Cupid's Invisible Hand

Alfred Galichon (Sciences Po, Paris)

Joint w Bernard Salanié (Columbia University)

Aalto University, June 6, 2014

1 Introduction

Estimation of matching models with fully transferable utility.

Econometrics of the marriage market: a few socio-economic variables (education, age, income, race, wealth, sometimes biometric data...) and a lot of unobserved heterogeneity both in terms of characteristics and in taste.

One would like to understand how the market clears in order to examine sociological theories (endogamy, assortativeness, transmission of social capital...) or economic/public policy issues (eg. impact on divorce law on gender inequality).

Revealed preference problem: What do observed marriage patterns reveal about mutual preferences of partners and the production function of the household? What is the surplus of both partners at equilibrium? In other words, how to identify matching surpluses based on observed actual marriages?

Beyond marriage, the analysis also extends to other matching settings, such as e.g. the market of CEOs. What are the complementarities between firm and CEO's characteristics?

The framework is a matching model with (fully or partially) transferable utility and unobserved heterogeneity. The object of interest is the joint surplus generated by a match between two partner types.

- We observe: socio-economic, sometimes biometric data on observed characteristics of both partners in matches and singles. Sometimes, transfers between them (salaries in the case of CEOs).

- We would like to: estimate a parametric form of the household production function, in particular in order to test cross-assortativeness or complementarities between the observable characteristics of both partners.

In most cases, we will observe a single market. Sometimes, we observe several disconnected markets where participants may be assumed to have similar characteristics distributions.

Becker (JPE 1973–74): Marriage as a competitive matching market with **transferable utility**.

On top of these assumptions, Becker presented an application with one dimensional characteristics and no heterogeneities. These assunptions yield stark predictions: "Positive assortative matching" on a single-dimensional "ability index".

A few years ago, Choo and Siow (JPE 2006) (hereafter, CS) have incorporated logit-type heterogeneities in Becker's model and show that in this framework, the marital surplus can be nonparametrically identified. This started a rich literature on identification and estimation in matching models:

- Echenique, Lee, Shum and Yenmez (2011)
- Fox (2011)
- Galichon and Salanié (2011)
- Decker, Lieb, McCann, and Stephens (2011)
- Chiappori, Salanié and Weiss (2011)
- Chiappori, Oreffice and Quintana-Domeque (2011)
- Dupuy and Galichon (2014)

1.1 The Becker-Shapley-Shubik theory of marriage

Transferable utility: surplus of a pair can be split without restrictions between man and woman. Static matching, no frictions. Observable types are discrete. We recall the Becker-Shapley-Shubik setting first. Consider a population with n_x men of type x, and m_y women of type y. Introduce:

- α_{xy} utility of man x with woman y, 0 if single

- γ_{xy} utility of woman y with man x, 0 if single

Transferable utility: τ_{xy} utility transfer from x to y. Then

- $\alpha_{xy} - \tau_{xy}$ post-transfer utility of man x

- $\gamma_{xy} + \tau_{xy}$ post-transfer utility of woman y

Then the market clears in order to maximize

$$\max_{\mu \geq 0} \sum_{x,y} \mu_{xy} \Phi_{xy} : \sum_{y} \mu_{xy} \leq n_x,\ \sum_{x} \mu_{xy} \leq m_y$$

where

- $\Phi_{xy} = \alpha_{xy} + \gamma_{xy}$ is the total gains to marriage,

- μ_{xy} is the number of (x, y) pairs,

$\mu_{x0} = n_x - \sum_y \mu_{xy}$ is the number of single men of type x, and $\mu_{0y} = m_y - \sum_x \mu_{xy}$ is the number of single women of type y.

1.2 Choo and Siow's model

$|\mathcal{X}|$ groups of men of same observable characteristics, indexed by x; $|\mathcal{Y}|$ groups of women, indexed by y (education, race, income, religion...). Market participants observe everybody's full characteristics – analyst does not.

Choo and Siow: utility of a man m of group x who marries a woman of group y can be written:

$$\alpha_{xy} - \tau_{xy} + \varepsilon_{xym},$$

where $\tau_{xy} =$ utility transfer in equilibrium, and ε_{xym} is a standard type-I E.V. unobserved heterogeneity. If single, gets utility

$$0 + \varepsilon_{x0m},$$

(0 is a choice of normalization w.l.o.g.). Similarly, the utility of a woman w of group y who marries a man of group x can be written as

$$\gamma_{xy} + \tau_{xy} + \eta_{xyw},$$

and she gets utility

$$0 + \eta_{0yw}$$

is she is single.

Denote μ_{xy} the number of marriages between men of group x and women of group y; μ_{x0} the number of single men of group x; and μ_{0y} the number of single women of group y.

Problem. *The matching surpluses (α_{xy} and γ_{xy}) are not observed; only matching patterns μ_{xy} are observed. What are the restrictions on the surpluses?*

Choo and Siow proved that, in equilibrium, if there are very large numbers of men and women within each group,

$$\exp(\frac{\Phi_{xy}}{2}) = \frac{\mu_{xy}}{\sqrt{\mu_{x0}\mu_{0y}}},$$

where Φ_{xy} denotes the *total systematic net gains to marriage*:

$$\Phi_{xy} = \alpha_{xy} + \gamma_{xy}.$$

Therefore marriage patterns directly identify the gains to marriage Φ in such a model.

