

Cupid's Invisible Hand: Social Surplus and Identification in Matching Models

Bernard Salanié, Alfred Galichon

► **To cite this version:**

Bernard Salanié, Alfred Galichon. Cupid's Invisible Hand: Social Surplus and Identification in Matching Models. 2012. hal-01053710

HAL Id: hal-01053710

<https://hal-sciencespo.archives-ouvertes.fr/hal-01053710>

Submitted on 1 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cupid's Invisible Hand:

Social Surplus and Identification in Matching Models

Alfred Galichon¹ Bernard Salanié²

September 17, 2012³

¹Economics Department, Sciences Po, Paris; e-mail: alfred.galichon@sciences-po.fr

²Department of Economics, Columbia University; e-mail: bsalanie@columbia.edu.

³The authors are grateful to Pierre-André Chiappori, Chris Conlon, Jim Heckman, Robert McCann, Jean-Marc Robin, and many seminar participants for useful comments and discussions. Galichon gratefully acknowledges support from Chaire “Assurance et Risques Majeurs” and FiME, Laboratoire de Finance des Marchés de l’Energie. Part of the research underlying this paper was done when Galichon was visiting the University of Chicago Booth School of Business and Columbia University, and when Salanié was visiting the Toulouse School of Economics. Galichon thanks the Alliance program for its support, and Salanié thanks the Georges Meyer endowment.

Abstract

We investigate a model of one-to-one matching with transferable utility when some of the characteristics of the players are unobservable to the analyst. We allow for a wide class of distributions of unobserved heterogeneity, subject only to a separability assumption that generalizes Choo and Siow (2006). We first show that the stable matching maximizes a social gain function that trades off the average surplus due to the observable characteristics and a generalized entropy term that reflects the impact of matching on unobserved characteristics. We use this result to derive simple closed-form formulæ that identify the joint surplus in every possible match and the equilibrium utilities of all participants, given any known distribution of unobserved heterogeneity. If transfers are observed, then the pre-transfer utilities of both partners are also identified. We also present a very fast algorithm that computes the optimal matching for any specification of the joint surplus. We conclude by discussing some empirical approaches suggested by these results.

Keywords: matching, marriage, assignment, hedonic prices.

JEL codes: C78, D61, C13.

Introduction

Since the seminal contribution of Becker (1973), economists have modeled the marriage market as a matching problem in which each potential match generates a marital surplus. Given transferable utilities, the distributions of tastes and of desirable characteristics determine equilibrium shadow prices, which in turn explain how partners share the marital surplus in any realized match. This insight is not specific of the marriage market: it characterizes the “assignment game” (Shapley and Shubik 1972), i.e. models of matching with transferable utilities. These models have also been applied to competitive equilibrium with hedonic pricing (Chiappori, McCann, and Nesheim 2010) and the market for CEOs (Tervio (2008) and Gabaix and Landier (2008).) We will show how our results can be used in these three contexts; but for concreteness, we often refer to partners as men and women in the exposition of the main results.

While Becker presented the general theory, he focused on the special case in which the types of the partners are one-dimensional and are complementary in producing surplus. As is well-known, the socially optimal matches then exhibit *positive assortative matching*: higher types pair up with higher types, and lower types with lower types. Moreover, the resulting configuration is stable, it is in the core of the corresponding matching game, and it can be efficiently implemented by classical optimal assignment algorithms.

This sorting result is both simple and powerful; but its implications are also quite unrealistic and at variance with the data, in which matches are observed between partners with quite different characteristics. To account for this wider variety of matching patterns, one could introduce search frictions, as in Shimer and Smith (2000) or Jacquemet and Robin (2011). But the resulting model is hard to handle, and under some additional conditions it still implies assortative matching. An alternative solution consists in allowing the joint surplus of a match to incorporate latent characteristics—heterogeneity that is unobserved by the analyst. Choo and Siow (2006) showed that it can be done in a way that yields a highly tractable model in large populations, provided that the unobserved heterogeneities enter the marital surplus quasi-additively and that they are distributed as standard type I

extreme value terms. Then the usual apparatus of multinomial logit discrete choice models applies, linking marriage patterns to marital surplus in a very simple manner. Choo and Siow (2006) used this model to link the changes in gains to marriage and abortion laws; Siow and Choo (2006) applied it to Canadian data to measure the impact of demographic changes. It has also been used to study increasing returns in marriage markets (Botticini and Siow 2008) and to test for complementarities across partner educations (Siow 2009); and, in a heteroskedastic version, to estimate the changes in the returns to education on the US marriage market (Chiappori, Salanié, and Weiss 2012).

We revisit here the theory of matching with transferable utilities in the light of Choo and Siow's insights; and we extend this framework to quite general distributions of unobserved variations in tastes. Our main contributions are twofold. First, we prove that the optimal matching maximizes a very simple function of the observable only. With quasi-additive surplus, the market equilibrium maximizes a social surplus function that consists of two terms: a term that describes assortativeness on the observed characteristics; and a generalized entropic term that describes the random character of matching conditional on observed characteristics. While the first term tends to match partners with complementary observed characteristics, the second one pulls towards randomly assigning partners to each other. The social gain from any matching patterns trades off these two terms. In particular, when unobserved heterogeneity is distributed as in Choo and Siow (2006), the generalized entropy is simply the usual entropy measure. The maximization of this social surplus function has very straightforward consequences in terms of identification, both when equilibrium transfers are observed and when they are not. In fact, most quantities of interest can be obtained from derivatives of the terms that constitute generalized entropy. We show in particular that the joint surplus from matching is (minus) a derivative of the generalized entropy, computed at the observed matching. The expected and realized utilities of all types of men and women follow just as directly. If moreover equilibrium transfers are observed, then we also identify the pre-transfer utilities on both sides of the market. To prove these results, we combine tools from duality and convexity theory, and we construct the Legendre-Fenchel

transform of the expected utilities of agents. A similar approach was used independently by Decker, Lieb, McCann, and Stephens (2012) to prove the uniqueness of the equilibrium and derive some of its properties in the Choo and Siow framework.

Our second contribution is to delineate an empirical approach to parametric estimation in this class of models, using maximum likelihood. Indeed, our nonparametric identification results rely on the strong assumption that the distribution of the unobservables is known, while in practice the analyst will want to estimate its parameters. The large number of covariates available to model the surplus function also limits the scope of nonparametric estimation. Maximum likelihood estimation is thus a natural recourse; but since evaluating the likelihood requires solving for the optimal matching, computational considerations loom large in matching models. We provide an efficient algorithm that maximizes the social surplus for any joint surplus function and computes the optimal matching, as well as the expected utilities in equilibrium. To do this, we adapt the Iterative Projection Fitting Procedure (known to some economists as RAS) to the structure of this problem. IPFP is a very stable and very simple algorithm; we report simulations that show that is much faster than alternative solvers. Finally, we show that in the context of the Choo and Siow (2006) model the maximum likelihood is a simple moment matching estimator, and we give a geometric interpretation.

There are other approaches to estimating matching models with unobserved heterogeneity; see the handbook chapter by Graham (2011). Fox (2010) in particular exploits a “rank-order property” and pools data across many similar markets; see Fox (2011) and Bajari and Fox (2009) for applications. More recently, Fox and Yang (2012) focus on identifying the complementarity between unobservable characteristics. We discuss the pros and cons of various methods in our conclusion.

Section 1 sets up the model and the notation. We prove our main results in Section 2, and we specialize them to leading examples in Section 3. Our results open the way to new and richer specifications; Section 4 explains how to estimate them using maximum likelihood, and how to use various restrictions to identify the underlying parameters. We

present in Section 5 our IPFP algorithm, which greatly accelerates computations. Finally, Section 6 specializes our results to a restricted but useful model¹.

1 The Assignment Problem with Unobserved Heterogeneity

1.1 The Populations

Throughout the paper, we maintain the basic assumptions of the transferable utility model of Choo and Siow (2006): utility transfers between partners are unconstrained, matching is frictionless, and there is no asymmetric information among potential partners. We call the partners “men” and “women”, but our results are clearly not restricted to the marriage market.

Men are denoted by $i \in \mathcal{I}$ and women by $j \in \mathcal{J}$. A *matching* $(\tilde{\mu}_{ij})$ is a matrix such that $\tilde{\mu}_{ij} = 1$ if man i and woman j are matched, 0 otherwise. A matching is *feasible* if for every i and j ,

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

with equality for individuals who are married. Single individuals are “matched with 0”: $\tilde{\mu}_{i0} = 1$ or $\tilde{\mu}_{0j} = 1$. For completeness, we should add the requirement that $\tilde{\mu}_{ij}$ is integral ($\tilde{\mu}_{ij} \in \{0, 1\}$). However it is known since at least Shapley and Shubik (1972) that this constraint is not binding, so we will omit it.

A hypothetical match between man i and woman j allows them to share a total utility $\tilde{\Phi}_{ij}$; the division of this total utility between them is done through utility transfers whose value is determined in equilibrium. Singles get utilities $\tilde{\Phi}_{i0}, \tilde{\Phi}_{0j}$. Following Gale and Shapley (1962) for matching with non-transferable utility, we focus on the set of *stable matchings*. A feasible matching is stable if there exists a division of the surplus in each realized match that makes it impossible for any man k and woman l to both achieve strictly higher utility by pairing up together, and for any agent to achieve higher utility by being single. More

¹This paper builds on and significantly extends our earlier discussion paper (Galichon and Salanié 2010), which is now obsolete.

formally, let \tilde{u}_i denote the utility man i gets in his current match; denote \tilde{v}_j the corresponding utilities for woman j . Then by definition $\tilde{u}_i + \tilde{v}_j = \tilde{\Phi}_{ij}$ if they are matched, that is if $\tilde{\mu}_{ij} > 0$; and $\tilde{u}_i = \tilde{\Phi}_{i0}$ (resp. $\tilde{v}_j = \tilde{\Phi}_{0j}$) if i (resp. j) is single. Stability requires that for every man k and woman l , $\tilde{u}_k \geq \tilde{\Phi}_{k0}$ and $\tilde{v}_l \geq \tilde{\Phi}_{0l}$, and $\tilde{u}_k + \tilde{v}_l \geq \tilde{\Phi}_{kl}$ for any potential match (k, l) .

Finally, a *competitive equilibrium* is defined as a set of prices \tilde{u}_i and \tilde{v}_j and a feasible matching $\tilde{\mu}_{ij}$ such that

$$\tilde{\mu}_{ij} > 0 \text{ iff } j \in \arg \max_{j \in \mathcal{J} \cup \{0\}} (\tilde{\Phi}_{ij} - \tilde{v}_j) \text{ and } i \in \arg \max_{i \in \mathcal{I} \cup \{0\}} (\tilde{\Phi}_{ij} - \tilde{u}_i). \quad (1.1)$$

Shapley and Shubik (1972) showed that the set of stable matchings coincides with the set of competitive equilibria (and with the core of the assignment game); and that moreover, any stable matching achieves the maximum of the total surplus²

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

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

The set of stable matchings is generically a singleton; on the other hand, the set of prices \tilde{u}_i and \tilde{v}_j (or, equivalently, the division of the surplus into \tilde{u}_i and \tilde{v}_j) that support it is a product of intervals.

1.2 Groups

The analyst only observes some of the payoff-relevant characteristics that determine the surplus matrix $\tilde{\Phi}$. Following Choo and Siow (2006), we assume that she can only observe which *group* each individual belongs to. Each man $i \in \mathcal{I}$ belongs to one group $x_i \in \mathcal{X}$; and, similarly, each woman $j \in \mathcal{J}$ belongs to one group $y_j \in \mathcal{Y}$. Groups are defined by the intersection of characteristics which are observed by all men and women, and also by the analyst. On the other hand, men and women of a given group differ along some dimensions that they all observe, but which do not figure in the analyst's dataset.

²See also Gretsky, Ostroy, and Zame (1992) for the extension to a continuum of agents.

As an example, observed groups $x, y = (E, R)$ may consist of education and income. Education could take values $E \in \{D, G\}$ (dropout or graduate), and income class R could take values 1 to n_R . Groups may also incorporate information that is sometimes available to the econometrician, such as physical characteristics, religion, and so on. In this paper we take the numbers of groups $|\mathcal{X}|$ and $|\mathcal{Y}|$ to be finite in number; we return to the case of continuous groups in the conclusion.

We denote n_x the number of men in group x , and m_y the number of women in group y . Like Choo and Siow, we assume that there are a large number of men in any group x , and of women in any group y . More precisely, our statements in the following hold exactly when the number of individuals goes to infinity and the proportions of genders and groups converge. To simplify the exposition, we consider the limit of a sequence of large economies where the proportion of each group remains constant:

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; and the ratios (n_x/N) and (m_y/N) are constant.

With finite N we would need to introduce corrective terms; however, since N is the size of the total population rather than that of the sample we use for estimation, these terms can safely be neglected in applications to the marriage market for instance.

Another benefit of Assumption 1 is that it mitigates concerns about agents misrepresenting their characteristics. This is almost always a profitable deviation in finite populations; but as shown by Gretsky, Ostroy, and Zame (1999), the benefit from such manipulations goes to zero as the population is replicated. In the large markets limit, the equilibrium prices \tilde{u}_i and \tilde{v}_j are unique. We will therefore write “the equilibrium” in what follows.

The analyst does not observe some of the characteristics of the players, and she can only compute quantities that depend on the observed groups of the partners in a match. Hence she cannot observe $\tilde{\mu}$, and she must focus instead on the matrix of matches across groups

(μ_{xy}) . This is related to $(\tilde{\mu}_{ij})$ by

$$\mu_{xy} = \sum_{i,j} \mathbf{1}(x_i = x, y_j = y) \tilde{\mu}_{ij}$$

with the obvious extension to μ_{x0} and μ_{0y} .

