theta}}; \hat{\boldsymbol{n}}, \hat{\boldsymbol{m}})$. It is often possible to circumvent that step, however. To see this, remember that the *generalized entropy* is only defined when $\boldsymbol{N}(\boldsymbol{\mu}) = \hat{\boldsymbol{n}}$ and $\boldsymbol{M}(\boldsymbol{\mu}) = \hat{\boldsymbol{m}}$. Now take any real-valued functions f and g such that $f(0) = g(0) = 0$, and consider the *extended entropy* function

$$E^{\boldsymbol{\beta}}(\boldsymbol{\mu}; \hat{\boldsymbol{n}}, \hat{\boldsymbol{m}}) = \mathcal{E}^{\boldsymbol{\beta}}(\boldsymbol{\mu}; \boldsymbol{N}(\boldsymbol{\mu}), \boldsymbol{M}(\boldsymbol{\mu})) + f(\boldsymbol{N}(\boldsymbol{\mu}) - \hat{\boldsymbol{n}}) + g(\boldsymbol{M}(\boldsymbol{\mu}) - \hat{\boldsymbol{m}}).$$

By construction, this function is well-defined for any $\boldsymbol{\mu}$, and it coincides with $\mathcal{E}^{\boldsymbol{\beta}}$ when $\boldsymbol{N}(\boldsymbol{\mu}) = \hat{\boldsymbol{n}}$ and $\boldsymbol{M}(\boldsymbol{\mu}) = \hat{\boldsymbol{m}}$. Therefore we can rewrite (12) as

$$\mathcal{W}^{\boldsymbol{\beta}}(\boldsymbol{\Phi}, \boldsymbol{n}, \boldsymbol{m}) = \max_{\boldsymbol{\mu} \geq 0} \left(\sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy} - E^{\boldsymbol{\beta}}(\boldsymbol{\mu}; \boldsymbol{n}, \boldsymbol{m}) \right)$$
$$s.t\ \boldsymbol{N}(\boldsymbol{\mu}) = \hat{\boldsymbol{n}} \text{ and } \boldsymbol{M}(\boldsymbol{\mu}) = \hat{\boldsymbol{m}}.$$

If moreover we choose f and g to be convex functions, this new program is also convex. As such, it has a dual formulation that can be written in terms of the Legendre-Fenchel transform $(E^{\boldsymbol{\beta}})^*$ of $E^{\boldsymbol{\beta}}$. Simple calculations show that the dual is:

$$\mathcal{W}^{\boldsymbol{\beta}}(\boldsymbol{\Phi}, \hat{\boldsymbol{n}}, \hat{\boldsymbol{m}}) = \min_{\boldsymbol{u}, \boldsymbol{v} \geq 0} \left(\langle \hat{\boldsymbol{n}}, \boldsymbol{u} \rangle + \langle \hat{\boldsymbol{m}}, \boldsymbol{v} \rangle + (E^{\boldsymbol{\beta}})^* (\boldsymbol{\Phi} - \boldsymbol{u} - \boldsymbol{v}, -\boldsymbol{u}, -\boldsymbol{v}) \right)$$

where we denote $\boldsymbol{\Phi} - \boldsymbol{u} - \boldsymbol{v} = (\Phi_{xy} - u_x - v_y)_{x,y}$.

Returning to (17), the program that defines the moment matching estimator can now be rewritten as follows:

$$\max_{\boldsymbol{\lambda}, \boldsymbol{u} \geq 0, \boldsymbol{v} \geq 0} \left(\sum_{x,y} \hat{\mu}_{xy} \Phi^{\boldsymbol{\lambda}}_{xy} - \langle \hat{\boldsymbol{n}}, \boldsymbol{u} \rangle - \langle \hat{\boldsymbol{m}}, \boldsymbol{v} \rangle - (E^{\boldsymbol{\beta}})^* (\boldsymbol{\Phi} - \boldsymbol{u} - \boldsymbol{v}, -\boldsymbol{u}, -\boldsymbol{v}) \right). \tag{18}$$