As we shall show, one can in fact extend significantly Choo and Siow's model to allow for a much larger class of unobservable heterogeneities and yet get tractable models.

1.3 The general framework

1.3.1 Introducing unobserved heterogeneity

To allow for unobserved heterogeneity, we index men by $i \in \mathcal{I}$ and women by $j \in \mathcal{J}$. A matching $\tilde{\mu}$ is now characterized by $\tilde{\mu}_{ij} = 1$ if man i and woman j are matched, 0 otherwise (with the usual convention that singles are matched with 0). We shall denote $\mathcal{I}_0 = \mathcal{I} \cup \{0\}$ and $\mathcal{J}_0 = \mathcal{J} \cup \{0\}$ the set of potential individual partners including singlehood. The feasibility constraints now write for every i and j,

$$\sum_{k \in \mathcal{J}} \tilde{\mu}_{ik} \leq 1 \text{ and } \sum_{k \in \mathcal{I}} \tilde{\mu}_{kj} \leq 1, \tag{1}$$

with equality for individuals who are married. The surplus of a match between individuals i and j is denoted $\tilde{\Phi}$ and an equilibrium matching achieves the maximum of the total surplus.

$$\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \tilde{\nu}_{ij} \tilde{\Phi}_{ij} + \sum_{i \in \mathcal{I}} \tilde{\nu}_{i0} \tilde{\Phi}_{i0} + \sum_{j \in \mathcal{J}} \tilde{\nu}_{0j} \tilde{\Phi}_{0j}$$

over all feasible matchings $\tilde{\nu}$.

The analyst does not observe all of the characteristics of the men and women, and she can only compute quantities that depend on the observed types of the partners in a match. Let x_i be the observable type of man i and y_j the observable type of woman j. The analyst cannot observe $\tilde{\mu}$. Rather, she observes

$$\hat{\mu}_{xy} = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \mathbf{1}\left(x_i = x, y_j = y\right) \tilde{\mu}_{ij}.$$

We assume that there are many individuals of a given observable type.

Assumption 1 (Large Market). *The number of individuals on the market*

$$N = \sum_{x \in \mathcal{X}} n_x + \sum_{y \in \mathcal{Y}} m_y$$

goes to infinity, while keeping the ratios (n_x/N) *and* (m_y/N) *constant.*

One now faces the problem of finding a surplus function Φ_{ij} that non trivially rationalizes $\hat{\mu}_{xy}$. Without further restrictions, any observed match between man i of observed type x_i and a woman j of observed type y_j can be attributed to interactions between man i's unobservable characteristics and woman j's unobserved characteristics. Hence, we have restored rationalizability at the expense of empirical content. To avoid this pitfall, we rule out interactions between unobserved characteristics of men and women with the following separability assumption.

Assumption 2 (Separability). *There exist deterministic Φ_{xy} and random ε_{iy} and η_{xj} such that the joint surplus from a match between a man i of observable type x and a woman j of observable type y is*

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

Assumption 2 means that conditional on the observable types x and y, the surplus function is additively separable in i and j and therefore there is no further interaction between the remaininig unobservable characteristics.

Hence, if two men i and i' have identical observable type x, and their respective partners j and j' have identical observable type y, then the total surplus generated by these two matches is unchanged if we shuffle partners.

In the sequel, we shall denote the vector $\eta_{\cdot j}$ of random terms in the woman's utility for each of the men's types and call Q_y its distribution if woman j is of type y. Similarly, we shall denote P_x the distribution of $\varepsilon_{i\cdot}$ when man i is of type x. We shall assume:

Assumption 3 (General distributions). *a) For any man i such that $x_i = x$, the ε_{iy} are drawn from a $(|\mathcal{Y}|+1)$-dimensional centered distribution $\mathbf{P}_x$;*

b) For any woman j such that $y_j = y$, the η_{xj} are drawn from an $(|\mathcal{X}|+1)$-dimensional centered distribution $\mathbf{Q}_y$;

c) The distributions $\mathbf{P}_x$ and $\mathbf{Q}_y$ all have full support.

In the sequel we shall show that there are vectors (U_{xy}, V_{xy}) such that $U_{xy} + V_{xy} \geq \Phi_{xy}$ and such that the payoff of a man i of type x matching with a woman of type y is

$$U_{xy} + \varepsilon_{iy}.$$

As a result, men are indifferent between all potential partners of the same observable type, and man i will choose the type of his partner as to maximize his utility

$$\max_{y \in \mathcal{Y}_0} \left(U_{xy} + \varepsilon_{iy} \right)$$

and the conditional probability of choosing y given one's type being x is

$$\mu_{y|x} = \arg\max_{y \in \mathcal{Y}_0} \left(U_{xy} + \varepsilon_{iy} \right),$$

with similar expression holding for the other side of the market.