The feasibility constraints on $\mu_{xy} \geq 0$ are

$$\forall x \in \mathcal{X}, \sum_{y \in \mathcal{Y}} \mu_{xy} \leq n_x ; \forall y \in \mathcal{Y}, \sum_{x \in \mathcal{X}} \mu_{xy} \leq m_y \quad (1.2)$$

which simply means that the number of married men (women) of group x (y) is not greater than the number of men (women) of group x (y). For future reference, we denote \mathcal{M} the set of $(|\mathcal{X}| |\mathcal{Y}| + |\mathcal{X}| + |\mathcal{Y}|)$ non-negative numbers (μ_{xy}) that satisfy these $(|\mathcal{X}| + |\mathcal{Y}|)$ equalities. Each element of \mathcal{M} is called a “matching” as it defines a feasible set of matches (and singles). For notational convenience, we shall denote μ_{x0} the number of single men of group x and μ_{0y} the number of single women of group y , and $\mathcal{X}_0 = \mathcal{X} \cup \{0\}$ and $\mathcal{Y}_0 = \mathcal{Y} \cup \{0\}$ the set of marital choices that are available to agents, including singlehood. Obviously,

$$\mu_{x0} = n_x - \sum_{y \in \mathcal{Y}} \mu_{xy} \text{ and } \mu_{0y} = m_y - \sum_{x \in \mathcal{X}} \mu_{xy}.$$

1.3 Matching Surpluses

Several approaches can be used to take this model to the data. A “brute force” method would use a parametric specification for the surplus $\tilde{\Phi}_{mw}$ and solve the system of equilibrium equations (1.1). The set of maximizers at the solution of this system defines the stable matchings, and can be compared to the observed matching in order to derive a minimum distance estimator of the parameters. However, there are two problems with this approach: it is very costly, and it is not clear at all what drives identification of the parameters. The literature has instead attempted to impose identifying assumptions that allow for more transparent identification. We follow here the framework of Choo and Siow (2006). We will

discuss other approaches in the conclusion, including those of Fox (2010) and Fox and Yang (2012).

Choo and Siow assumed that the utility surplus of a man i of group x (that is, such that $x_i = x$) who marries a woman of group y can be written as

$$\alpha_{xy} + \tau + \varepsilon_{iy}, \quad (1.3)$$

where α_{xy} is the systematic part of the surplus τ represents the utility transfer (possibly negative) that the man gets from his partner in equilibrium, and ε_{iy} is a standard type I extreme value random variation. If such a man remains single, he gets utility ε_{i0} ; that is to say, the systematic utilities of singles α_{x0} are normalized to zero. Similarly, the utility of a woman j of group $y_j = y$ who marries a man of group x can be written as

$$\gamma_{xy} - \tau + \eta_{xj}, \quad (1.4)$$

where τ is the utility transfer she leaves to her partner. A woman of group y gets utility η_{0j} if she is single, that is we adopt normalization $\gamma_{0y} = 0$.

As shown in Chiappori, Salanié, and Weiss (2012), the key assumption here is that the joint surplus created when a man i of group x marries a woman j of group y rule out interactions between their unobserved characteristics, conditional on (x, y) . This leads us to assume:

Assumption 2 (Separability) *There exists a vector Φ_{xy} such that the joint surplus from a match between a man i in group x and a woman j in group y is*

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

In Choo and Siow's formulation, the vector Φ is simply

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

which they call the *total systematic net gains to marriage*; and note that by construction, Φ_{x0} and Φ_{0y} are zero. It is easy to see that Assumption 2 is equivalent to specifying that if

two men i and i' belong to the same group x , and their respective partners j and j' belong to the same group y , then the total surplus generated by these two matches is unchanged if we shuffle partners: $\tilde{\Phi}_{ij} + \tilde{\Phi}_{i'j'} = \tilde{\Phi}_{ij'} + \tilde{\Phi}_{i'j}$. Note that in this form it is clear that we needn't adopt Choo and Siow's original interpretation of ε as a preference shock of the husband and η as a preference shock of the wife. To take an extreme example, we could equally have men who are indifferent over partners and are only interested in the transfer they receive, so that their ex post utility is τ ; and women who also care about some attractiveness characteristic of men, in a way that may depend on the woman's group. The net utility of women of group y would be $\varepsilon_{iy} - \tau$; the resulting joint surplus would satisfy Assumption 2 and all of our results would apply³. In other words, there is no need to assume that the term ε_{iy_j} was "created" by man i , nor that the term η_{jx_i} was "created" by the woman j ; it may perfectly be the opposite.

1.4 Choo and Siow's Identification Result

In addition to Assumptions 1 and 2, Choo and Siow (2006) specified the distribution of unobserved heterogeneities to be independent products of standard type I extreme values distributions:

Assumption 3 (Type-I extreme values distribution)

a) For any man i , the $(\varepsilon_{iy})_{y \in \mathcal{Y}_0}$ are drawn independently from a standard type I extreme value distribution;

b) For any woman j , the $(\eta_{xj})_{x \in \mathcal{X}_0}$ are drawn independently from a standard type I extreme value distribution;

c) These draws are independent across men and women.

³It is easy to see that in such a model, a man i who is married in equilibrium is matched with a woman in the group that values his attractiveness most, and he receives a transfer $\tau_i = \max_{y \in \mathcal{Y}} \varepsilon_{iy}$.

Choo and Siow proved that under Assumptions 1-3, there is a simple equilibrium relationship between group preferences, as defined by α and γ , and equilibrium marriage patterns. To state their result, we 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 . Then:

Theorem 1 (Choo and Siow) *Under Assumptions 1-3, in equilibrium, for all $x, y \geq 1$*

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

Therefore marriage patterns μ directly identify the gains to marriage Φ .

1.5 Unobserved Heterogeneity

One of our goals in this paper is to extend identification to quite general distributions of unobserved heterogeneity. We continue to assume separability (Assumption 2) and a large market (Assumption 1); but we allow for quite general distributions of unobserved heterogeneity:

Assumption 4 (Distribution of Unobserved Variation in Surplus)

- 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) These draws are independent across men and women.*

Assumption 4 clearly is a substantial generalization with respect to Choo and Siow (2006). It relaxes Assumption 3 in three important ways: it allows for different families of distributions, with any form of heteroskedasticity, and with any pattern of correlation across partner groups.

2 Social Surplus, Utilities, and Identification

We derive most of our results by considering the “primal” problem, which maximizes the total joint surplus. As Choo and Siow (2006) remind us (p. 177): “A well-known property of transferable utility models of the marriage market is that they maximize the sum of marital output in the society”. This is true when marital output is defined as it is evaluated by the participants: the market equilibrium in fact maximizes $\sum_{i,j} \tilde{\mu}_{ij} \tilde{\Phi}_{ij}$ over the set of feasible matchings ($\tilde{\mu}_{ij}$). A very naive evaluation of the sum of marital output, computed from the groups of partners only, would be

$$\sum_{xy} \mu_{xy} \Phi_{xy}, \quad (2.1)$$

but this is clearly misleading. Realized matches by nature have a value of the unobserved marital surplus ($\varepsilon_{iy} + \eta_{xj}$) that is more favorable than an unconditional draw; and as a consequence, the equilibrium marriage patterns (μ) do not maximize $\sum_{xy} \mu_{xy} \Phi_{xy}$ over \mathcal{M} . In order to find the expression of the value function that (μ) maximizes, we need to account for terms that reflect the conditional expectation of the unobserved parts of the surplus, given a match on observable groups.

2.1 Separability and Discrete Choice

We first prove that separability (Assumption 2) reduces the choice of partner to a one-sided discrete choice problem. To see this, note that by standard results in the literature (Shapley and Shubik 1972), the equilibrium utilities solve the system of functional equations

$$\tilde{u}_i = \max_j \left(\tilde{\Phi}_{ij} - \tilde{v}_j \right) \quad \text{and} \quad \tilde{v}_j = \max_i \left(\tilde{\Phi}_{ij} - \tilde{u}_i \right),$$

where the maximization includes the option of singlehood.

Focus on the first one. It states that the utility man i gets in equilibrium trades off the surplus his match with woman j creates and the share of the joint surplus he has to give her, which is given by her own equilibrium utility. There may be several pairs of vectors (u, v) that solve this system; and for each of them, any feasible matching whose support is

contained in the set of maximizers is an optimal matching. Now use Assumption 2: for a man i in group x ,

$$\tilde{\Phi}_{ij} = \Phi_{xyj} + \varepsilon_{iyj} + \eta_{xj}$$

so that

$$\tilde{u}_i = \max_j \left(\tilde{\Phi}_{ij} - \tilde{v}_j \right) = \max_y \max_{j|y_j=y} \left(\tilde{\Phi}_{ij} - \tilde{v}_j \right)$$

can be rewritten as

$$\tilde{u}_i = \max_y \left(\Phi_{xy} + \varepsilon_{iy} - \min_{j|y_j=y} (\tilde{v}_j - \eta_{xj}) \right).$$

Denoting

$$V_{xy} = \min_{j|y_j=y} (\tilde{v}_j - \eta_{xj})$$

and $U_{xy} = \Phi_{xy} - V_{xy}$, it follows that:

Proposition 1 (*Splitting the Surplus*)

There exist two vectors U_{xy} and V_{xy} such that $\Phi_{xy} = U_{xy} + V_{xy}$ and in equilibrium:

(i) Man i in group x achieves utility

$$\tilde{u}_i = \max_{y \in \mathcal{Y}_0} (U_{xy} + \varepsilon_{iy})$$

and he matches with some woman whose group y achieves the maximum;

(ii) Woman j in group y achieves utility

$$\tilde{v}_j = \max_{x \in \mathcal{X}_0} (V_{xy} + \eta_{xj})$$

and she matches with some man whose group x achieves the maximum.

This result also appears in Chiappori, Salanié, and Weiss (2012), with a different proof. It reduces the two-sided matching problem to a series of one-sided discrete choice problems that are only linked through the adding-up formula $U_{xy} + V_{xy} = \Phi_{xy}$. Men of a given group x match women of different groups, since they have idiosyncratic ε_{iy} shocks. But as a consequence of the separability assumption, if a man of group x matches with a woman of group y , then he would be equally well-off with any other woman of this group. The vectors

U_{xy} and V_{xy} depend on all of the primitives of the model (the vector Φ_{xy} , the distributions of the utility shocks ε and η , and the number of groups n and m .) They are only a useful construct, and they should not be interpreted as utilities. As we will see in Section 2.3, there are at least three relevant definitions of utility, and U and V do not measure any of them.

2.2 Social surplus

In the following, for any (A_{xy}) we denote $\mathbf{A}_{\mathbf{x}} = (A_{x0}, A_{x1}, \dots, A_{x|\mathcal{Y}|})$ and $\mathbf{A}_{\mathbf{y}} = (A_{0y}, A_{1y}, \dots, A_{|\mathcal{X}|y})$. Consider a randomly chosen man in group x . His expected utility (conditional to belonging to this group) is

$$G_x(\mathbf{U}_{\mathbf{x}\cdot}) = \mathbb{E}_{\mathbf{P}_x} \left(\max_{y \in \mathcal{Y}_0} (U_{xy} + \varepsilon_y) \right),$$

where the expectation is taken over the random vector $(\varepsilon_0, \dots, \varepsilon_{|\mathcal{Y}|}) \sim \mathbf{P}_x$. Similarly, a randomly chosen woman of group y expects to get utility

$$H_y(\mathbf{V}_{\cdot\mathbf{y}}) = \mathbb{E}_{\mathbf{Q}_y} \left(\max_{x \in \mathcal{X}_0} (V_{xy} + \eta_x) \right).$$

The social surplus is simply the sum of the expected utilities of all groups of men and women:

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

Of course the vectors \mathbf{U} and \mathbf{V} are unobserved; we must find a way to write them in terms of the matching patterns μ .

For simplicity, we will focus on the case when all groups of matches are possible at the optimal matching: for every $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, μ_{xy} is positive. A simple way to guarantee this is to assume

Assumption 5 (Full support) *The distributions \mathbf{P}_x and \mathbf{Q}_y all have full support.*

Assumption 5 can be relaxed in the obvious way: all that matters is that the supports of the distributions are wide enough relative to the variation in Φ .