This is still a globally convex program; and if we can choose f and g such that the extended entropy $(E^{\boldsymbol{\beta}})^*$ has a simple Legendre-Fenchel transform, it will serve as a computationally attractive estimation procedure. In addition to estimating the parameters $\boldsymbol{\lambda}$ of the *joint surplus*, it directly yields estimates of the expected utilities $\boldsymbol{u}$ and $\boldsymbol{v}$ of each type. Moreover, after estimation the matching patterns can be obtained by:

$$\begin{cases} \mu^{\boldsymbol{\theta}}_{xy} = \frac{\partial (E^{\boldsymbol{\beta}})^*}{\partial z_{xy}} (\boldsymbol{\Phi} - \boldsymbol{u} - \boldsymbol{v}, -\boldsymbol{u}, -\boldsymbol{v}) \\ \mu^{\boldsymbol{\theta}}_{x0} = \frac{\partial (E^{\boldsymbol{\beta}})^*}{\partial z_{x0}} (\boldsymbol{\Phi} - \boldsymbol{u} - \boldsymbol{v}, -\boldsymbol{u}, -\boldsymbol{v}) \\ \mu^{\boldsymbol{\theta}}_{0y} = \frac{\partial (E^{\boldsymbol{\beta}})^*}{\partial z_{0y}} (\boldsymbol{\Phi} - \boldsymbol{u} - \boldsymbol{v}, -\boldsymbol{u}, -\boldsymbol{v}) \end{cases} \tag{19}$$

The logit model of Section 2.3 provides an illustration of this approach.

3.3 Estimating the Logit Model

Plugging in estimates $\hat{\boldsymbol{\mu}}$ of the matching patterns in formula (15) gives a closed-form estimator $\hat{\boldsymbol{\Phi}}$ of the *joint surplus* matrix in the *logit model*. On the other hand, determining the equilibrium matching patterns $\boldsymbol{\mu}$ for given primitive parameters $\boldsymbol{\Phi}, \boldsymbol{n}, \boldsymbol{m}$ 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 4.1 below. In addition, the analyst may want to assume that the joint surplus matrix $\boldsymbol{\Phi}$ belongs in a parametric family $\boldsymbol{\Phi}^{\lambda}$. While this could be done by finding the value of $\boldsymbol{\lambda}$ that minimize the distance between $\boldsymbol{\Phi}^{\lambda}$ and the $\hat{\boldsymbol{\Phi}}$ obtained from (15), the approach sketched in Section 3.2 is more appealing.

To construct an *extended entropy* function E in the logit model, we rely on the primitive of the logarithm $\mathcal{L}(t) = t \log t - t$; we define $f(\boldsymbol{T}) = \sum_x \mathcal{L}(T_x)$, and similarly for g. They are clearly convex functions. 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 cancel out. In fact, simple calculations give

$$E^{*}(\boldsymbol{z}) = 2 \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \exp\left(\frac{z_{xy}}{2}\right) + \sum_{x \in \mathcal{X}} \exp(z_{x0}) + \sum_{y \in \mathcal{Y}} \exp(z_{0y}). \tag{20}$$

Substituting into (18), the moment matching estimator and associated utilities solve

$$\min_{\boldsymbol{\lambda}, \boldsymbol{u} \geq 0, \boldsymbol{v} \geq 0} F(\boldsymbol{\lambda}, \boldsymbol{u}, \boldsymbol{v})$$

where

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

This is the objective function of a *Poisson regression* with two-way fixed effects. Minimizing F is a very easy task; we give some specialized algorithms in Section 4, but problems of moderate size can also be treated using statistical packages handling generalized linear models. Denote $\boldsymbol{\alpha} = (\boldsymbol{\lambda}, \boldsymbol{u}, \boldsymbol{v})$ the set of arguments of F. The asymptotic distribution of the estimator of $\boldsymbol{\alpha}$ is given in Appendix B.

3.4 The Maximum-score Method

In most one-sided random utility models of discrete choice, the probability that a given alternative is chosen increases with its mean utility. Assume that alternative k has utility $U(x_{kl}, \theta_0) + u_{kl}$ for individual l. Let $K(l)$ be the choice of individual l and for any given θ, denote

$$R_l(\theta) \equiv \sum_{k \neq K(l)} \mathbb{1}\left(U(x_{l,K(l)},\theta) > U(x_{kl},\theta)\right)$$

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

$$\sum_l F\left(R_l(\theta)\right)$$

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

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?

If transfers are observed, then each individual's choices is just a one-sided choice model and the maximum score estimator 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 *logit* specification of Section 2.3. Formula (14) implies that for any (x,x',y,y'), the double log-odds ratio $2\log((\mu_{xy}\mu_{x'y'})/(\mu_{x,y'}\mu_{x'y}))$ equals the double difference

$$D_\Phi(x,x',y,y') \equiv \Phi_{xy} + \Phi_{x'y'} - \Phi_{x'y} - \Phi_{xy'}.$$

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

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

where C is a subset of the pairs that can be formed from the data.