1.3.2 Entropy of a Discrete Choice problem

In this section we deal with the problem of recovering U_{xy} from $\mu_{y|x}$, and we introduce a general methodology to do so based on "generalized entropy," a name which arises from reasons which will soon become clear. In the following, for any (A_{xy}) we denote $\mathbf{A_{x\cdot}} = (A_{x1}, \ldots, A_{x|\mathcal{Y}|})$ and $\mathbf{A_{\cdot y}} = (A_{1y}, \ldots, A_{|\mathcal{X}|y})$. Consider a randomly chosen man in group x. His expected utility (conditional to belonging to this group) is

$$G_x(\mathbf{U_{x\cdot}}) = \mathbb{E}_{\mathbf{P}_x}\left(\max_{y\in\mathcal{Y}}(U_{xy} + \varepsilon_y, \varepsilon_0)\right),$$

where the expectation is taken over the random vector $(\varepsilon_0, \ldots, \varepsilon_{|\mathcal{Y}|}) \sim \mathbf{P}_x$. First note that for any two numbers a, b and random variables (ε, η), the derivative of $E\max(a+\varepsilon, b+\eta)$ with respect to a is simply the probability that $a+\varepsilon$ is larger than $b+\eta$. Applying this to the function G_x, we get

$$\frac{\partial G_x}{\partial U_{xy}}(\mathbf{U_{x\cdot}}) = \Pr(U_{xy}+\varepsilon_{iy} \geq U_{xz}+\varepsilon_{iz} \text{ for all } z \in \mathcal{Y}_0).$$

But the right-hand side is simply the probability that a man of group x partners with a woman of group y; and therefore

$$\frac{\partial G_x}{\partial U_{xy}}(\mathbf{U_{x\cdot}}) = \frac{\mu_{xy}}{n_x} = \mu_{y|x}. \tag{2}$$

As the expectation of the maximum of linear functions of the (U_{xy}), G_x is a convex function of $\mathbf{U_{x\cdot}}$. Now consider the function

$$G_x^*(\boldsymbol{\mu}_{\cdot|\mathbf{x}}) = \max_{\tilde{\mathbf{U}}_{\mathbf{x}\cdot}=(\tilde{U}_{x1},\dots,\tilde{U}_{x|\mathcal{Y}|})} \left(\sum_{y\in\mathcal{Y}} \mu_{y|x}\tilde{U}_{xy} - G_x(\tilde{\mathbf{U}}_{\mathbf{x}\cdot}) \right) \tag{3}$$

whenever $\sum_{y\in\mathcal{Y}} \mu_{y|x} \leq 1$, $G_x^*(\boldsymbol{\mu}_{\cdot|\mathbf{x}}) = +\infty$ otherwise. Mathematically speaking, G_x^* is the *Legendre-Fenchel transform*, or *convex conjugate* of G_x. Like G_x and for the same reasons, it is a convex function. By the envelope theorem, at the optimum in the definition of G_x^*

$$\frac{\partial G_x^*}{\partial \mu_{y|x}}(\boldsymbol{\mu}_{\cdot|\mathbf{x}}) = U_{xy} \tag{4}$$

As a consequence, for any $y \in \mathcal{Y}$, U_{xy} is identified from $\boldsymbol{\mu}_{\cdot|\mathbf{x}}$, the observed matching patterns of men of group x.

Going back to (3), convex duality implies that if $\boldsymbol{\mu}_{\cdot|\mathbf{x}}$ and $\mathbf{U}_{\mathbf{x}\cdot}$ are related by (2), then

$$G_x(\mathbf{U}_{\mathbf{x}\cdot}) = \sum_{y \in \mathcal{Y}} \mu_{y|x} U_{xy} - G_x^*(\boldsymbol{\mu}_{\cdot|\mathbf{x}}). \tag{5}$$

Letting Y_i^* be the optimal choice of y given $\varepsilon_{i\cdot}$, one has

$$-G_x^*\left(\boldsymbol{\mu}_{\cdot|\mathbf{x}}\right) = G_x(\mathbf{U}_{\mathbf{x}\cdot}) - \sum_{y \in \mathcal{Y}} \mu_{y|x} U_{xy} = \mathbb{E}\left[\varepsilon_{iY_i^*}\right]. \tag{6}$$

Hence, $G_x^*\left(\mu_{.|x}\right)$ is interpreted as the expected amount of heterogeneity needed to rationalize the choice probabilities of an agent of type x, and we call it the *generalized entropy* of the corresponding discrete choice problem.

1.3.3 Social welfare

The notion of generalized entropy aggregates in a very convenient way, and leads us to derive an expression for the social welfare. Define H_y similarly as G_x: a randomly chosen woman of group y expects to get utility

$$H_y(\mathbf{V}_{\cdot \mathbf{y}}) = \mathbb{E}_{\mathbf{Q}_y}\left(\max_{x \in \mathcal{X}}(V_{xy} + \eta_x, \eta_0)\right),$$

so that the social surplus $\mathcal{W}$ is simply the sum of the expected utilities of all groups of men and women:

$$\mathcal{W} = \sum_{x \in \mathcal{X}} n_x G_x(\mathbf{U}_{\mathbf{x}\cdot}) + \sum_{y \in \mathcal{Y}} m_y H_y(\mathbf{V}_{\cdot \mathbf{y}}),$$

but by identity (5), we get

$$\begin{aligned} G_x(\mathbf{U}_{\mathbf{x}\cdot}) &= \sum_{y \in \mathcal{Y}} \mu_{y|x} U_{xy} - G_x^*\left(\boldsymbol{\mu}_{\cdot|\mathbf{x}}\right) \text{ and} \\ H_y(\mathbf{V}_{\cdot \mathbf{y}}) &= \sum_{x \in \mathcal{X}} \mu_{x|y} V_{xy} - H_y^*(\boldsymbol{\mu}_{\cdot|\mathbf{y}}), \end{aligned}$$