We first give a heuristic derivation⁴ of our main result, Theorem 2 below. Focusing on the function G_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

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

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

$$\frac{\partial G_x}{\partial U_{xy}}(\mathbf{U}_{\mathbf{x}\cdot}) = \frac{\mu_{xy}}{n_x}. \quad (2.2)$$

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

$$G_x^*(\boldsymbol{\mu}_{\mathbf{x}\cdot}) = \max_{\tilde{\mathbf{U}}_{\mathbf{x}\cdot} = (\tilde{U}_{x0}, \dots, \tilde{U}_{x|\mathcal{Y}|})} \left(\sum_{y \in \mathcal{Y}_0} \mu_{xy} \tilde{U}_{xy} - \left(\sum_{y \in \mathcal{Y}_0} \mu_{xy} \right) G_x(\tilde{\mathbf{U}}_{\mathbf{x}\cdot}) \right) \quad (2.3)$$

whenever $\sum_{y \in \mathcal{Y}_0} \mu_{xy} = n_x$, $G_x^*(\boldsymbol{\mu}_{\mathbf{x}\cdot}) = +\infty$ otherwise. This is just the Legendre-Fenchel transform of G_x rescaled by the factor $\sum_{y \in \mathcal{Y}_0} \mu_{xy}$. 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^*

$$\frac{\partial G_x^*}{\partial \mu_{xy}}(\boldsymbol{\mu}_{\mathbf{x}\cdot}) = U_{xy} - G_x(\mathbf{U}_{\mathbf{x}\cdot}) \quad (2.4)$$

$$\frac{\partial G_x^*}{\partial \mu_{x0}}(\boldsymbol{\mu}_{\mathbf{x}\cdot}) = -G_x(\mathbf{U}_{\mathbf{x}\cdot}) \quad (2.5)$$

where the second equality comes from the fact that we have normalized U_{x0} at zero. As a consequence, for any $y \in \mathcal{Y}$

$$U_{xy} = \frac{\partial G_x^*}{\partial \mu_{xy}}(\boldsymbol{\mu}_{\mathbf{x}\cdot}) - \frac{\partial G_x^*}{\partial \mu_{x0}}(\boldsymbol{\mu}_{\mathbf{x}\cdot})$$

is identified from $\boldsymbol{\mu}_{\mathbf{x}\cdot}$, the observed matching patterns of men of group x . Going back to (2.3), convex duality implies that

$$n_x G_x(\mathbf{U}_{\mathbf{x}\cdot}) = \left(\sum_{y \in \mathcal{Y}_0} \mu_{xy} \right) G_x(\mathbf{U}_{\mathbf{x}\cdot}) = \sum_{y \in \mathcal{Y}_0} \mu_{xy} U_{xy} - G_x^*(\boldsymbol{\mu}_{\mathbf{x}\cdot}). \quad (2.6)$$

⁴Appendix A gives rigorous proofs of all of our results.

Defining $H_y^*(\boldsymbol{\mu}_{\cdot y})$ similarly for women of group y and using $U_{xy} + V_{xy} = \Phi_{xy}$, we get the value of the total surplus:

$$\sum_{x \in \mathcal{X}} n_x G_x(\mathbf{U}_{x\cdot}) + \sum_{y \in \mathcal{Y}} m_y H_y(\mathbf{V}_{\cdot y}) = \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy} - \sum_{x \in \mathcal{X}} G_x^*(\boldsymbol{\mu}_{x\cdot}) - \sum_{y \in \mathcal{Y}} H_y^*(\boldsymbol{\mu}_{\cdot y}).$$

This can readily be computed from the data: it only requires specifying the distributions \mathbf{P}_x and \mathbf{Q}_y and observing the margins n and m and the matching patterns μ . Finally, the expected equilibrium utility of the average man in group x is $G_x(\mathbf{U}_{x\cdot})$, which equals $-\frac{\partial G_x^*}{\partial \mu_{x0}}(\boldsymbol{\mu}_{x\cdot})$ by (2.5). Once again this is easy to compute from the data, given a specification of the error distributions.

Let us now turn to a formal statement of the main welfare and identification result, which is proved in Appendix A. Note that since the functions G_x^* and H_y^* are convex, they are differentiable almost everywhere—and very mild assumptions on the distributions \mathbf{P}_x and \mathbf{Q}_y would make them differentiable everywhere. We will use their derivatives in stating our results; they should be replaced with subgradients at hypothetical points of non-differentiability.

Theorem 2 (Social Surplus and identification)

(i) Under assumptions 1, 2, 4, and 5, the optimal matching μ maximizes the social gain

$$\mathcal{W}(\mu) = \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy} + \mathcal{E}(\mu) \tag{2.7}$$

over all feasible matchings, where \mathcal{E} is defined by

$$\mathcal{E}(\mu) = - \sum_{x \in \mathcal{X}} G_x^*(\boldsymbol{\mu}_{x\cdot}) - \sum_{y \in \mathcal{Y}} H_y^*(\boldsymbol{\mu}_{\cdot y}).$$

In equilibrium, for any $x \in \mathcal{X}, y \in \mathcal{Y}$

$$\Phi_{xy} = \frac{\partial \mathcal{E}}{\partial \mu_{x0}}(\mu) - \frac{\partial \mathcal{E}}{\partial \mu_{0y}}(\mu) - \frac{\partial \mathcal{E}}{\partial \mu_{xy}}(\mu), \tag{2.8}$$

that is

$$\Phi_{xy} = \frac{\partial G_x^*}{\partial \mu_{xy}}(\boldsymbol{\mu}_{\mathbf{x}\cdot}) + \frac{\partial H_y^*}{\partial \mu_{xy}}(\boldsymbol{\mu}_{\cdot\mathbf{y}}) - \frac{\partial G_x^*}{\partial \mu_{x0}}(\boldsymbol{\mu}_{\mathbf{x}\cdot}) - \frac{\partial H_y^*}{\partial \mu_{0y}}(\boldsymbol{\mu}_{\cdot\mathbf{y}}).$$

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

$$U_{xy} + \varepsilon_{iy} = \max_{z \in \mathcal{Y}_0} (U_{xz} + \varepsilon_{iz})$$

where

$$U_{xz} = \frac{\partial G_x^*}{\partial \mu_{xz}}(\boldsymbol{\mu}_{\mathbf{x}\cdot}) - \frac{\partial G_x^*}{\partial \mu_{x0}}(\boldsymbol{\mu}_{\mathbf{x}\cdot})$$

can also be computed by solving the system of equations

$$\frac{\partial G_x}{\partial U_{xz}}(\mathbf{U}_{\mathbf{x}\cdot}) = \frac{\mu_{xz}}{n_x} \text{ for } z \in \mathcal{Y}_0$$

given the normalization $U_{x0} = 0$.

iii) The average expected utility of the men of group x is

$$u_x = G_x(\mathbf{U}_{\mathbf{x}\cdot}) = -\frac{\partial G_x^*}{\partial \mu_{x0}}(\boldsymbol{\mu}_{\mathbf{x}\cdot}). \quad (2.9)$$

(iv) Parts (ii) and (iii) transpose to women with the obvious changes.

(v) Denote the systematic parts of pre-transfer utilities (α, γ) and of transfers τ as in Section 1. Then

$$U_{xy} = \alpha_{xy} + \tau_{xy} \text{ and } V_{xy} = \gamma_{xy} - \tau_{xy}.$$

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

In fact, the functions G_x^* and H_y^* can also be evaluated by solving an associated matching problem. Take men of group x for instance. Denote $\mu_{Y|X=x}$ the probability distribution giving mass $\mu_{xy} / \left(\sum_{y' \in \mathcal{Y}_0} \mu_{xy'} \right)$ to option $y \in \mathcal{Y}_0$. Define \mathcal{M}_x as the set of probability distributions of the random joint vector $(\boldsymbol{\varepsilon}_{\mathbf{i}\cdot}, Y)$, where:

- $\boldsymbol{\varepsilon}_{\mathbf{i}\cdot} = (\varepsilon_{iy})_{y \in \mathcal{Y}_0}$ a random vector taking values in $\mathbb{R}^{|\mathcal{Y}_0|}$ such that $\boldsymbol{\varepsilon}_{\mathbf{i}\cdot} \sim \mathbf{P}_x$
- Y is a random variable taking values in \mathcal{Y}_0 such that $Y \sim \mu_{Y|X=x}$.

Then:

Proposition 2 *The function G_x^* is $(-n_x)$ times the value of matching optimally the distribution \mathbf{P}_x of random vector $\boldsymbol{\varepsilon}_i$. and the distribution $\mu_{Y|X=x}$ when the surplus from a match is the unobserved heterogeneity $(\boldsymbol{\varepsilon}_i, y) \rightarrow \varepsilon_{iy}$:*

$$\begin{aligned} G_x^*(\boldsymbol{\mu}_{\mathbf{x}}) &= -n_x \max_{\pi \in \mathcal{M}_x} \mathbb{E}_\pi [\varepsilon_{iY}] \text{ if } \sum_{y \in \mathcal{Y}_0} \mu_{xy} = n_x \\ &= +\infty \text{ else.} \end{aligned}$$

Proof. See Appendix A. ■

As a result, all of the quantities in Theorem 2 can be computed by solving simple linear programming problems.

2.3 Discussion

The right-hand side of equation (2.7) 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 dissimilar partners.

To interpret the formula, start with the case when unobserved heterogeneity is dwarfed by variation due to observable characteristics: $\tilde{\Phi}_{ij} \simeq \Phi_{xy}$ if $x_i = x$ and $y_j = y$. Then we know that the observed matching μ must maximize the value in (2.1); 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. If on the other hand 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. Information theory tells us that entropy is a natural measure of statistical disorder; and as we will see in Example 1, in the simple case analyzed by Choo and Siow (2006) the function \mathcal{E} is just the usual notion of entropy. For this reason, we call it the *generalized entropy* of the matching.

In the intermediate case in which some of the variation in marital surplus is driven by group characteristics (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 group characteristics (as in (2.1)) and matching on unobserved characteristics, as measured by the generalized entropy terms in $\mathcal{E}(\mu)$.

The data directly yield the numbers of participants of each group n_x and m_y and their matching patterns μ ; and the specification of the distribution of unobserved heterogeneity determines the functions G_x^* and H_y^* . Part (i) of the result shows that the social surplus can be identified by computing these functions. As we show in section 3, this can be done in closed form in many usual cases. Parts (ii) and (iii) of Theorem 2 show that participant utilities are also identified and easily computed. Note that there are three measures of utilities:

- ex ante utility u_x is the expected utility of a man, conditional on his being in group x . Part (iii) gives a very simple formula to compute it;
- ex interim utility, if we also condition on this man marrying a woman of group y , is

$$\mathbb{E}[U_{xy} + \varepsilon_{iy} | U_{xy} + \varepsilon_{iy} \geq U_{xz} + \varepsilon_{iz} \text{ for all } z \in \mathcal{Y}];$$

this can be computed since the U_{xz} 's are identified from part (ii), although it may require simulation for general distributions;

- ex post utility $U_{xy} + \varepsilon_{iy}$ for these men, whose distribution can also be simulated.

In the special case studied by Choo and Siow, ex post utility is distributed as type I extreme value with mean $(-\log \frac{\mu_{x0}}{n_x})$, which is the common value u_x of ex ante and ex interim utility; but the three definitions give different results in general (de Palma and Kilani 2007).

3 Examples

While Proposition 2 provides a general way of computing surplus and utilities, they can often be derived in closed form. Appendix B gives the resulting formulæ for McFadden's Generalized Extreme Value (GEV) framework. This comprises most specifications used in discrete choice studies. The first one is, obviously, Choo and Siow's model. This obtains when the \mathbf{P}_x and \mathbf{Q}_y distributions are all standard type I extreme value and i.i.d.:

Example 1 (Choo and Siow) *Under assumptions 1, 2, and 3 (which implies 5), for feasible matchings the function \mathcal{E} is simply (see Appendix C.1 for details)*

$$\mathcal{E}(\mu) = - \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}_0}} \mu_{xy} \log \frac{\mu_{xy}}{n_x} - \sum_{\substack{x \in \mathcal{X}^0 \\ y \in \mathcal{Y}}} \mu_{xy} \log \frac{\mu_{xy}}{m_y},$$

so that

$$\mathcal{W}(\mu) = \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy} - \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}_0}} \mu_{xy} \log \frac{\mu_{xy}}{n_x} - \sum_{\substack{x \in \mathcal{X}^0 \\ y \in \mathcal{Y}}} \mu_{xy} \log \frac{\mu_{xy}}{m_y}. \quad (3.1)$$

Moreover, surplus and matching patterns are linked by

$$\Phi_{xy} = 2 \log \mu_{xy} - \log \mu_{x0} - \log \mu_{0y}, \quad (3.2)$$

which is Choo and Siow's result (Theorem 1 above.)

Note that as announced after Theorem 2, the generalized entropy \mathcal{E} boils down here to the usual definition of entropy. As a more complex example of a GEV distribution, consider a nested logit.

Example 2 (A two-level nested logit) Suppose for instance that men of group x are concerned about the social group of their partner and her education. There are G social groups and there are L levels of education, so that a group $x = (g, e)$. We can allow for correlated preferences by modeling this as a nested logit in which educations are nested within social groups. Consider men of group x ; let $\sigma_{x,g'}$ measure the dispersion in the surplus they generate with partners of different education levels within social group g' . We show in Appendix C.2 that

$$U_{x,g'e'} = \log \frac{\mu_{x,g'}}{\mu_{x,0}} + \sigma_{x,g'} \log \frac{\mu_{x,g'e'}}{\mu_{x,g'}}$$

so that the social surplus is identified as

$$\Phi_{ge,g'e'} = \log \frac{\mu_{ge,g'} \mu_{g,g'e'}}{\mu_{ge,0} \mu_{0,g'e'}} + \sigma_{ge,g'} \log \frac{\mu_{ge,g'e'}}{\mu_{ge,g'}} + \sigma_{g,g'e'} \log \frac{\mu_{ge,g'e'}}{\mu_{g,g'e'}}.$$

While the GEV framework is convenient, it is common in the applied literature to allow, say, for random variation in preferences over characteristics of products. Our last example incorporates this idea. We develop the computations in Appendix C.3 on a slightly more general case.

Example 3 (A “mixed” model) Assume that for man i in group x , $\varepsilon_{iy} = \varepsilon_i \phi_x(y)$, where $\phi_x(y)$ is an index of the observable characteristics of women that is common to all men in the same group. The random term ε_i , which we take to be uniformly distributed over $[0, 1]$ for simplicity, denotes unobserved variation in how men in group x “weigh” this index.

