Matching with Transfers: Applications*

Bernard Salanié†

July 2, 2024

Introduction

Over the past twenty years, the empirical applications of matching models with market-clearing transfers have expanded dramatically. This growth has been driven by methodological advancements that elucidated the sources of identification in these models and proposed manageable specifications. This chapter will present these methodological developments, discuss their limitations, and highlight some of the key applications.

The Costs and Benefits of TU

Most of this chapter focuses on "perfectly transferable utility", which, like most of the literature, I will simply refer to as "TU". (Perfect) TU is a very demanding assumption: it requires that the partners in any given match be able to transfer utility freely to each other, without limits, and without loss or friction. TU implies the existence of a numéraire good that has the same, constant marginal value for all partners. Taken literally, it precludes many things that do matter in the real world: differing marginal utilities, frictions in exchange, and taxes on transfers among others. As such, it is not appropriate in all applications. The reader should turn to chapter 7

*Draft chapter for volume 1 of the *Handbook of the Economics of Matching*.

†Department of Economics, Columbia University, bsalanie@columbia.edu. I am grateful to Aloysius Siow and to Sean Elliott for their very useful comments on a first draft.

of this *Handbook* volume for the analysis of matching when the limitations on transfers cannot be neglected.

When TU is an acceptable assumption, however, it is a remarkably useful one in applications. Chapter 5 showed that given TU, the utility possibility frontier within a match is an entire hyperplane on which the sum of the utilities of the partners add up to a constant: the *joint surplus*. The fact that the frontier is a linear subspace greatly simplifies the properties of stable matchings, as compared with NTU. Whereas in NTU models the set of stable matchings is a lattice whose cardinality may grow exponentially with the number of agents, it is (generically) a singleton with TU. More importantly for the purpose of this chapter, TU opens the way to simple and powerful inference methods.

Since TU models presuppose the existence of a utility scale that is common to all agents, welfare analysis is also much more straightforward than under NTU. Once such a model is estimated, the analyst can put numbers on the trade offs that agents face in choosing partners with various characteristics, and also evaluate the effect of policy counterfactuals on the welfare of the various types of agents.

Consider the marriage market as an illustration. Becker's (1973, 1974)'s classic analysis made it the first major application of TU models in economics; I will use it as a running thread in this chapter. We could approach it with a tightly specified theory of individual preferences in and out of marriage; or we could try to recover preferences over a set of observable characteristics such as age, education, etc. Macroeconomists tend to favor the first approach. Given my own background, I will focus on the latter. How do different subgroups of women value the characteristics of potential partners? have more educated agents become keener on highly educated partners? how does the legal framework on divorce or income taxation affect the matching patterns and welfare of different groups? what of immigration? these are all topics that have been explored in empirical applications.

Taking the Theory to the Data

Becker assumed that all members of a household sought to maximize their share of a composite good which he called "household production", and which we now call the joint surplus. Suppose that the joint surplus only depends on a single characteristic of the partners (call it education), and educations are complementary inputs in household production. As explained in Chapter 5, Becker showed that under these conditions, any allocation in the core of the marriage market game must display positive assortative matching.

Under pure positive assortative matching (hereafter PAM), marriage patterns are strikingly simple: if we order men and women on two vertical axes by education, matches can be made sequentially by matching the highest education man left with the highest education woman left (unless one of the two would prefer to remain single). If any man "marries up" (that is, marries a woman of a higher education level), then no man with the same education level will "marry down"[1]. This is illustrated in Table 1 with four ordered levels of education: any non-zero entry off the diagonal is reflected in a symmetric zero entry.

	Women			
Men	Education 1	Education 2	Education 3	Education 4
Education 1	μ_{11}	0	μ_{13}	0
Education 2	μ_{21}	μ_{22}	0	μ_{24}
Education 3	0	0	μ_{33}	μ_{34}
Education 4	μ_{41}	0	0	μ_{44}

Table 1: Numbers of matches between men and women

While the data consistently show that education levels are positively correlated within couples, matching patterns rarely look like Table 1. Cells may be empty because a sample is small, but they are typically nonzero in the population. There are at least two ways to reconcile the theory and the observations. One is to appeal to frictions. If it takes time and effort to find a mate, it is optimal to stop searching once a "good enough" partner is found. Given random meetings, this acceptable partner may be well educated or not.

An alternative is to recognize that individuals differ in their preferences in ways that the analyst cannot observe. College graduates tend to prefer to match with college graduates, but they differ in the intensity of this educational preference. A

[1] This statement also applies after interchanging "down" and "up", or "man" with "woman".

college graduate whose educational preference is moderate may end up in a match with a high-school graduate who likes college graduates, and compensate by getting a larger share of the joint surplus. Preferences may also differ along other lines: religion, physical characteristics etc, which are unlikely to be all recorded in our datasets.

These two ways of explaining the observed matching patterns are obviously compatible, and in fact complementary. This chapter will emphasize unobserved heterogeneity; Chapter 11 has much more to say on matching with frictions.