so summing over the total number of men and women, and using $U_{xy} + V_{xy} = \mathbf{\Phi}_{xy}$, and defining

$$\mathcal{E}(\mu) := \sum_{x \in \mathcal{X}} n_x G_x^*(\boldsymbol{\mu}_{\cdot|\mathbf{x}}) + \sum_{y \in \mathcal{Y}} m_y H_y^*(\boldsymbol{\mu}_{\cdot|\mathbf{y}}),$$

we get an expression for the value of the total surplus:

$$\begin{aligned}\mathcal{W} &= \sum_{x\in\mathcal{X}} n_x \underbrace{G_x(\mathbf{U_{x\cdot}})}_{u_x} + \sum_{y\in\mathcal{Y}} m_y \underbrace{H_y(\mathbf{V_{\cdot y}})}_{v_y} \\ &= \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \mu_{xy}\Phi_{xy} - \mathcal{E}(\mu).\end{aligned}$$

The first part of this expression explains how the total surplus $\mathcal{W}$ is broken down at the individual level: the average expected equilibrium utility of men in group x is $u_x = G_x(\mathbf{U_{x\cdot}})$, and similarly for women. The second part of this expression explains how the total surplus is broken down at the level of the couples. We turn this into a formal statement.

Theorem 1. *(***Social and Individual Surpluses***) Under Assumptions 1, 2 and 3, the following holds:*

(i) the optimal matching μ maximizes the social gain over all feasible matchings $\mu \in \mathcal{M}$, that is

$$\mathcal{W} = \max_{\mu\in\mathcal{M}} \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \mu_{xy}\Phi_{xy} - \mathcal{E}(\mu). \qquad (7)$$

and equivalently, $\mathcal{W}$ is given by its dual expression

$$\begin{aligned} \mathcal{W} &= \min_{U_{xy},V_{xy}} \sum_{x\in\mathcal{X}} n_x G_x\left(U_{x.}\right) + \sum_{y\in\mathcal{Y}} m_y H_y\left(V_{.y}\right) \quad (8) \\ & \quad s.t.\ U_{xy} + V_{xy} \geq \Phi_{xy}. \end{aligned}$$

(ii) A man i of group x who marries a woman of group y^* obtains utility

$$U_{xy^*} + \varepsilon_{iy^*} = \max_{y\in\mathcal{Y}_0} \left(U_{xy} + \varepsilon_{iy}\right)$$

where $U_{x0} = 0$, and the U_{xy}'s are solution to (8).

(iii) The average expected utility of the men of group x is $u_x = G_x(\mathbf{U_{x}}.)$.

(iv) Parts (ii) and (iii) transpose to the other side of the market with the obvious changes.

The right-hand side of equation (7) gives the value of the social surplus when the matching patterns are (μ_{xy}).

The first term $\sum_{xy} \mu_{xy} \Phi_{xy}$ reflects "group preferences": if groups x and y generate more surplus when matched, then they should be matched with higher probability. On the other hand, the second and the third term reflect the effect of the dispersion of individual affinities, conditional on observed characteristics: those men i in a group x that have more affinity to women of group y should be matched to this group with a higher probability. In the one-dimensional Beckerian example, a higher x or y could reflect higher education. If the marital surplus is complementary in the educations of the two partners, Φ_{xy} is supermodular and the first term is maximized when matching partners with similar education levels (as far as feasibility constraints allow.) But because of the dispersion of marital surplus that comes from the ε and η terms, it will be optimal to have some marriages between PhDs and high-school drop-outs.

To interpret the formula further, start with the case when unobserved heterogeneity is neglectable with respect to the magnitude of the observable surplus: $\tilde{\Phi}_{ij} \simeq \Phi_{xy}$ if

$x_i = x$ and $y_j = y$. Then we know that the observed matching μ must maximize the first term in (7); but this is precisely what the more complicated expression $\mathcal{W}(\mu)$ above boils down to if we scale up the values of Φ to infinity. On the other hand, if the data is so poor that unobserved heterogeneity dominates ($\Phi \simeq 0$), then the analyst should observe something that, to her, looks like completely random matching. In the intermediate case in which some of the variation in marital surplus is driven by observable types (through the Φ_{xy}) and some is carried by the unobserved heterogeneity terms ε_{iy} and η_{xj}, the market equilibrium trades off matching on observable types and matching on unobserved characteristics, as measured by $\mathcal{E}(\mu)$.