For $y \in \mathcal{Y}_0$, denote

$$F_{Y|x}(y) = \Pr(\phi(Y) \leq \phi(y) | X = x), \quad (3.3)$$

where Y has the distribution $\mu_{Y|X=x}$, which is observed in the data. We show in Appendix C.3 that

$$G_x^*(\boldsymbol{\mu}_x) = - \sum_{y \in \mathcal{Y}_0} \phi_x(y) \left(F_{Y|x}(y) - \frac{\mu_{xy}}{2n_x} \right) \mu_{xy},$$

and that the expected utility of men of group x is

$$u_x = \mathbb{E}[\max(\phi_x(Y), \phi_x(0))],$$

while the first term of the surplus of men of group x matching with women of group y is

$$U_{xy} = -\mathbb{E}[\max(\phi_x(Y), \phi_x(y))] + \mathbb{E}[\max(\phi_x(Y), \phi_x(0))].$$

4 Parametric Inference

Theorem 2 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, 4, and 5 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 4 can then be used to infer information about these distributions.

In the present paper, we are assuming that a single market is being observed. While the formula in Theorem 2(i) gives a straightforward nonparametric estimator of the systematic surplus function Φ , with multiple surplus-relevant observable groups it will be very unreliable. Even our toy education/income example of Section 1 already has $4n_R^2$ cells; and realistic applications will require many more. 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:

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

Let μ^λ be the optimal matching. We will show in Section 5 how it can be computed very efficiently. For now we focus on statistical inference on λ . 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

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

and a similar expression for each woman of group y . Under Assumptions 1, 2, 4, and 5, 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:

$$\begin{aligned} \log L(\lambda) &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}_0} \hat{\mu}_{xy} \log \frac{\mu_{xy}^\lambda}{\hat{n}_x} + \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}_0} \hat{\mu}_{xy} \log \frac{\mu_{xy}^\lambda}{\hat{m}_y} \\ &= 2 \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \hat{\mu}_{xy} \log \frac{\mu_{xy}^\lambda}{\sqrt{\hat{n}_x \hat{m}_y}} + \sum_{x \in \mathcal{X}} \hat{\mu}_{x0} \log \frac{\mu_{x0}^\lambda}{\hat{n}_x} + \sum_{y \in \mathcal{Y}} \hat{\mu}_{0y} \log \frac{\mu_{0y}^\lambda}{\hat{m}_y}. \end{aligned}$$

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

$$\frac{\partial \log L}{\partial \lambda}(\hat{\lambda}) = 0$$

is consistent, asymptotically normal and asymptotically efficient under the usual set of assumptions. As we will see in Section 6, it becomes a very simple Moment Matching estimator in a special but useful class of models.

5 Computation

Maximizing the conditional likelihood requires computing the optimal matching μ^λ for a large number of values of λ . But the optimal matching will be a large-dimensional object

in realistic applications; and it is itself the maximizer of \mathcal{W} in (2.7). It is therefore crucial to be able to compute μ^λ efficiently. We show here how the Iterative Proportional Fitting Procedure (IPFP) provides a solution to this problem.

Take the Choo and Siow (2006) model of Example 1 for instance. Fix a value of λ and dropping it from the notation: let the joint surplus function be Φ , with optimal matching μ . Formula (3.2) can be rewritten as

$$\mu_{xy} = \exp\left(\frac{\Phi_{xy}}{2}\right) \sqrt{\mu_{x0}\mu_{0y}}. \quad (5.1)$$

In principle we could just plug this into the feasibility constraints $\sum_y \mu_{xy} + \mu_{x0} = \hat{n}_x$ and $\sum_x \mu_{xy} + \mu_{0y} = \hat{m}_y$ and solve for the numbers of singles μ_{x0} and μ_{0y} . Unfortunately, the resulting equations are still high-dimensional and highly nonlinear, which makes them hard to handle. Even proving the uniqueness of the solution to this system of equations is a hard problem. This is done in Decker, Lieb, McCann, and Stephens (2012), who also derive some comparative statics results on the variation of μ with Φ , \hat{n} and \hat{m} .

On the other hand, to find a feasible solution of (3.2), we could start from an infeasible solution and project it somehow on the set of feasible matchings $\mathcal{M}(\hat{n}, \hat{m})$. Moreover, IPFP was precisely designed to find projections on intersecting sets of constraints, by projecting iteratively on each constraint. It is used for instance as the RAS method to impute missing values in data.

Now if we start from any infeasible solution to (3.2) and we use for instance the Euclidean distance to project on the feasible matchings, we would be very lucky to end up at the optimal matching. To get the algorithm there we will need to carefully choose both the infeasible solution and the projection distance. As it turns out, we can do it in the general case studied in this paper. We can formally describe the algorithm as follows.

Algorithm 1

1. Set $k = 0$, and take any $\mu^{(0)}$ that solves (2.8):

$$\Phi_{xy}^\lambda = \frac{\partial \mathcal{E}^\lambda}{\partial \mu_{x0}}(\mu^{(0)}) + \frac{\partial \mathcal{E}^\lambda}{\partial \mu_{0y}}(\mu^{(0)}) - \frac{\partial \mathcal{E}^\lambda}{\partial \mu_{xy}}(\mu^{(0)})$$

and that has the same total number of individuals as in the sample:

$$2 \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \mu_{xy}^{(0)} + \sum_{x \in \mathcal{X}} \mu_{x0}^{(0)} + \sum_{y \in \mathcal{Y}} \mu_{0y}^{(0)} = N_{ind}$$

where $N_{ind} = \sum_x \hat{n}_x + \sum_y \hat{m}_y$ is the number of individuals in the sample.

2. Compute $\mu^{(2k+1)}$ by

$$\frac{\partial \mathcal{E}}{\partial \mu_{xy}}(\mu^{(2k+1)}) = \frac{\partial \mathcal{E}}{\partial \mu_{xy}}(\mu^{(2k)}) - a_x \text{ for } y \in \mathcal{Y} \text{ and } \frac{\partial \mathcal{E}}{\partial \mu_{0y}}(\mu^{(2k+1)}) = \frac{\partial \mathcal{E}}{\partial \mu_{0y}}(\mu^{(2k)})$$

where (a_x) is such that $\sum_{y \in \mathcal{Y}_0} \mu_{xy}^{(2k+1)} = \hat{n}_x$ for every $x \in \mathcal{X}$.

3. Compute $\mu^{(2k+2)}$ by

$$\frac{\partial \mathcal{E}}{\partial \mu_{xy}}(\mu^{(2k+2)}) = \frac{\partial \mathcal{E}}{\partial \mu_{xy}}(\mu^{(2k+1)}) - b_y \text{ for } x \in \mathcal{X} \text{ and } \frac{\partial \mathcal{E}}{\partial \mu_{x0}}(\mu^{(2k+2)}) = \frac{\partial \mathcal{E}}{\partial \mu_{x0}}(\mu^{(2k+1)})$$

where (b_y) is such that $\sum_{x \in \mathcal{X}_0} \mu_{xy}^{(2k+2)} = \hat{m}_y$ for every $y \in \mathcal{Y}$.

4. If $\mu^{(2k+2)}$ is close enough to $\mu^{(2k)}$, then go to step 5; otherwise add one to k and go to step 2.

5. Take $\mu^\lambda = \mu^{(2k+2)}$ to be the optimal matching.

We prove in Appendix A that

Theorem 3 *The algorithm converges to the solution μ of (2.7).*

To illustrate, take the Choo and Siow example again. Choose a constant C , define $\mu^{(0)}$ by $\mu_{x0}^{(0)} = \mu_{0y}^{(0)} = C$ for all $x \in \mathcal{X}, y \in \mathcal{Y}$, and apply (5.1) to get $\mu_{xy}^{(0)}$. By construction $\mu^{(0)}$ solves (2.8), and since it is proportional to C we can choose C so as to have the same

number of individuals as in the sample. Then $\mu^{(0)}$ is suitable for step 1. The updating step 3 gives for $x, y \neq 0$

$$\mu_{xy}^{(2k+1)} = \mu_{xy}^{(2k)} \exp\left(\frac{a_x}{2}\right) \text{ and } \mu_{x0}^{(2k+1)} = \mu_{x0}^{(2k)} \exp(a_x),$$

along with $\mu_{0y}^{(2k+1)} = \mu_{0y}^{(2k)}$; so that we need to solve in a_x the equation

$$\sum_y \mu_{xy}^{(2k)} \exp\left(\frac{a_x}{2}\right) + \mu_{x0}^{(2k)} \exp(a_x) = \hat{n}_x.$$

This is a quadratic equation in only one unknown, $\exp(a_x/2)$; as such it can be solved in closed form.

The convergence is extremely fast. We tested it on a simulation in which we let the number of categories $|\mathcal{X}| = |\mathcal{Y}|$ increase from 100 to 1,000. For each of these ten cases, we draw 50 samples, with the n_x and m_y drawn uniformly in $\{1, \dots, 100\}$; and for each (x, y) match we draw Φ_{xy} from $N(0, 1)$. To have a basis for comparison, we also ran two nonlinear equation solvers on the system of $(|\mathcal{X}| + |\mathcal{Y}|)$ equations

$$t_x^2 + t_x \left(\sum_y \exp(\Phi_{xy}/2) T_y \right) = n_x$$

and

$$T_y^2 + T_y \left(\sum_x \exp(\Phi_{xy}/2) t_x \right) = m_y,$$

which characterizes the optimal matching with $\mu_{xy} = \exp(\Phi_{xy}/2) \sqrt{\mu_{x0} \mu_{0y}}$, $\mu_{x0} = t_x^2$, and $\mu_{0y} = T_y^2$ (see the IPFP formulæ above with $a_x = b_y = 0$ at the fixed point, or Decker, Lieb, McCann, and Stephens (2012).)

To solve this system, we used both Minpack and Knitro. Minpack is probably the most-used solver in scientific applications, and underlies many statistical and numerical packages. Knitro⁵ is a constrained optimization software; but minimizing the function zero under constraints that correspond to the equations one wants to solve has become popular recently.

⁵See Byrd, Nocedal, and Waltz (2006).

For all three methods, we used C/C^{++} programs, run on a single processor of a Mac desktop. We set the convergence criterion for the three methods as a relative estimated error of 10^{-6} . This is not as straightforward as one would like: both Knitro and Minpack rescale the problem before solving it, while we did not attempt to do it for IPFP. Still, varying the tolerance within reasonable bounds hardly changes the results, which we present in Figure 1. Each panel gives the distribution of CPU times over 50 samples (20 for Knitro) for the ten experiments, in the form of a Tukey box-and-whiskers graph⁶.

The performance of IPFP stands out clearly—note the different vertical scales. While IPFP has more variability than Minpack and Knitro (perhaps because we did not rescale the problem beforehand), even the slowest convergence times for each problem size are at least three times smaller than the fastest sample under Minpack, and fifteen times smaller than the fastest time with Knitro. This is all the more remarkable that we fed the code for the Jacobian of the system of equations into Minpack, and for both the Jacobian and the Hessian into Knitro.

6 The Linear Choo and Siow Model

Assume that the analyst has chosen K *basis surplus vectors*

$$\Phi_{xy}^1, \dots, \Phi_{xy}^K$$

which are linearly independent: no linear combination of these vectors is identically equal to zero.

The analyst then specifies the systematic surplus function Φ_{xy}^λ as a linear combination of these basis surplus vectors, with unknown weights $\lambda \in \mathbb{R}^K$:

$$\Phi_{xy}^\lambda = \sum_{k=1}^K \lambda_k \Phi_{xy}^k \tag{6.1}$$

⁶The box goes from the first to the third quartile; the horizontal bar is at the median; the lower (resp. upper) whisker is at the first (resp. third) quartile minus (resp. plus) 1.5 times the interquartile range, and the circles plot all points beyond that.