More generally, one can prove the following result.

Theorem 6 (Comonotonicity of double-differences)**.** *Assume that the surplus is separable and that the distribution of the unobservable heterogeneity vectors is exchangeable. Then for all (x,y,x',y'), the log-odds ratio $D_\Phi(x,x',y,y')$ and the double difference $\log((\mu_{xy}\mu_{x'y'})/(\mu_{x,y'}\mu_{x'y}))$ have the same sign.*

While this is clearly a weaker result than in the *logit* model, it is enough to apply the same maximum-score estimator.

One of the main advantages of the maximum-score method is that it extends to more complex matching markets. 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 converges slowly[8]. Second, the underlying monotonicity property only holds for distributions of unobserved heterogeneity that exclude nested logit models and random coefficients for instance.

4 Computation

We now turn to the efficient evaluation of the stable matching and the associated utilities for given values of the parameters. In all of this section, we consider any distributional parameters $\boldsymbol{\beta}$ as fixed and we omit them from the notation.

4.1 Solving for equilibrium with coordinate descent

First consider the determination of the equilibrium matching patterns for a given matrix $\boldsymbol{\Phi}$. In several important models, this can be done by adapting formula (18). A slight modification of the arguments that lead to this formula shows that for given $\boldsymbol{\Phi}$, maximizing the following function yields the equilibrium utilities of all *types*:

$$\bar{F}(\boldsymbol{u}, \boldsymbol{v}) := \sum_{x,y} \hat{\mu}_{xy} \Phi_{xy} - \langle \hat{\boldsymbol{n}}, \boldsymbol{u} \rangle - \langle \hat{\boldsymbol{m}}, \boldsymbol{v} \rangle - E^* \left(\boldsymbol{\Phi} - \boldsymbol{u} - \boldsymbol{v}, -\boldsymbol{u}, -\boldsymbol{v} \right).$$

Coordinate descent consists of maximizing $\bar{F}$ iteratively with respect to the two argument vectors: with respect to $\boldsymbol{u}$ keeping $\boldsymbol{v}$ fixed, then with respect to $\boldsymbol{v}$ keeping $\boldsymbol{u}$ fixed at its new value, etc.

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

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

These equations can in turn 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.

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

The *logit* model constitutes an important special case in which these equations can be solved with elementary calculations, for any joint surplus matrix $\mathbf{\Phi}$. 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 $\boldsymbol{u}^{(t+1)}$ 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 $\boldsymbol{b}^{(t+1)}$ from $\boldsymbol{a}^{(t+1)}$. Note that $a_x^{(0)} = \sqrt{\hat{\mu}_{x0}}$ and $b_y^{(0)} = \sqrt{\hat{\mu}_{0y}}$ are obvious good choices for initial values.

This procedure generalizes the *Iterative Proportional Fitting Procedure (IPFP)*, also known as *Sinkhorn's algorithm*. It converges globally and very fast. Once the solutions $\boldsymbol{a}$ and $\boldsymbol{b}$ are obtained, the equilibrium matching patterns for this $\mathbf{\Phi}$ are given by $\mu_{x0} = a_x^2$, $\mu_{0y} = b_y^2$ and $\mu_{xy} = a_x b_y S_{xy}$.

4.2 Gradient descent

Suppose that the analyst has chosen to use (18) for estimation. The simplest approach to maximizing the objective function is through *gradient descent*. Denoting $\boldsymbol{\alpha} = (\boldsymbol{\lambda}, \boldsymbol{u}, \boldsymbol{v})$, we start from a reasonable[9] $\boldsymbol{\alpha}^{(0)}$ and we iterate:

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

where $\epsilon^{(t)} > 0$ is a small enough parameter. This gives

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

denoting $\boldsymbol{\mu}^{(t)}$ the result of plugging $(\boldsymbol{u}^{(t)}, \boldsymbol{v}^{(t)}, \boldsymbol{\lambda}^{(t)})$ into (19).

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 $\boldsymbol{\lambda}$ in proportion of the mismatch between the k-th moment predicted by $\boldsymbol{\alpha}$ and the observed k-th moment.

4.3 Hybrid Algorithms

The approaches in the previous two subsections can also be combined. Carlier et al. (forthcoming) suggest alternating between coordinate descent steps on $\boldsymbol{u}$

[9] In the 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.

and $\boldsymbol{v}$ and gradient descent steps on $\boldsymbol{\lambda}$. In the *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. (forthcoming), in a more general setting that allows for model selection based on penalty functions.