1.3.4 Identification

A consequence for identification of Theorem 1, is the following:

Theorem 2. *The following hold:*

(i) U_{xy} *is identified by the equivalent set of relations*

$$U_{xy} = \frac{\partial G_x^*}{\partial \mu_{y|x}}\left(\boldsymbol{\mu}_{\cdot|\mathbf{x}}\right) \text{ for } y \in \mathcal{Y}, \tag{9}$$

or equivalently

$$\mu_{y|x} = \frac{\partial G_x}{\partial U_{xy}}\left(\mathbf{U}_{\mathbf{x}\cdot}\right) \text{ for } y \in \mathcal{Y}, \tag{10}$$

and a similar result holds for V_{xy}*.*

(ii) As a result, Φ_{xy} *is identified by*

$$\begin{aligned} \Phi_{xy} &= \frac{\partial G_x^*}{\partial \mu_{y|x}}\left(\boldsymbol{\mu}_{\cdot|\mathbf{x}}\right) + \frac{\partial H_y^*}{\partial \mu_{x|y}}\left(\boldsymbol{\mu}_{\cdot|\mathbf{y}}\right) \\ &= \frac{\partial \mathcal{E}}{\partial \mu_{xy}}(\mu). \end{aligned}$$

Proof. (i) By definition of G^* and H^*, one has (Fenchel inequality)

$$\begin{aligned} G_x(\mathbf{U}_{\mathbf{x}\cdot}) &\geq \sum_{y\in\mathcal{Y}} \mu_{y|x} U_{xy} - G_x^*\left(\boldsymbol{\mu}_{\cdot|\mathbf{x}}\right) \quad (11) \\ H_y(\mathbf{V}_{\cdot\mathbf{y}}) &\geq \sum_{x\in\mathcal{X}} \mu_{x|y} V_{xy} - H_y^*(\boldsymbol{\mu}_{\cdot|\mathbf{y}}) \quad (12) \end{aligned}$$

with equality if and only if (9), or equivalently (10) holds. Summing over the population, one has

$$\begin{aligned} & \sum_{x\in\mathcal{X}} n_x G_x(\mathbf{U}_{\mathbf{x}\cdot}) + \sum_{y\in\mathcal{Y}} m_y H_y(\mathbf{V}_{\cdot\mathbf{y}}) \\ \geq \; & \sum_{\substack{x\in\mathcal{X} \\ y\in\mathcal{Y}}} \mu_{xy} \boldsymbol{\Phi}_{xy} - \sum_{x\in\mathcal{X}} n_x G_x^*\left(\boldsymbol{\mu}_{\cdot|\mathbf{x}}\right) - \sum_{y\in\mathcal{Y}} m_y H_y^*(\boldsymbol{\mu}_{\cdot|\mathbf{y}}) \end{aligned}$$

but at optimum, both sides of this inequality coincide with $\mathcal{W}$, thus equality holds in (11) and (12). (9) and (10) follow.

(ii) follows from $\boldsymbol{\Phi}_{xy} = U_{xy} + V_{xy}$. □

The previous result does not assume that transfers are observed. In the case when transfers are observed, the systematic parts of pre-transfer utilities (α, γ) are also observed. This case is unlikely to occur in the context of family economics, where the econometrician typically does not observe transfers between partners, but it is typically the case in other settings where matching theory has been successfully applied, as the CEO compensation literature, for instance, where the compensation amount is often available. We state the following corollary:

Corollary 1. *Under the previous assumptions, denote (α, γ) the systematic parts of pre-transfer utilities and w the transfers from y to x. Then α_{xy} and γ_{xy} are identified by*

$$\alpha_{xy} = \frac{\partial G_x^*}{\partial \mu_{y|x}} \left(\boldsymbol{\mu}_{\cdot|\mathbf{x}}\right) - w_{xy} \text{ and } \gamma_{xy} = \frac{\partial H_y^*}{\partial \mu_{x|y}} \left(\boldsymbol{\mu}_{\cdot|\mathbf{y}}\right) + w_{xy}.$$

Therefore if transfers w_{xy} are observed, both pre-transfer utilities α_{xy} and γ_{xy} are also identified.

Proof. This follows from $U_{xy} = \alpha_{xy} + w_{xy}$ and $V_{xy} = \gamma_{xy} - w_{xy}$. □

All of the quantities in Theorem 1 can be computed by solving simple linear programming problems. This makes identification and estimation feasible in practice.

1.3.5 Comparative statics

Social welfare in matching markets depend on the surplus function and the distribution of types which we highlight by using the notation $\mathcal{W}(\Phi, n, m)$. This raises many intresting question about how welfare would be affected if the distribution of types were to change or if the shape of the surplus function were to be changed. In this section, we use the results of Theorem 1 and derive comparative statics results. Recall that $\mathcal{W}(\Phi, n, m)$ is given by the dual expressions

$$\mathcal{W} = \max_{\mu \in \mathcal{M}(n,m)} \sum_{xy} \mu_{xy} \Phi_{xy} - \mathcal{E}(\mu) \text{, and} \tag{13}$$

$$\mathcal{W} = \min_{U_{xy}+V_{xy}=\Phi_{xy}} \sum n_x G_x (U_{xy}) + \sum m_y H_y (V_{xy}) \tag{14}$$

where we have dropped the implicit dependence with respect to Φ, n and m. We observe on these formulae that $\mathcal{W}$ is a convex function of Φ_{xy}, and a concave function of (n, m).