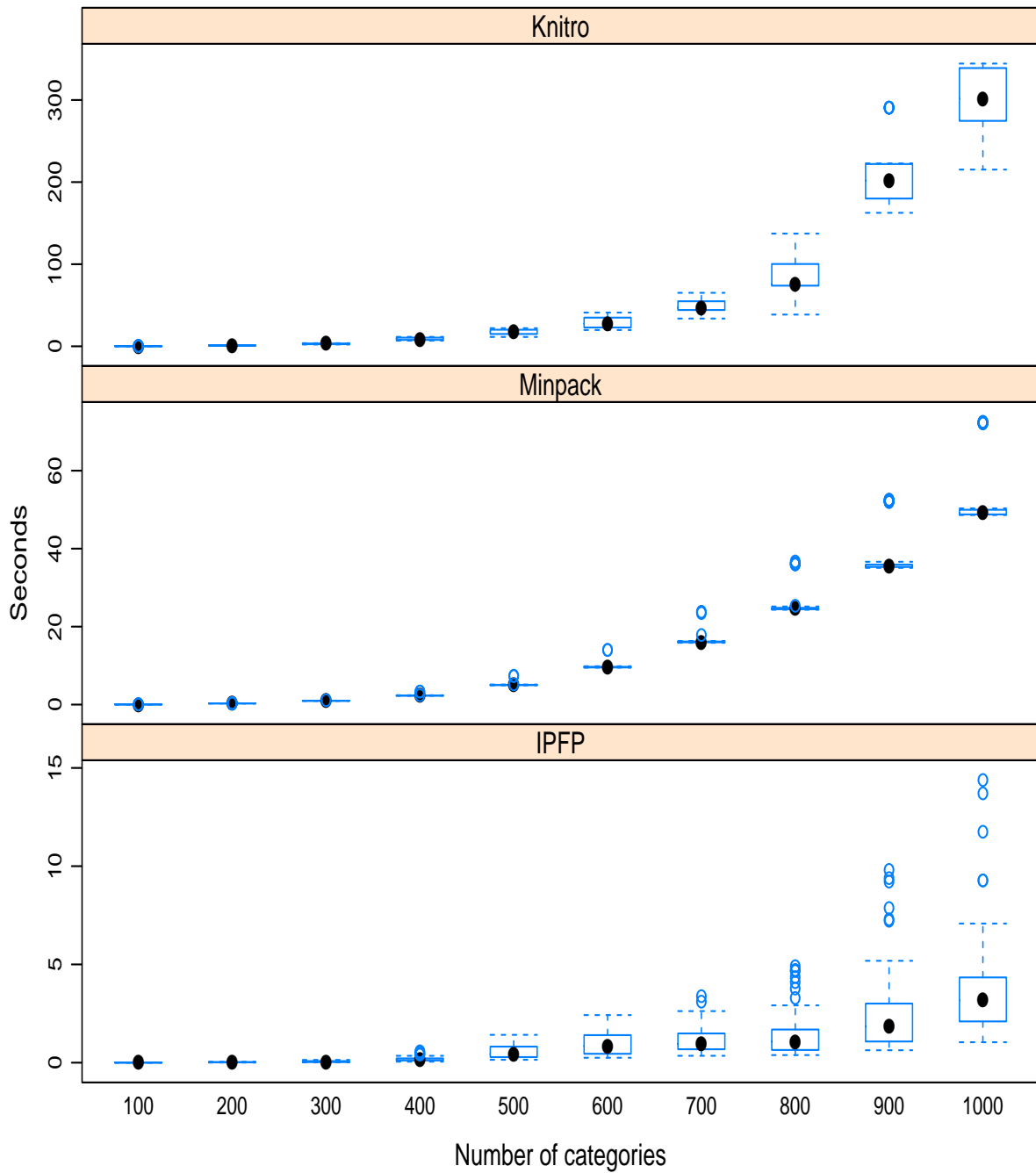


Figure 1: Solving for the optimal matching

where the sign of each λ_k is unrestricted. We call this specification the “linear Choo and Siow model” because the surplus depends linearly on the parameters. Quite obviously, if the set of basis surplus vectors is large enough, this specification covers the full set without restriction; however, parsimony is often valuable in applications.

To return to the education/income example, we could for instance assume that a match between man i and woman j creates a surplus that depends on whether partners are matched on both education and income dimensions. The corresponding specification would have basis functions like $\mathbf{1}(E_i = E_j = e)$ and $\mathbf{1}(R_i = R_j = r)$, along with “one-sided” basis functions to account for different probabilities of marrying: $\mathbf{1}(R_i = r, E_i = e)$ and $\mathbf{1}(R_j = r, E_j = e)$. This specification only has $(5n_R + 2)$ parameters, while an unrestricted specification⁷ would have $4n_R^2$. With more, multi-valued criteria the reduction in dimensionality would be much larger. It is clear that the relative importance of the λ 's reflects the relative importance of the criteria. They indicate how large the systematic preference for complementarity of incomes of partners is relative to the preference for complementarity in educations.

Suppose that the unobserved heterogeneity satisfies Assumption 3. Then the linear structure underlying model (6.1) makes it very easy to analyze optimal matchings. All inference can be based on only $(K + 1)$ numbers: the entropy \mathcal{E} will be used to test the specification, and K *comoments* will be used to estimate λ . Moreover, computing the Conditional Maximum Likelihood (CML) estimator of λ becomes very easy.

For any feasible matching μ , we define the *comoments*

$$C^k(\mu) = \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy}^k.$$

We prove in Appendix A that

Theorem 4 1. *The Maximum Likelihood Estimator is characterized by either of the two equivalent properties:*

⁷Such an unrestricted specification would for instance allow the effect of matching a man in income class 3 with a woman in income class 2 to also depend on both of their education levels.

(a) $\hat{\lambda}$ maximizes over Λ the concave function

$$\sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \hat{\mu}_{xy} \Phi_{xy}^{\lambda} - \mathcal{W}(\lambda) \quad (6.2)$$

(b) $\hat{\lambda}$ solves

$$C^k(\hat{\mu}) = C^k(\mu^{\hat{\lambda}}) \text{ for all } k = 1, \dots, K.$$

2. Moreover, the entropy of the optimal matching for the CML cannot be smaller than the entropy of the observed matching:

$$\mathcal{E}(\hat{\mu}) \leq \mathcal{E}(\mu^{\hat{\lambda}})$$

and the two entropies are equal if and only if $\mu^{\hat{\lambda}} = \hat{\mu}$, that is, if and only if the surplus function in (6.1) rationalizes the observed matching for the CML $\hat{\lambda}$.

Therefore under the assumptions that drive the Choo and Siow model, if the surplus function is linear in the parameters the Conditional Maximum Likelihood maximizes a very well-behaved (globally concave) function; and it matches the observed comoments to those that are predicted by the model. Entropy is a sufficient statistic to test the specification; and if we cannot reject that the model is well-specified (so that the true data-generating process is of the form (6.1) for the set of basis functions chosen by the analyst), then the K comoments form a sufficient statistic to estimate λ .

Our approach to inference has a simple geometric interpretation. Consider the set of comoments associated to every feasible matching

$$\mathcal{F} = \left\{ (C^1, \dots, C^K) : C^k = \sum_{xy} \mu_{xy} \Phi_{xy}^k, \mu \in \mathcal{M}(\hat{n}, \hat{n}) \right\}$$

This is a convex polyhedron, which we call the *covariogram*; and if the model is well-specified the covariogram must contain the observed matching $\hat{\mu}$. For any value of the parameter vector λ , the optimal matching μ^{λ} generates a vector of comoments C^{λ} that

belongs to the covariogram; and it also has an entropy $\mathcal{E}^\lambda \equiv \mathcal{E}(\mu^\lambda)$. We already know that this model is just-identified from the comoments: the mapping $\lambda \rightarrow C^\lambda$ is invertible on the covariogram. Denote $\lambda(C)$ its inverse. The corresponding optimal matching has entropy

$$\mathcal{E}_r(C) = \mathcal{E}^{\lambda(C)}.$$

The level sets of $\mathcal{E}_r(\cdot)$ are the isoentropy curves in the covariogram; they are represented on Figure 2. The figure assumes $K = 2$ dimensions; then λ can be represented in polar coordinates as $\lambda = r \exp(it)$. For $r = 0$, the model is uninformative and entropy is highest; the matching is random and generates comoments C_0 . At the other extreme, the boundary ∂F of the covariogram corresponds to $r = \infty$. Then there is no unobserved heterogeneity and generically over t , the comoments generated by λ must belong to a finite set of vertices, so that λ is only set-identified.

As r decreases for a given t , the corresponding comoments follow a trajectory indicated by the dashed line on Figure 2, from the boundary ∂F to the point C_0 . At the same time, the entropy \mathcal{E}^λ increases, and the trajectory crosses contours of higher entropy (\mathcal{E}' then \mathcal{E}'' on the figure.) We prove in Appendix A that the CML Estimator $\hat{\lambda}$ could also be obtained by taking the normal to the isoentropy contour that goes through the observed comoments \hat{C} , as shown on Figure 2:

Theorem 5 (Geometric identification and estimation) *The estimator $\hat{\lambda}$ of the parameter vector is given by the gradient of $-\mathcal{E}_r(\cdot)$ at the point \hat{C} .*

Concluding Remarks

While the framework we used here is bipartite, one-to-one matching, our results open the way to possible extensions to other matching problems. Among these, the “roommate problem” drops the requirement that the two partners of a match are drawn from distinct populations. Chiappori, Galichon, and Salanié (2012) have shown that this problem is in

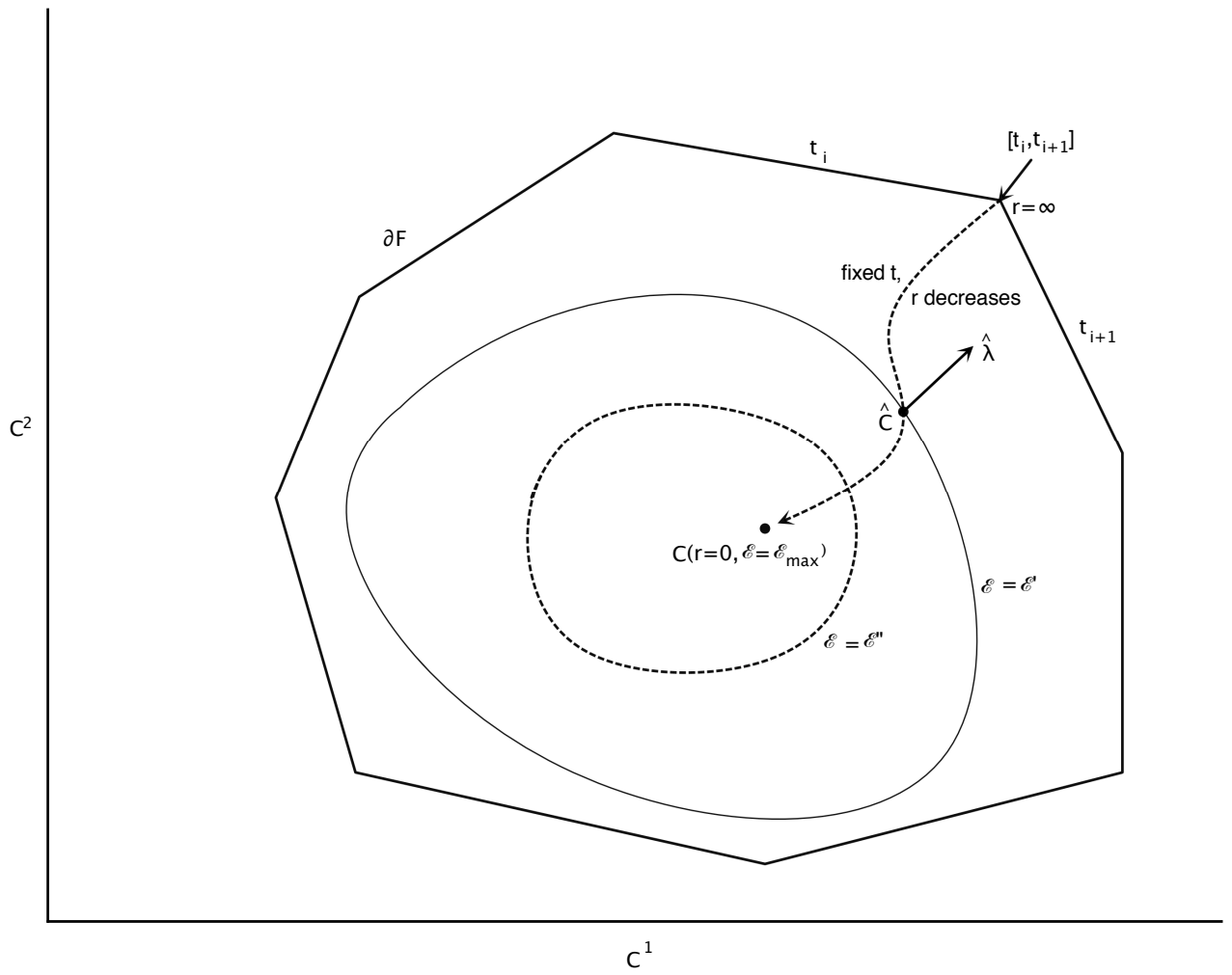


Figure 2: The covariogram and related objects

fact isomorphic, in a large population, to an associated bipartite matching problem; as a consequence, the empirical tools from the present paper can be extended to the study of the roommate problem. Although an extension to situations of “one-to-many matching” where one entity on one side of the market (such as a firm) may match with several agents on the other (such as employees) seems less direct, it is likely that the present approach would be useful. It may also be insightful in the study of trading on networks, when transfers are allowed (thus providing an empirical counterpart to Hatfield, Kominers, Nichifor, Ostrovsky, and Westkamp (2011), Hatfield and Kominers (2012).)

As mentioned earlier, several other approaches to estimating matching models with heterogeneity exist. One could directly specify the equilibrium utilities of each man and woman, as Hitsch, Hortacsu, and Ariely (2010) did in a non-transferable utility model. Under separability, this would amount to choosing a distribution \mathbf{P}_x and a parameterization λ of U and fitting the multinomial choice model

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

to the observed matches of men of type x . The downside is that unlike the joint surplus, the utilities U and V are not primitive objects; and it is very difficult to justify a specification of equilibrium utilities.

An alternative class of approaches pools data from many markets in which the surplus from a match is assumed to be the same. Fox (2010) starts from the standard monotonicity property of single-agent choice models, in which under very weak assumptions, the probability of choosing an alternative increases with its mean utility. By analogy, he posits a “rank-order property” for matching models with transferable utility: given the characteristics of the populations of men and women, a given matching is more likely than another when it produces a higher expected surplus.

Unlike the results we derived from the Choo and Siow (2006) framework, the rank-order property is not implied by any theoretical model we know of. In our framework, it holds only when the generalized entropy is a constant function, that is when there is no matching on unobservable characteristics. The attraction of the identification results based

on the rank-order property, on the other hand, is that they extend easily to models with many-to-one or many-to-many matching.

Finally, Fox and Yang (2012) take an approach that is somewhat dual to ours: while we use separability to restrict the distribution of unobserved heterogeneity so we can focus on the surplus over observables, they restrict the latter in order to recover the distribution of complementarities across unobservables. To do this, they rely on pooling data across many markets; in fact given the very high dimensionality of unobservable shocks, their method, while very ingenious, has yet to be tested on real data.