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. It is possible to incorporate continuous types in a separable model that feels very similar to the logit model of Section 2.3. The idea is to model the choice of possible partners as generated by the points of a specific *Poisson process*. An interesting special case has a bilinear *joint surplus function* $\Phi(x, y) = x^\top Ay$. It is easy to see that at the optimum, the Hessian of the logarithm of the matching patterns equals A everywhere: for all $x \in \mathbb{R}^{d_x}$ and $y \in \mathbb{R}^{d_y}$,

$$\frac{\partial^2 \ln \mu}{\partial x \partial y}(x, y) = \frac{A}{2}.$$

As a consequence, the model is overidentified and therefore testable. Among other things, it makes it possible to test for the rank of the matrix A. If it is some $r < \min(d_x, d_y)$, then one can identify the "salient" combination of types that generate the joint surplus.

If moreover the distribution P of x and the distribution Q of y are Gaussians, 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 Notes

Matching with perfectly transferable utility was introduced by Koopmans and Beckmann (1957) and its theoretical properties were elucidated by Shapley and Shubik (1972). Becker (1973, 1974) made it the cornerstone of his analysis of marriage. Sections 2 and 3 of this chapter are based on the approach developed in Galichon and Salanié (2020). The extension of the logit model to continuous types was proposed by Dupuy and Galichon (2014), following Dagsvik (2000). They applied it to study how the joint surplus from marriage depends on the Big Five psychological traits of the partners. Guadalupe et al. (2020) combine continuous and discrete types to model mergers between European firms. The results for the bilinear Gaussian models appear in Bojilov and Galichon (2016).

The maximum-score method for matching models was proposed by Fox (2010), taking inspiration from Manski (1975)'s classic paper on one-sided discrete choice models. Bajari and Fox (2013) used this estimator to study the FCC spectrum auctions. Graham (2011, 2014) proved Theorem 6 for independent and identically distributed variables and Fox (2018) extended it to exchangeable variables.

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.

Becker, G. 1974. A theory of marriage, part II. *Journal of Political Economy*, **82**, S11–S26.

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. forthcoming. SISTA: learning optimal transport costs under sparsity constraints. *Communications on Pure and Applied Mathematics*.

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.

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.

Koopmans, Tjalling C., and Beckmann, Martin. 1957. Assignment Problems and the Location of Economic Activities. *Econometrica*, **25**, 53–76.

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.

Shapley, L., and Shubik, M. 1972. The Assignment Game I: The Core. *International Journal of Game Theory*, **1**, 111–130.

Appendix A: reminders on convex analysis

We focus here on the results that our chapter relies on. For an economic interpretation in terms of matching, see Chapter 6 of Galichon (2016).

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

$$\varphi(\tilde{x}) \geq \varphi(x) + y^\top (\tilde{x} - x) \quad \forall \tilde{x} \in \mathbb{R}^n, \tag{21}$$

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

Appendix B: asymptotic distribution of the logit moment-matching estimator

In this appendix, we provide the explicit formulas for the asymptotic distribution of the estimator of the matching surplus in the logit model of Section 3.3. The asymptotic distribution of the estimator $\hat{\boldsymbol{\alpha}}$ is easy to derive by totally differentiating the first order conditions $F_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\pi}}) = 0$. This yields

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

where

$$V_{\boldsymbol{\alpha}} = (F_{\boldsymbol{\alpha}\boldsymbol{\alpha}})^{-1} F_{\boldsymbol{\alpha}\boldsymbol{\pi}} V_\pi F_{\boldsymbol{\alpha}\boldsymbol{\pi}}^\top (F_{\boldsymbol{\alpha}\boldsymbol{\alpha}})^{-1}.$$

In this formula, V_π is as in (16) and the F_{ab} represent the blocks of the Hessian of F at $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\pi}})$. Easy calculations show that $F_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$ in turn decomposes into

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

where

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

Moreover,

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

Once the estimates $\hat{\boldsymbol{\alpha}}$ are obtained, we can apply (19) to compute the estimated matching patterns:

$$\begin{cases} \mu_{xy}^{\hat{\alpha}} = \exp\left(\Phi_{xy}^{\hat{\lambda}} - \hat{u}_x - \hat{v}_y)/2\right) \\ \mu_{x0}^{\hat{\alpha}} = \exp\left(-\hat{u}_x\right) \\ \mu_{0y}^{\boldsymbol{\alpha}} = \exp\left(-\hat{v}_y\right). \end{cases}$$

Index