By the envelope theorem in the primal problem (13), we get

$$\frac{\partial \mathcal{W}}{\partial \Phi_{xy}} = \mu_{xy}$$

while the envelope theorem in the dual problem (14) yields

$$\frac{\partial \mathcal{W}}{\partial n_x} = G_x\left(U_{xy}\right) = u_x \text{ and } \frac{\partial \mathcal{W}}{\partial m_y} = H_y\left(V_{xy}\right) = v_y.$$

A second differentiation yields

$$\frac{\partial u_x}{\partial n_{x'}} = \frac{\partial^2 \mathcal{W}}{\partial n_x \partial n_{x'}} = \frac{\partial u_{x'}}{\partial n_x} \tag{15}$$

(and similarly $\frac{\partial u_x}{\partial m_y} = \frac{\partial v_y}{\partial n_x}$ and $\frac{\partial v_y}{\partial m_{y'}} = \frac{\partial v_{y'}}{\partial m_y}$). Next, the cross-derivative of $\mathcal{W}$ with respect to $n_{x'}$ and Φ_{xy} yields

$$\frac{\partial u_{x'}}{\partial \Phi_{xy}} = \frac{\partial^2 \mathcal{W}}{\partial n_{x'} \partial \Phi_{xy}} = \frac{\partial \mu_{xy}}{\partial n_{x'}}. \tag{16}$$

Finally, differentiating $\mathcal{W}$ twice with respect to Φ_{xy} and

$\Phi_{x'y'}$ yields

$$\frac{\partial \mu_{xy}}{\partial \Phi_{x'y'}} = \frac{\partial^2 \mathcal{W}}{\partial \Phi_{xy} \partial \Phi_{x'y'}} = \frac{\partial \mu_{x'y'}}{\partial \Phi_{xy}}. \tag{17}$$

1.3.6 Examples

While Theorem 2 provides a general way of computing surplus and utilities, they can often be derived in closed form. We give some useful examples below. In all formulæ below, the proportions and numbers of single men in feasible matchings are computed as

$$\mu_{0|x} = 1 - \sum_{y \in Y} \mu_{y|x} \quad \text{and} \quad \mu_{x0} = n_x - \sum_{y \in Y} \mu_{xy}, \tag{18}$$

and similarly for women.

Our first example is the classical model of Choo and Siow, which is obtained as a particular case of the results in the previous section when the $\mathbf{P}_x$ and $\mathbf{Q}_y$ distributions are iid standard type I extreme value:

Example 1 (Choo and Siow). *Maintain Assumptions 1 and 2. Assume further that $\mathbf{P}_x$ and $\mathbf{Q}_y$ are the distributions of i.i.d. standard type I extreme value random* *variables (which is a particular case of Assumption 3). Then*

$$
\begin{aligned}
G_x(\mathbf{U_{x\cdot}}) &= \log\left(1 + \sum_{y\in\mathcal{Y}} \exp(U_{xy})\right) \\
\text{and } G_x^*(\boldsymbol{\mu}_{\cdot|x}) &= \mu_{0|x}\log(\mu_{0|x}) + \sum_{y\in\mathcal{Y}} \mu_{y|x}\log\mu_{y|x}.
\end{aligned}
$$

where $\mu_{0|x}$ *is defined in (18). Expected utilities are* $u_x = -\log\mu_{0|x}$ *and* $v_y = -\log\mu_{0|y}$. *The generalized entropy is*

$$
\mathcal{E}(\mu) = \sum_{\substack{x\in\mathcal{X} \\ y\in\mathcal{Y}_0}} \mu_{xy}\log\mu_{y|x} + \sum_{\substack{y\in\mathcal{Y} \\ x\in\mathcal{X}_0}} \mu_{xy}\log\mu_{x|y}, \quad (19)
$$

and surplus and matching patterns are linked by

$$
\Phi_{xy} = 2\log\mu_{xy} - \log\mu_{x0} - \log\mu_{0y}, \quad (20)
$$

which is Choo and Siow's (2006) identification result.

Note that as announced after Theorem 1, the generalized entropy $\mathcal{E}$ boils down here to the usual definition of entropy. The Choo and Siow model is the simplest example which fits into McFadden's Generalized Extreme Value (GEV) framework in the Appendix. This framework includes most specifications used in discrete choice studies. As a more complex example of a GEV distribution, we turn to a nested logit model.

Example 2 (Nested Logit). *Suppose for instance that men of a given group x are concerned about the social group of their partner and her education, so that the type $y = (s, e)$. We can allow for correlated preferences by modeling this as a nested logit in which educations are nested within social groups. Let $\mathbf{P}_x$ have cdf*

$$F(w) = \exp\left(-\exp(-w_0) - \sum_s \left(\sum_e \exp(-w_{se}/\sigma_s)\right)^{\sigma_s}\right)$$

This is a particular case of the Generalized Extreme Value (GEV) framework. The numbers $1/\sigma_s$ describe the correlation in the surplus generated with partners of different education levels within social group s. Then (dropping the x indices for notational simplicity, so that for instance μ_s denotes the number of matches with women in social *group s)*

$$
\begin{aligned}
G(\mathbf{U}.) &= \log\left(1+\sum_s\left(\sum_e \exp(U_{se}/\sigma_s)\right)^{\sigma_s}\right), \textit{ and} \\
G^*(\boldsymbol{\mu}.) &= \mu_0 \log \mu_0 + \sum_s (1-\sigma_s)\,\mu_s \log \mu_s \\
&\quad + \sum_s \sigma_s \sum_e \mu_{se} \log \mu_{se}.
\end{aligned}
$$

where $\mu_{0|x}$ is again defined in (18). The expected utility is still $u = -\log \mu_0$.