References

- BAJARI, P., AND J. FOX (2009): “Measuring the Efficiency of an FCC Spectrum Auction,” University of Michigan mimeo.
- BAUSCHKE, H., AND J. BORWEIN (1997): “Legendre Functions and the Method of Random Bregman Projections,” *Journal of Convex Analysis*, 4, 27–67.
- BECKER, G. (1973): “A Theory of Marriage, part I,” *Journal of Political Economy*, 81, 813–846.
- BOTTICINI, M., AND A. SIOW (2008): “Are there Increasing Returns in Marriage Markets?,” University of Toronto mimeo.
- BYRD, R., J. NOCEDAL, AND R. WALTZ (2006): “KNITRO: An Integrated Package for Nonlinear Optimization,” in *Large-Scale Nonlinear Optimization*, p. 3559. Springer Verlag.
- CHIAPPORI, P.-A., A. GALICHON, AND B. SALANIÉ (2012): “The Roommate Problem is More Stable than You Think,” mimeo.
- CHIAPPORI, P.-A., R. MCCANN, AND L. NESHEIM (2010): “Hedonic Price Equilibria, Stable Matching, and Optimal Transport: Equivalence, Topology, and Uniqueness,” *Economic Theory*, 42, 317–354.
- CHIAPPORI, P.-A., B. SALANIÉ, AND Y. WEISS (2012): “Partner Choice and the Marital College Premium,” mimeo Columbia University.
- CHOO, E., AND A. SIOW (2006): “Who Marries Whom and Why,” *Journal of Political Economy*, 114, 175–201.
- CSISZÁR, I. (1975): “ I -divergence Geometry of Probability Distributions and Minimization Problems,” *Annals of Probability*, 3, 146–158.

- DE PALMA, A., AND K. KILANI (2007): “Invariance of Conditional Maximum Utility,” *Journal of Economic Theory*, 132, 137–146.
- DECKER, C., E. LIEB, R. MCCANN, AND B. STEPHENS (2012): “Unique Equilibria and Substitution Effects in a Stochastic Model of the Marriage Market,” forthcoming in the *Journal of Economic Theory*.
- FOX, J. (2010): “Identification in Matching Games,” *Quantitative Economics*, 1, 203–254.
- (2011): “Estimating Matching Games with Transfers,” mimeo.
- FOX, J., AND C. YANG (2012): “Unobserved Heterogeneity in Matching Games,” mimeo University of Michigan.
- GABAIX, X., AND A. LANDIER (2008): “Why Has CEO Pay Increased So Much?,” *Quarterly Journal of Economics*, 123, 49–100.
- GALE, D., AND L. SHAPLEY (1962): “College Admissions and the Stability of Marriage,” *American Mathematical Monthly*, 69, 9–14.
- GALICHON, A., AND B. SALANIÉ (2010): “Matching with Trade-offs: Revealed Preferences over Competing Characteristics,” Discussion Paper 7858, CEPR.
- GRAHAM, B. (2011): “Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers,” in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. Jackson. Elsevier.
- GRETSKY, N., J. OSTROY, AND W. ZAME (1992): “The nonatomic assignment model,” *Economic Theory*, 2(1), 103–127.
- (1999): “Perfect competition in the continuous assignment model,” *Journal of Economic Theory*, 88, 60–118.
- HATFIELD, J. W., AND S. D. KOMINERS (2012): “Matching in Networks with Bilateral Contracts,” *American Economic Journal: Microeconomics*, 4, 176–208.

- HATFIELD, J. W., S. D. KOMINERS, A. NICHIFOR, M. OSTROVSKY, AND A. WESTKAMP (2011): “Stability and competitive equilibrium in trading networks,” mimeo Stanford GSB.
- HITSCH, G., A. HORTACSU, AND D. ARIELY (2010): “Matching and Sorting in Online Dating,” *American Economic Review*, 100, 130–163.
- JACQUEMET, N., AND J.-M. ROBIN (2011): “Marriage with Labor Supply,” mimeo, Sciences Po.
- SHAPLEY, L., AND M. SHUBIK (1972): “The Assignment Game I: The Core,” *International Journal of Game Theory*, 1, 111–130.
- SHIMER, R., AND L. SMITH (2000): “Assortative matching and Search,” *Econometrica*, 68, 343–369.
- SIOW, A. (2009): “Testing Becker’s Theory of Positive Assortative Matching,” Discussion paper, University of Toronto.
- SIOW, A., AND E. CHOO (2006): “Estimating a Marriage Matching Model with Spillover Effects,” *Demography*, 43(3), 463–490.
- TERVIO, M. (2008): “The difference that CEO make: An Assignment Model Approach,” *American Economic Review*, 98, 642–668.

Appendix

A Proofs

A.1 Proof of Theorem 2

(i) By the classical dual formulation of the matching problem, the market equilibrium assigns utilities u_{xi} to man i such that $x_i = x$ and v_{yj} to woman j such that $y_j = j$ so as to solve

$$\mathcal{W} = \min \left(\sum_i \tilde{u}_i + \sum_j \tilde{v}_j \right)$$

where the minimum is taken under the set of constraints

$$\tilde{u}_i + \tilde{v}_j \geq \Phi_{xy} + \varepsilon_{iy} + \eta_{xj}, \quad \tilde{u}_i \geq \varepsilon_{0i}, \quad \tilde{v}_j \geq \eta_{0j}.$$

Denote

$$U_{xy} = \min_{i:x_i=x} \{ \tilde{u}_i - \varepsilon_{iy} \}, \quad x \in \mathcal{X}, y \in \mathcal{Y}_0$$

$$V_{xy} = \min_{j:y_j=y} \{ \tilde{v}_j - \eta_{xj} \}, \quad x \in \mathcal{X}_0, y \in \mathcal{Y}$$

so that

$$\tilde{u}_i = \max_{y \in \mathcal{Y}_0} \{ U_{x_i y} + \varepsilon_{iy} \} \quad \text{and} \quad \tilde{v}_j = \max_{x \in \mathcal{X}_0} \{ V_{x y_j} + \eta_{xj} \}$$

Then

$$\mathcal{W} = \min \left(\sum_i \max_{y \in \mathcal{Y}_0} \{ U_{x_i y} + \varepsilon_{iy} \} + \sum_j \max_{x \in \mathcal{X}_0} \{ V_{x y_j} + \eta_{xj} \} \right)$$

under the set of constraints

$$U_{xy} + V_{xy} \geq \Phi_{xy}, \quad U_{x0} \geq 0, \quad V_{0y} \geq 0.$$

Assign non-negative multipliers $\mu_{xy}, \mu_{x0}, \mu_{0y}$ to these constraints. By duality in Linear

Programming, we can rewrite

$$\mathcal{W} = \max_{\mu_{xy} \geq 0} \left(\sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy} - \max_{U_{xy}} \left\{ \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}_0}} \mu_{xy} U_{xy} - \sum_i \max_{y \in \mathcal{Y}_0} \{U_{x_i y} + \varepsilon_{iy}\} \right\} \right. \\ \left. - \max_{V_{xy}} \left\{ \sum_{\substack{x \in \mathcal{X}_0 \\ y \in \mathcal{Y}}} \mu_{xy} V_{xy} - \sum_j \max_{x \in \mathcal{X}_0} \{V_{xy_j} + \eta_{xj}\} \right\} \right).$$

Now,

$$\sum_i \max_{y \in \mathcal{Y}_0} \{U_{x_i y} + \varepsilon_{iy}\} = \sum_x n_x \mathbb{E}_{\mathbf{P}_x} \max_{y \in \mathcal{Y}_0} \{U_{x_i y} + \varepsilon_{iy}\} = n_x G_x(\mathbf{U}_{\mathbf{x}.}),$$

where $\mathbb{E}_{\mathbf{P}_x}$ denotes the expectation over the population of men in group x , and where we have invoked Assumption 1 and the law of large numbers in order to replace the sum by an expectation. Adding the similar expression for women, we get

$$\mathcal{W} = \max_{(\mu_{xy})} \left(\sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy} - A(\mu) - B(\mu) \right)$$

where

$$A(\mu) = \max_{(U_{xy})} \left\{ \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}_0}} \mu_{xy} U_{xy} - \sum_{x \in \mathcal{X}} n_x G_x(\mathbf{U}_{\mathbf{x}.}) \right\} \\ B(\mu) = \max_{(V_{xy})} \left\{ \sum_{\substack{x \in \mathcal{X}_0 \\ y \in \mathcal{Y}}} \mu_{xy} V_{xy} - \sum_{y \in \mathcal{Y}} m_y H_y(\mathbf{V}_{\cdot y}) \right\}$$

Consider the term with first subscript x in $A(\mu)$. It is

$$\sum_{y \in \mathcal{Y}_0} \mu_{xy} U_{xy} - n_x G_x(\mathbf{U}_{\mathbf{x}.}).$$

It is easy to see that since G_x is the expected maximum utility, for any number t we have $G_x(\mathbf{U}_{\mathbf{x}.} + t) = G_x(\mathbf{U}_{\mathbf{x}.}) + t$; therefore if $\sum_{y \in \mathcal{Y}_0} \mu_{xy} \neq n_x$, the term is plus infinity. This

implies that at the maximum in \mathcal{W} , the feasibility constraints in (1.2) must hold, and we can rewrite $A(\mu)$ and $B(\mu)$ in terms of the rescaled Legendre-Fenchel transforms:

$$A(\mu) = \sum_{x \in \mathcal{X}} G_x^*(\boldsymbol{\mu}_{\mathbf{x}\cdot}) \quad \text{and} \quad B(\mu) = \sum_{y \in \mathcal{Y}} H_y^*(\boldsymbol{\mu}_{\cdot\mathbf{y}}).$$

It follows that

$$\mathcal{W} = \max_{\mu \in \mathcal{M}(n,m)} \left\{ \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy} - \sum_{x \in \mathcal{X}} G_x^*(\boldsymbol{\mu}_{\mathbf{x}\cdot}) - \sum_{y \in \mathcal{Y}} H_y^*(\boldsymbol{\mu}_{\cdot\mathbf{y}}) \right\}.$$

Assigning multipliers a_x and b_y to the feasibility constraints, the first order conditions of this problem are

$$\Phi_{xy} + a_x + b_y = \frac{\partial G_x^*}{\partial \mu_{xy}}(\boldsymbol{\mu}_{\mathbf{x}\cdot}) + \frac{\partial H_y^*}{\partial \mu_{xy}}(\boldsymbol{\mu}_{\cdot\mathbf{y}})$$

and

$$a_x = \frac{\partial G_x^*}{\partial \mu_{x0}}(\boldsymbol{\mu}_{\mathbf{x}\cdot}), \quad b_y = \frac{\partial H_y^*}{\partial \mu_{0y}}(\boldsymbol{\mu}_{\cdot\mathbf{y}}).$$

Combining them gives formula (2.8).

(ii) From the proof of part (i), by the envelope theorem

$$U_{xy} = \frac{\partial A}{\partial \mu_{xy}}(\mu);$$

and since

$$A(\mu) = \sum_{x \in \mathcal{X}} G_x^*(\boldsymbol{\mu}_{\mathbf{x}\cdot}),$$

adding the normalization $U_{x0} = 0$ gives the formula for U_{xy} in the theorem.

(iii) By duality, $u_x = G_x(\mathbf{U}_{\mathbf{x}\cdot})$ is the multiplier of the feasibility constraint for group x ; and the proof of (i) shows that this is

$$a_x = \frac{\partial G_x^*}{\partial \mu_{x0}}(\boldsymbol{\mu}_{\mathbf{x}\cdot}).$$

The proof of (iv) is the same as for (ii) and (iii).