If the heterogeneity structure is the same for all men and all women (with possibly different dispersion parameters σ for men and τ for women), then the expressions of $\mathcal{E}(\mu)$ and $\mathcal{W}(\mu)$ can easily be obtained. The social surplus from a match between a man of group $x = (s, e)$ and a woman of group $y = (s', e')$ is identified by

$$
\Phi_{xy} = \log \frac{\mu_{xy}^{\sigma^x_{s'}+\tau^y_s}\, \mu_{x,s'}^{1-\sigma^x_{s'}}\, \mu_{s,y}^{1-\tau^y_s}}{\mu_{x0}\mu_{0y}}.
$$

In particular, when there is only one possible value of s, we get a heteroskedastic model, where $\mathbf{P}_x$ is an extreme value type I distribution with scale parameter σ_x and $\mathbf{Q}_y$ with scale parameter τ_y; the expected utilities are

$$u_x = -\sigma_x \log \mu_{0|x} \quad \text{and } v_y = -\tau_y \log \mu_{0|y}$$

and the general identification formula gives

$$\Phi_{xy} = (\sigma_x + \tau_y) \log \mu_{xy} - \sigma_x \log \mu_{x0} - \tau_y \log \mu_{0y}. \tag{21}$$

We investigate a particular case of this specification in the next example: the Random Scalar Coefficient (RSC) model, where the dimension of $\zeta_x(y)$ and $\boldsymbol{\varepsilon}_i$ is one. As we argue below, this assumption much simplifies the computations. Assuming further that the distribution of $\boldsymbol{\varepsilon}_i$ is uniform, one is led to what we call the Random Uniform Scalar Coefficient Model (RUSC). This last model has one additional advantage: it yields simple closed-form expressions, even though it does not belong to the Generalized Extreme Value (GEV) class.

Example 3 (Random [Uniform] Scalar Coefficient (RSC/RUSC) models). *Assume that for each man i in group x,*

$$\varepsilon_{iy} = \varepsilon_i \zeta_x(y),$$

where $\zeta_x(y)$ is a scalar index of the observable characteristics of women which is the same for all men in the same group x, and the ε_i's are iid random variables which are assumed to be continuously distributed according to a c.d.f. F_ε (which could also depend on x.) We call this model the Random Scalar Coefficient (RSC) model; and the entropy is

$$\mathcal{E}(\mu) = \sum_{xy} \mu_{xy} \left(\zeta_x(y) \bar{e}_x(y) + \xi_y(x) \bar{f}_y(x) \right),$$

where $\bar{e}_x(y)$ is the expected value of ε on the interval $[a, b]$ defined by

$$F_\varepsilon(a) = \sum_{z|\zeta_x(z)<\zeta_x(y)} \mu_{z|x} \text{ and } F_\varepsilon(b) = \sum_{z|\zeta_x(z)\leq\zeta_x(y)} \mu_{z|x},$$

and $\bar{f}_y(x)$ is defined similarly.

Assuming further that the ε_i are uniformly distributed over $[0, 1]$, we call this model the Random Uniform Scalar Coefficient (RUSC) model. In this case, simpler formulæ can be given. For any $x \in \mathcal{X}$, let S^x be the square matrix with elements $S^x_{yy'} = \max(\zeta_x(y), \zeta_x(y'))$ for

$y, y' \in \mathcal{Y}_0$. Define T^x by $T^x_{yy'} = S^x_{y0} + S^x_{0y'} - S^x_{yy'} - S^x_{00}$, and let $\sigma^x_y = S^x_{00} - S^x_{y0}$.

Then G^*_x is quadratic with respect to $\boldsymbol{\mu}_{\cdot|\mathbf{x}}$:

$$G^*_x(\boldsymbol{\mu}_{\cdot|\mathbf{x}}) = \frac{1}{2}(\boldsymbol{\mu}_{\cdot|\mathbf{x}}{}' T^x \boldsymbol{\mu}_{\cdot|\mathbf{x}} + 2\sigma^x . \boldsymbol{\mu}_{\cdot|\mathbf{x}} - S^x_{00}).$$

If we now assume that preferences have such a structure for every group x of men and for every group y of women (so that $\eta_{xj} = \eta_j \xi_y(x)$), then the generalized entropy is quadratic in μ:

$$\mathcal{E}(\mu) = \frac{1}{2}(\mu' A \mu + 2B\mu + c),$$

where the expressions for A, B and c have an expression in closed form. As a consequence, the optimal matching solves a simple quadratic problem.

1.3.7 Parametric Inference

Theorem 1 shows that, given a specification of the distribution of the unobserved heterogeneities $\mathbf{P}_x$ and $\mathbf{Q}_y$, the model spelled out by Assumptions 1, 2, and 3 is exactly nonparametrically identified from the observation of a single market, as long as μ_{xy}, μ_{x0} and $\mu_{y0} > 0$ for each x and y. In particular, one recovers the fact that the model of Choo and Siow is identified from the observation of a single market. There is therefore no way to test separability from the observation of a single market. When multiple, but similar markets (in the sense that Φ_{xy}, $\mathbf{P}_x$ and $\mathbf{Q}_y$ are identical across them) are observed, the model is nonparametrically overidentified given a fixed specification of $\mathbf{P}_x$ and $\mathbf{Q}_y$. The flexibility allowed by Assumption 3 can then be used to infer information about these distributions.

In this section, it is assumed that a single market is observed. While the formula in Theorem 1(i) gives a straightforward nonparametric estimator of the systematic surplus function Φ, with multiple surplus-relevant observable groups it will be very unreliable. In addition, we do not know the distributions $\mathbf{P}_x$ and $\mathbf{Q}_y$. Both of these remarks point towards the need to specify a parametric model in most applications. Such a model would be described by a family of joint surplus functions and distributions

$$\Phi_{xy}^{\lambda},\ \mathbf{P}_x^{\lambda},\ \mathbf{Q}_y^{\lambda}$$

for λ in some subset of a finite-dimensional parameter space Λ.

We observe a sample of $\hat{N}_{ind}$ individuals; $N_{ind} = \sum_x \hat{n}_x + \sum_y \hat{m}_y$, where $\hat{n}_x$ (resp. $\hat{m}_y$) denotes the number of men of group x (resp. women of group y) in the sample. Let $\hat{\mu}$ the observed matching; we assume that the data was generated by the parametric model above, with an interior parameter vector λ_0.

Recall the expression of the social surplus:

$$\mathcal{W}\left(\Phi^{\lambda}\right) = \max_{\mu \in \mathcal{M}(\hat{n},\hat{m})} \left(\sum_{x,y} \mu_{xy} \Phi_{xy}^{\lambda} - \mathcal{E}^{\lambda}(\mu) \right)$$

Let μ^λ be the optimal matching. Of course, computing μ^λ is a crucial issue. We will show in Section **??** how it can be computed, in some cases very efficiently. For now we focus on statistical inference on λ and we propose a very general Maximum Likelihood method.

We will use Conditional Maximum Likelihood (CML) estimation, where we condition on the observed margins $\hat{n}_x$ and $\hat{m}_y$. The log-likelihood of marital choice is

$$\sum_{y \in \mathcal{Y}_0} \frac{\hat{\mu}_{xy}}{\hat{n}_x} \log \frac{\mu^\lambda_{xy}}{\hat{n}_x} \text{ for each man of group } x$$

and a similar expression for each woman of group y. Under Assumptions 1, 2 and 3, the choice of each individual is stochastic in that it depends on his vector of unobserved heterogeneity, and these vectors are independent across men and women. Hence the log-likelihood of the sample is the sum of the individual log-likelihood elements:

$$\log L(\lambda)$$

$$= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_0} \hat{\mu}_{xy} \log \frac{\mu_{xy}^{\lambda}}{\hat{n}_x} + \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}_0} \hat{\mu}_{xy} \log \frac{\mu_{xy}^{\lambda}}{\hat{m}_y}$$

$$= \left\{ \begin{array}{c} 2 \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \hat{\mu}_{xy} \log \frac{\mu_{xy}^{\lambda}}{\sqrt{\hat{n}_x \hat{m}_y}} \\ + \sum_{x \in \mathcal{X}} \hat{\mu}_{x0} \log \frac{\mu_{x0}^{\lambda}}{\hat{n}_x} + \sum_{y \in \mathcal{Y}} \hat{\mu}_{0y} \log \frac{\mu_{0y}^{\lambda}}{\hat{m}_y} \end{array} \right\}.$$

The Conditional Maximum Likelihood Estimator $\hat{\lambda}^{MLE}$ given by

$$\frac{\partial \log L}{\partial \lambda} \left(\hat{\lambda}^{MLE} \right) = 0$$

is consistent, asymptotically normal and asymptotically efficient under the usual set of assumptions.

Example 2 continued. In the Nested Logit model of Example 2, where the type of men and women are respectively (s_x, e_x) *and* (s_y, e_y)*, one can take* $\sigma_{s_y}^{s_x e_x}$ *and*

$\sigma_{s_x}^{s_y,e_y}$ as parameters. Assume that there are N_s social categories and N_e classes of education. There are $N_s^2 \times N_e^2$ equations, so one can parameterize the surplus function Φ^θ by a parameter θ of dimension less or equal than $N_s^2 \times N_e^2 - N_s^2 \times N_e - N_s \times N_e^2$. Letting $\lambda = \left(\sigma_{s_y}^{s_x e_x}, \sigma_{s_x}^{s_y,e_y}, \theta\right)$, μ^λ is the solution in M to the system of equations

$$\Phi_{xy}^{\theta} = \log \frac{\mu_{xy}^{\sigma_{s'}^{x}+\tau_s^y} \mu_{x,s'}^{1-\sigma_{s'}^{x}} \mu_{s,y}^{1-\tau_s^y}}{(n_x - \sum_y \mu_{xy})(m_y - \sum_x \mu_{xy})}, \ \forall x \in \mathcal{X}, \ y \in \mathcal{Y}$$

and the log-likelihood can be deduced. □

The Maximum Likelihood estimation has many advantages: (i) it allows for joint parametric estimation of the surplus function and of the unobserved heterogeneities; (ii) it enjoys desirable statistical properties in terms of statistical efficiency; (iii) its asymptotic properties are well-known. However, there is no guarantee that the log-likelihood shall be a concave function in general, and hence maximization of the likelihood may lead to practical problems in some situations. In some of these cases, an alternative method, based on moments, is available.