(v) follows from the fact that $U_{xy} = \alpha_{xy} + \tau_{xy}$ and $V_{xy} = \gamma_{xy} - \tau_{xy}$; thus if U_{xy} and V_{xy} are identified and τ_{xy} is observed, then α and γ are identified by

$$\alpha_{xy} = U_{xy} - \tau_{xy} \quad \text{and} \quad \gamma_{xy} = V_{xy} + \tau_{xy}.$$

A.2 Proof of Proposition 2

By definition,

$$G_x(\bar{\mathbf{U}}_{\mathbf{x}\cdot}) = \mathbb{E}_{\mathbf{P}_x} \left(\max_{y \in \mathcal{Y}_0} (U_{xy} + \varepsilon_{iy}) \right)$$

and

$$G_x^*(\boldsymbol{\mu}_{\mathbf{x}\cdot}) = \max_{\tilde{\mathbf{U}}_{\mathbf{x}\cdot} = (\tilde{U}_{x0}, \dots, \tilde{U}_{x|\mathcal{Y}|})} \left(\sum_{y \in \mathcal{Y}_0} \mu_{xy} \tilde{U}_{xy} - \left(\sum_{y \in \mathcal{Y}_0} \mu_{xy} \right) G_x(\tilde{\mathbf{U}}_{\mathbf{x}\cdot}) \right).$$

Now, using the feasibility constraint $\sum_{y \in \mathcal{Y}_0} \mu_{xy} = n_x$:

$$G_x^*(\boldsymbol{\mu}_{\mathbf{x}\cdot}) = -n_x \min_{\tilde{\mathbf{U}}_{\mathbf{x}\cdot}} \left(\mathbb{E}_{\mathbf{P}_x} \left(\max_{y \in \mathcal{Y}_0} (\tilde{U}_{xy} + \varepsilon_{iy}) \right) - \sum_{y \in \mathcal{Y}_0} \frac{\mu_{xy}}{n_x} \tilde{U}_{xy} \right),$$

and defining $\bar{U}_{xy} = -\tilde{U}_{xy}$, this is also

$$G_x^*(\boldsymbol{\mu}_{\mathbf{x}\cdot}) = -n_x \min_{\bar{\mathbf{U}}_{\mathbf{x}\cdot}} \left(\sum_{y \in \mathcal{Y}_0} \frac{\mu_{xy}}{n_x} \bar{U}_{xy} + \mathbb{E}_{\mathbf{P}_x} \left(\max_{y \in \mathcal{Y}_0} (\varepsilon_{iy} - \bar{U}_{xy}) \right) \right)$$

The first term in the minimand is the expectation of $\bar{\mathbf{U}}_{\mathbf{x}\cdot}$ under the distribution $\mu_{Y|X=x}$; therefore this can be rewritten as

$$G_x^*(\boldsymbol{\mu}_{\mathbf{x}\cdot}) = -n_x \min_{\bar{U}_{xy} + k_x(\varepsilon_{i\cdot}) \geq \varepsilon_{iy}} \left(E_{\mu_{Y|X=x}} \bar{U}_{xY} + \mathbb{E}_{\mathbf{P}_x} k_x(\varepsilon_{i\cdot}) \right)$$

where the minimum is taken over all pairs of functions $(\bar{\mathbf{U}}_{\mathbf{x}\cdot}, k_x(\varepsilon_{i\cdot}))$ that satisfy the inequality. We recognize the value of the dual of a matching problem in which the margins are $\mu_{Y|X=x}$ and \mathbf{P}_x and the surplus is ε_{iy} . By the equivalence of the primal and the dual, this gives

$$G_x^*(\boldsymbol{\mu}_{\mathbf{x}\cdot}) = -n_x \max_{\pi \in \mathcal{M}_x} \mathbb{E}_{\pi} [\varepsilon_{iy}].$$

A.3 Proof of Theorem 3

The proof uses results in Bauschke and Borwein (1997), which builds on Csiszár (1975).

For any matching μ , consider the function

$$\varphi(\mu) = -\mathcal{E}(\mu).$$

Since generalized entropy \mathcal{E} is concave in μ , φ is a convex function. In fact, it satisfies the conditions in Bauschke and Borwein (1997); in particular it is a Legendre function⁸. Introduce D the associated “Bregman divergence” as

$$D(\mu; \nu) = \varphi(\mu) - \varphi(\nu) - \langle \nabla \varphi(\nu), \mu - \nu \rangle.$$

Bregman divergences are often called “Bregman distances”; they are not distances, but they are useful for our purposes because one can generalize the concept of a projection to Bregman divergences.

Now take any surplus function Φ and margins \hat{n} and \hat{m} . Step 1 of the algorithm constructs a matching $\mu^{(0)}$ such that for all $x \neq 0, y \neq 0$

$$\Phi_{xy} = \frac{\partial \varphi}{\partial \mu_{xy}}(\mu^{(0)}) - \frac{\partial \varphi}{\partial \mu_{x0}}(\mu^{(0)}) - \frac{\partial \varphi}{\partial \mu_{0y}}(\mu^{(0)})$$

and $2 \sum_{xy} \mu_{xy}^{(0)} + \sum_x \mu_{x0}^{(0)} + \sum_y \mu_{0y}^{(0)} = \sum_x \hat{n}_x + \sum_y \hat{m}_y$. Note that while $\mu^{(0)}$ adds up to the total number of men and women, it needn’t satisfy any of the other feasibility constraints. Moreover, by Theorem 2, the optimal matching for Φ given margins n and m (which always exists) satisfies all of these constraints; therefore we can always find such a $\mu^{(0)}$.

Then

$$\langle \nabla \varphi(\mu^{(0)}), \mu - \mu^{(0)} \rangle = \sum_{xy} \frac{\partial \varphi}{\partial \mu_{xy}}(\mu^{(0)}) (\mu_{xy} - \mu_{xy}^{(0)}) + \sum_x \frac{\partial \varphi}{\partial \mu_{x0}}(\mu^{(0)}) (\mu_{x0} - \mu_{x0}^{(0)}) + \sum_y \frac{\partial \varphi}{\partial \mu_{0y}}(\mu^{(0)}) (\mu_{0y} - \mu_{0y}^{(0)})$$

becomes

$$\sum_{xy} \mu_{xy} \Phi_{xy} + \sum_x \frac{\partial \varphi}{\partial \mu_{x0}}(\mu^{(0)}) \left(\sum_y \mu_{xy} + \mu_{x0} \right) + \sum_y \frac{\partial \varphi}{\partial \mu_{0y}}(\mu^{(0)}) \left(\sum_x \mu_{xy} + \mu_{0y} \right),$$

up to additive terms that depend on $\mu^{(0)}$ but not on μ . But if $\mu \in \mathcal{M}(\hat{n}, \hat{m})$, then $\sum_y \mu_{xy} + \mu_{x0} = \hat{n}_x$ and $\sum_x \mu_{xy} + \mu_{0y} = \hat{m}_y$, so that up to irrelevant terms,

$$\langle \nabla \varphi(\mu^{(0)}), \mu - \mu^{(0)} \rangle = \sum_{xy} \mu_{xy} \Phi_{xy}$$

⁸A Legendre function is essentially smooth and essentially strictly convex—see section 2 of Bauschke and Borwein (1997) for details.

and

$$D(\mu, \mu^{(0)}) = \varphi(\mu) + \sum_{xy} \mu_{xy} \Phi_{xy} = \sum_{xy} \mu_{xy} \Phi_{xy} - \mathcal{E}(\mu)$$

on $\mathcal{M}(\hat{n}, \hat{m})$.

Since the optimal matching for surplus Φ and margins \hat{n} and \hat{m} maximizes

$$\sum_{xy} \mu_{xy} \Phi_{xy} + \mathcal{E}(\mu)$$

it can also be found by minimizing $D(\mu, \mu^{(0)})$ over $\mu \in \mathcal{M}(\hat{n}, \hat{m})$; that is, by projecting $\mu^{(0)}$ on $\mathcal{M}(\hat{n}, \hat{m})$ using the Bregman divergence.

Introduce the linear subspaces $\mathcal{M}(\hat{n})$ and $\mathcal{M}(\hat{m})$ by

$$\mathcal{M}(\hat{n}) = \left\{ \mu \geq 0 : \forall x \in \mathcal{X}, \sum_{y \in \mathcal{Y}} \mu_{xy} + \mu_{x0} = \hat{n}_x \right\}$$

$$\mathcal{M}(\hat{m}) = \left\{ \mu \geq 0 : \forall y \in \mathcal{Y}, \sum_{x \in \mathcal{X}} \mu_{xy} + \mu_{0y} = \hat{m}_y \right\}$$

so that

$$\mathcal{M}(\hat{n}, \hat{m}) = \mathcal{M}(\hat{n}) \cap \mathcal{M}(\hat{m}).$$

We define $\mu^{(k)}$ recursively by iteratively projecting with respect to D on the linear subspaces $\mathcal{M}(\hat{n})$ and on $\mathcal{M}(\hat{m})$:

$$\mu^{(2k+1)} = \arg \min_{\mu \in \mathcal{M}(\hat{n})} D(\mu; \mu^{(2k)}) \tag{A.1}$$

$$\mu^{(2k+2)} = \arg \min_{\mu \in \mathcal{M}(\hat{m})} D(\mu; \mu^{(2k+1)}) \tag{A.2}$$

By Theorem 8.4 of Bauschke and Borwein (1997), the iterated projection algorithm converges⁹ to the projection μ of $\mu^{(0)}$ on $\mathcal{M}(\hat{n}, \hat{m})$, which is also the maximizer μ of (2.7).

⁹In the notation of their Theorem 8.4, the hyperplanes (C_i) are $\mathcal{M}(p)$ and $\mathcal{M}(q)$; and the Bregman/Legendre function f is our ϕ .

The updating formulas of the Theorem are easily obtained; let us work out (A.1). Introduce multipliers a_x for the feasibility constraints $\mu^{(2k+1)} \in \mathcal{M}(\hat{n})$. Neglecting irrelevant terms again, the Bregman divergence is

$$D(\mu, \mu^{(2k)}) = \varphi(\mu) - \sum_{xy} \frac{\partial \varphi}{\partial \mu_{xy}}(\mu^{(2k)}) \mu_{xy} - \sum_x \frac{\partial \varphi}{\partial \mu_{x0}}(\mu^{(2k)}) \mu_{x0} - \sum_y \frac{\partial \varphi}{\partial \mu_{0y}}(\mu^{(2k)}) \mu_{0y}$$

and the constraints are $\sum_y \mu_{xy} + \mu_{x0} = \hat{n}_x$ for all x .

The first order conditions are simply

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial \mu_{xy}}(\mu^{(2k+1)}) - \frac{\partial \mathcal{E}}{\partial \mu_{xy}}(\mu^{(2k)}) &= a_x \text{ for } x \in \mathcal{X}, y \in \mathcal{Y}_0 \\ \frac{\partial \mathcal{E}}{\partial \mu_{0y}}(\mu^{(2k+1)}) - \frac{\partial \mathcal{E}}{\partial \mu_{0y}}(\mu^{(2k)}) &= 0 \text{ for } y \in \mathcal{Y}. \end{aligned}$$

A.4 Proof of Theorem 4

Since \mathcal{W} is 1-homogeneous in (\hat{n}, \hat{m}, μ) ,

$$\mathcal{W} = \sum_x \hat{n}_x \frac{\partial \mathcal{W}}{\partial n_x} + \sum_y \hat{m}_y \frac{\partial \mathcal{W}}{\partial m_y}.$$

Therefore

$$\frac{\partial \mathcal{W}}{\partial \lambda} = \sum_x \hat{n}_x \frac{\partial^2 \mathcal{W}}{\partial n_x \partial \lambda} + \sum_y \hat{m}_y \frac{\partial^2 \mathcal{W}}{\partial m_y \partial \lambda}.$$

Now by (iii) of Theorem 2,

$$\frac{\partial \mathcal{W}}{\partial n_x} = u_x = -\log \frac{\mu_{x0}^\lambda}{\hat{n}_x}$$

hence

$$\frac{\partial^2 \mathcal{W}}{\partial n_x \partial \lambda} = -\frac{\partial \log \mu_{x0}^\lambda}{\partial \lambda}.$$

Therefore

$$\frac{\partial \mathcal{W}}{\partial \lambda} = -\sum_x \hat{n}_x \frac{\partial \log \mu_{x0}^\lambda}{\partial \lambda} - \sum_y \hat{m}_y \frac{\partial \log \mu_{0y}^\lambda}{\partial \lambda}. \quad (\text{A.3})$$

Now consider the derivative of the log-likelihood:

$$\frac{\partial \log L}{\partial \lambda} = 2 \sum_{xy} \hat{\mu}_{xy} \frac{\partial \log \mu_{xy}^\lambda}{\partial \lambda} + \sum_x \hat{\mu}_{x0} \frac{\partial \log \mu_{x0}^\lambda}{\partial \lambda} + \sum_y \hat{\mu}_{0y} \frac{\partial \log \mu_{0y}^\lambda}{\partial \lambda}.$$

Since $\hat{\mu}_{x0} = \hat{n}_x - \sum_y \hat{\mu}_{xy}$, we get

$$\sum_y \hat{\mu}_{xy} \frac{\partial \log \mu_{xy}^\lambda}{\partial \lambda} + \hat{\mu}_{x0} \frac{\partial \log \mu_{x0}^\lambda}{\partial \lambda} = \hat{n}_x \frac{\partial \log \mu_{x0}^\lambda}{\partial \lambda} + \sum_y \hat{\mu}_{xy} \left(\frac{\partial \log \mu_{xy}^\lambda}{\partial \lambda} - \frac{\partial \log \mu_{x0}^\lambda}{\partial \lambda} \right);$$

adding up with similar terms for women gives

$$\frac{\partial \log L}{\partial \lambda} = \sum_{xy} \hat{\mu}_{xy} \left(2 \frac{\partial \log \mu_{xy}^\lambda}{\partial \lambda} - \frac{\partial \log \mu_{x0}^\lambda}{\partial \lambda} - \frac{\partial \log \mu_{0y}^\lambda}{\partial \lambda} \right) + \sum_x \hat{n}_x \frac{\partial \log \mu_{x0}^\lambda}{\partial \lambda} + \sum_y \hat{n}_y \frac{\partial \log \mu_{0y}^\lambda}{\partial \lambda};$$

or, using (A.3),

$$\frac{\partial \log L}{\partial \lambda} = \sum_{xy} \hat{\mu}_{xy} \left(2 \frac{\partial \log \mu_{xy}^\lambda}{\partial \lambda} - \frac{\partial \log \mu_{x0}^\lambda}{\partial \lambda} - \frac{\partial \log \mu_{0y}^\lambda}{\partial \lambda} \right) - \frac{\partial \mathcal{W}}{\partial \lambda}.$$

But given Theorem 1, this is just

$$\frac{\partial \log L}{\partial \lambda} = \sum_{xy} \hat{\mu}_{xy} \frac{\partial \Phi_{xy}^\lambda}{\partial \lambda} - \frac{\partial \mathcal{W}}{\partial \lambda},$$

which establishes part 1(a) of the Theorem.

Now by the envelope theorem,

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

since the entropy term does not depend on λ in the Choo and Siow model; this proves part 1(b) since with the linear specification

$$\frac{\partial \Phi_{xy}^\lambda}{\partial \lambda^k} = C^k(\lambda).$$

To prove part 2, note that since $\mu^{\hat{\lambda}}$ maximizes \mathcal{W} when $\lambda = \hat{\lambda}$,

$$\sum_{x,y} \hat{\mu}_{xy} \Phi_{xy}^{\hat{\lambda}} + \mathcal{E}(\hat{\mu}) \leq \sum_{x,y} \mu_{xy}^{\hat{\lambda}} \Phi_{xy}^{\hat{\lambda}} + \mathcal{E}(\mu^{\hat{\lambda}})$$

and, since \mathcal{E} is strictly concave in μ , equality holds if and only if $\mu^{\hat{\lambda}} = \hat{\mu}$. But

$$\sum_{x,y} \hat{\mu}_{xy} \Phi_{xy}^{\hat{\lambda}} = \sum_{x,y} \mu_{xy}^{\hat{\lambda}} \Phi_{xy}^{\hat{\lambda}}$$

by construction, hence

$$\mathcal{E}(\hat{\mu}) \leq \mathcal{E}(\mu^{\hat{\lambda}})$$

with equality if and only if $\mu^{\hat{\lambda}} = \hat{\mu}$.

A.5 Proof of Theorem 5

Let us first prove that

$$\mathcal{E}_r(\hat{C}) = \min_{\lambda} \left(\mathcal{W}(\lambda) - \sum_{k=1}^K \lambda_k \hat{C}^k \right). \quad (\text{A.4})$$

Indeed, the optimum is reached at $\hat{\lambda} = \lambda(\hat{C})$, and there

$$\mathcal{E}_r(\hat{C}) = \mathcal{W}(\hat{\lambda}) - \sum_{k=1}^K \hat{\lambda}_k \hat{C}^k = \mathcal{E}(\mu^{\hat{\lambda}})$$

which shows (A.4). This implies that $\mathcal{E}_r(\hat{C})$ is a concave function; and by the envelope theorem in (A.4), we get

$$\frac{\partial \mathcal{E}_r}{\partial C^k}(\hat{C}) = \hat{\lambda}^k.$$

B The Generalized Extreme Values Framework

Consider a family of functions $g_x : \mathbb{R}^{|\mathcal{J}|+1} \rightarrow \mathbb{R}$ such that the following four conditions hold:

- (i) g_x are positive homogeneous of degree one; (ii) they go to $+\infty$ whenever any of their arguments goes to $+\infty$ (iii) their partial derivatives of order k exist outside of 0 and have sign $(-1)^k$ (iv) and the functions defined by

$$\mathbf{P}_x(w_0, \dots, w_J) = \exp(-g_x(e^{-w_0}, \dots, e^{-w_J}))$$

are multivariate cumulative distribution functions. Then introducing utility shocks $\varepsilon_x \sim \mathbf{P}_x$, we have by a theorem of McFadden (1978):

$$\frac{G_x(w)}{n_x} = \mathbb{E}_{\mathbf{P}_x} \left[\max_{y \in \mathcal{Y}_0} \{w_y + \varepsilon_y\} \right] = \log g_x(e^w) + \gamma$$

where γ is the Euler constant $\gamma \simeq 0.577$. Therefore, if $\sum_{y \in \mathcal{Y}_0} a_y = n_x$, then

$$G_x^*(n_x, a) = \sum_{y \in \mathcal{Y}_0} a_y w_y^x(n_x, a) - n_x \left(\log g_x \left(e^{w^x(n_x, a)} \right) + \gamma \right)$$

where for $x \in \mathcal{X}_0$, the vector $w^x(n_x, a)$ solves the system

$$a_y = n_x \frac{\partial}{\partial w_y^x} \log g_x(e^{w^x}), \quad y \in \mathcal{Y}_0. \quad (\text{B.1})$$

Hence, the part of the expression of $\mathcal{E}(n, m, \mu)$ arising from the heterogeneity on the men side is

$$\sum_{x \in \mathcal{X}} \left(n_x \log g_x \left(e^{w^x(n_x, \boldsymbol{\mu}_{x \cdot})} \right) - \sum_{y \in \mathcal{Y}_0} \mu_{xy} w_y^x(n_x, \boldsymbol{\mu}_{x \cdot}) \right) + C$$

where $C = \gamma \sum_{x \in \mathcal{X}} n_x$, whose derivative with respect to μ_{xy} ($x, y \geq 1$) is $-w_y^x(n_x, \boldsymbol{\mu}_{x \cdot})$.

C Computations for the Examples

C.1 Computations for Example 1

With type I extreme values iid distributions, the expected utility is

$$G_x(\mathbf{U}_{x \cdot}) = \log \sum_{y \in \mathcal{Y}_0} \exp(U_{xy}),$$

and the maximum in the program that defines $G_x^*(\boldsymbol{\mu}_{x \cdot})$ is achieved in

$$U_{xy} = U_{x0} + \log \frac{\mu_{xy}}{\mu_{x0}}.$$

This yields

$$G_x^*(\boldsymbol{\mu}_{x \cdot}) = \sum_{y \in \mathcal{Y}_0} \mu_{xy} \log \mu_{xy} - \left(\sum_{y \in \mathcal{Y}_0} \mu_{xy} \right) \log \left(\sum_{y \in \mathcal{Y}_0} \mu_{xy} \right)$$

which gives equation (3.1). Equation (3.2) obtains by straightforward differentiation.

C.2 Computations for Example 2

Consider a man of a group x . Such a man marries a woman of education e' within social group g' with conditional probability

$$\frac{\mu_{x,g'e'}}{\mu_{x,g'}} = \frac{\exp(U_{x,g'e'}/\sigma_{x,g'})}{\sum_{e'=1}^L \exp(U_{x,g'e'}/\sigma_{x,g'})};$$

and his probability of marrying within group g' is

$$\frac{\mu_{x,g'}}{n_x} = \frac{\left(\sum_{e'=1}^L \exp(U_{x,g'e'}/\sigma_{x,g'})\right)^{\sigma_{x,g'}}}{1 + \sum_{g''=1}^G \left(\sum_{e'=1}^L \exp(U_{x,g''e'}/\sigma_{x,g''})\right)^{\sigma_{x,g''}}}.$$

Then, taking logs and subtracting,

$$U_{x,g'e'} = \sigma_{x,g'} \log \frac{\mu_{x,g'e'}}{\mu_{x,g'}} + t_{x,g'}$$

so that only the constants $t_{x,g'}$ remain to be determined by solving (with $U_{x0} = 0$ as usual)

$$\frac{\mu_{x,g'}}{n_x} = \frac{\exp(t_{x,g'})}{1 + \sum_{g''=1}^G \exp(t_{x,g''})}.$$

This gives $\exp(t_{x,g'}) = \frac{\mu_{x,g'}}{\mu_{x,0}}$, so that

$$U_{x,g'e'} = \log \frac{\mu_{x,g'}}{\mu_{x,0}} + \sigma_{x,g'} \log \frac{\mu_{x,g'e'}}{\mu_{x,g'}}.$$

The expected utility of this man is

$$u_x = \log \left(1 + \sum_{g=1}^G \sigma_{x,g'} \log \left(\sum_{e'=1}^L \exp(U_{x,g'e'}/\sigma_{x,g'}) \right) \right)$$

which is

$$u = \log \left(1 + \sum_{g=1}^G \frac{\mu_{x,g'}}{\mu_{x,0}} \sum_{e'=1}^L \left(\frac{\mu_{x,g'e'}}{\mu_{x,g'}} \right)^{\sigma_{x,g'}} \right).$$

Finally, the formula for $U_{x,ge}$ shows that the surplus from a marriage between a man (g, e) and a woman (g', e') is identified by the formula in the text.

C.3 Computations for Example 3

We start by giving a formula for G_x^* in the slightly more general case where ε is continuously distributed according to a c.d.f. F_ε . From Proposition 2,

$$G_x^*(\boldsymbol{\mu}_{\mathbf{x}\cdot}) = -n_x \max_{\pi \in \mathcal{M}_x} \mathbb{E}_\pi [\phi_x(Y) \varepsilon].$$

where π has margins F_ε and $\mu(Y|X=x)$. Since the function $(\varepsilon, \phi) \rightarrow \varepsilon\phi$ is supermodular, the optimal matching must exhibit positive assortative matching—larger ε 's must be matched with y 's with larger values of the index $\phi_x(y)$. Let

$$\phi_{(1)} < \dots < \phi_{(m)}$$

denote the distinct values that the index takes over \mathcal{Y}_0 ; value $\phi_{(k)}$ has probability

$$\Pr(\phi_x(Y) = \phi_{(k)} | X = x) = \frac{\sum_{\phi_x(y)=\phi_{(k)}} \mu_{xy}}{n_x}. \quad (\text{C.1})$$

By positive assortative matching, there exists a sequence

$$\varepsilon_{(0)} = \inf \varepsilon < \varepsilon_{(1)} < \dots < \varepsilon_{(m-1)} \varepsilon_{(m)} = \sup \varepsilon$$

such that ε matches with an y with $\phi_x(y) = \phi_{(k)}$ if and only if $\varepsilon \in [\varepsilon_{(k-1)}, \varepsilon_{(k)}]$; and since probability is conserved, the sequence is constructed recursively by

$$F_\varepsilon(\varepsilon_{(k)}) - F_\varepsilon(\varepsilon_{(k-1)}) = \frac{\sum_{\phi_x(y)=\phi_{(k)}} \mu_{xy}}{n_x}. \quad (\text{C.2})$$

The resulting surplus is

$$\sum_{k=1}^m \phi_{(k)} E(\varepsilon \mathbf{1}(\varepsilon \in [\varepsilon_{(k-1)}, \varepsilon_{(k)}])).$$

Now assume that ε is distributed uniformly over $[0, 1]$. Then (C.2) gives

$$\varepsilon_{(k)} = \frac{\sum_{\phi_x(y) \leq \phi_{(k)}} \mu_{xy}}{n_x} = \Pr(\phi_x(Y) \leq \phi_{(k)} | X = x),$$

and

$$E(\varepsilon \mathbf{1}(\varepsilon \in [\varepsilon_{(k-1)}, \varepsilon_{(k)}])) = \frac{\varepsilon_{(k)}^2 - \varepsilon_{(k-1)}^2}{2},$$

which is also

$$\left(\Pr(\phi_x(Y) \leq \phi_{(k)} | X = x) - \frac{\Pr(\phi_x(Y) \leq \phi_{(k)} | X = x)}{2} \right) \Pr(\phi_x(Y) = \phi_{(k)} | X = x).$$

Therefore $G_x^*(\boldsymbol{\mu}_{\mathbf{x}})$ is

$$-\sum_{k=1}^m \sum_{y | \phi_x(y) = \phi_{(k)}} \phi_{(k)} \left(\Pr(\phi_x(Y) \leq \phi_{(k)} | X = x) - \frac{\Pr(\phi_x(Y) \leq \phi_{(k)} | X = x)}{2} \right) \Pr(\phi_x(Y) = \phi_{(k)} | X = x);$$

changing the order of sums and using (C.1) turns this into

$$G_x^*(\boldsymbol{\mu}_{\mathbf{x}}) = - \sum_{y \in \mathcal{Y}_0} \mu_{xy} \phi_x(y) \frac{\sum_{\phi_x(y') \leq \phi_x(y)} \mu_{xy'} - \sum_{\phi_x(y') = \phi_x(y)} \mu_{xy'} / 2}{n_x}.$$

The derivative in μ_{xy} has two terms:

$$-\phi_x(y) \left(\frac{\sum_{\phi_x(y') \leq \phi_x(y)} \mu_{xy'} - \sum_{\phi_x(y') = \phi_x(y)} \mu_{xy'} / 2}{n_x} \right)$$

and

$$-\sum_{y' \in \mathcal{Y}_0} \mu_{xy'} \phi_x(y') \frac{\mathbf{1}(\phi_x(y) \leq \phi_x(y')) - \mathbf{1}(\phi_x(y) = \phi_x(y')) / 2}{n_x}.$$

Combining them,

$$\begin{aligned} \frac{\partial G_x^*(\boldsymbol{\mu}_{\mathbf{x}})}{\partial \mu_{xy}} &= - \sum_{\phi_x(y') \leq \phi_x(y)} \phi_x(y) \frac{\mu_{xy'}}{n_x} - \sum_{\phi_x(y) \leq \phi_x(y')} \frac{\mu_{xy'}}{n_x} \phi_x(y') + \sum_{\phi_x(y') = \phi_x(y)} \phi_x(y) \frac{\mu_{xy'}}{n_x} \\ &= - \left(\sum_{\phi_x(y') \leq \phi_x(y)} \phi_x(y) \frac{\mu_{xy'}}{n_x} + \sum_{\phi_x(y) < \phi_x(y')} \frac{\mu_{xy'}}{n_x} \phi_x(y') \right) \\ &= -\mathbb{E}[\max(\phi_x(Y), \phi_x(y))] \end{aligned}$$

so that

$$u_x = - \frac{\partial G_x^*(\boldsymbol{\mu}_{\mathbf{x}})}{\partial \mu_{x0}} = \mathbb{E}[\max(\phi_x(Y), \phi_x(0))]$$

and

$$\begin{aligned} U_{xy} &= \frac{\partial G_x^*(\boldsymbol{\mu}_{\mathbf{x}})}{\partial \mu_{xy}} - \frac{\partial G_x^*(\boldsymbol{\mu}_{\mathbf{x}})}{\partial \mu_{x0}} \\ &= -\mathbb{E}[\max(\phi_x(Y), \phi_x(y))] + \mathbb{E}[\max(\phi_x(Y), \phi_x(0))]. \end{aligned}$$