Identifying the Parameters

Let us return to the example that underlies Table 1, where education is the only payoff-relevant observed characteristics. Suppose that we only observe the μ_{xy} numbers for all pairs of education levels $x, y \in \{1, 2, 3, 4\}$. To further simplify inference, let these numbers be known for the population, rather than a random sample. If this is all the information we have—so that transfers are not observed, as is typical with data on marriages—then we can only recover at most 16 parameters of the distribution of the joint surplus. This clearly restricts how unobserved heterogeneity can be introduced in the joint surplus. It should not come as a surprise; we are dealing with a two-sided discrete-choice problem, and we know that even the one-sided version can only be identified under restrictive assumptions[2].

Following on the pioneering contribution of Choo and Siow (2006), much of the literature has adopted a "separability" assumption that rules out interactions between the unobservables. Galichon and Salanié (2022) showed that in separable models, *without further information* the joint surplus is just identified once a distribution of the observables is specified. They also proposed inference procedures that build on their identification results, as well as a fast computational device to solve for the stable matching.

To put it differently, if all one observes is "who matches with whom" then even under separability there are very many ways to rationalize the data, and no way to

[2]Existing identification results for the discret choice model specify the distribution of the unobservables tightly, resort to a special regressor, or combine these two approaches; see e.g. Lewbel (2000) and Matzkin (2007).

test the theory[3]. As always, one must add more assumptions and/or more data to refine inference; Section 4 illustrates it on some applications.

Roadmap

Section 1 sets up the one-to-one bipartite discrete model; it introduces the crucial *separability* property and shows how separable models can be identified. Section 2 explains how to solve for equilibrium and to estimate the parameters of separable models; and in Section 3 i discuss extensions. I present some applications in Section 4. Section 5 looks at attempts to go beyond separability. Finally, I briefly discuss models with frictions in Section 6.

In the following, a *matching* consists of a list of *matches*: who ends up partnered with whom, and/or remains single. I use the term *matching patterns* to refer to the observed data: we may for instance only observe how many college-educated women marry a man with only a high-school degree, but not their income. Tildes on notation generally denote that the corresponding variable depends on unobserved heterogeneity. The subscript 0 for a match denotes that the individual is single. I denote $|S|$ the number of elements of a finite set S. I write $m \in x$ for "m belongs to group x". Finally, I use bold symbols to denote vectors or matrices, and I use the dot notation: $Z_{x\cdot}$ is row x and $Z_{\cdot y}$ is column y of the matrix $\boldsymbol{Z}$.

I have developed a Python package called `cupid_matching` that solves and estimates separable models of matching with perfect transferable utility. It is freely available on the usual repositories[4].

1 Separable Models

Becker's (traditional) marriage market remains the simplest framework to study matching with perfectly transferable utility. In the jargon of the field, it is *one-to-one*: each individual has at most one partner. It is also *bipartite*: the two partners in a match belong to separate populations. As these two features simplify the analysis

[3]The model without unobserved heterogeneity is the one exception, as Table 1 shows.

[4]See Salanié (2023).

and the notation, I will start with this model.

1.1 The One-to-One Bipartite Model

The population consists of men $m \in M$ and women $w \in W$. Each man m has a *full type* $\tilde{x}_m$ that is perfectly observed by all women; on the other hand, the analyst only observes that $\tilde{x}_m$ belongs to a *group* x_m. As an example, x_m could be the education level and $\tilde{x}_m$ could be the pair formed by the education level and the religion. Similarly, each woman w has a full type $\tilde{y}_w$ which all men observe, but the analyst only observes her group y_w.

Note that the unobserved characteristics could be discrete or continuous, and multidimensional; they could also be much more complex, for instance a whole profile of preferences over partners. The full type needs to include all characteristics that are payoff-relevant in the sense that they contribute to the joint surplus in a couple. If for instance the joint surplus is higher when partners have similar political beliefs, then the full types should include them. The assumption that each individual can observe the full type of all of his/her potential partners is obviously very strong. The literature has not yet found a way to deal with private information in a manner that is tractable enough for empirical studies.

In many datasets, the characteristics of the partners that are known at the time of marriage are discrete-valued, as in the case of education and other demographics. Income may be recorded as continuous, when available; but since it is typically measured at the time of the survey rather than when the match is formed, it is less directly useful. Sometimes the data contains information on physical characteristics or personality scores, which are (somewhat) persistent over time. These should be treated as continuous. For now we assume that there are a finite number of groups ($x \in X$ and $y \in Y$); we will deal with continuous groups in Section 3.5.

We summarize this in Assumption 1.

Assumption 1 (The Bipartite Discrete Model)**.** *The population consists of men $m \in M$ and women $w \in W$. A man m (resp. a woman w) has an observed type $x_m \in X$ (resp. $y_w \in Y$) and a full type $\tilde{x}_m$ (resp. $\tilde{y}_w$). There are n_x men in group x and m_y women in group y. We will collect these group numbers in a vector $\boldsymbol{q} = (\boldsymbol{n}, \boldsymbol{m})$.*

A match between a man m and w generates a joint utility $\tilde{\Phi}_{mw}$. Single men (resp. women) have utilities $\tilde{\Phi}_{m0}$ and $\tilde{\Phi}_{0w}$.

We denote $X_0 = X \cup \{0\}$ and $Y_0 = Y \cup \{0\}$.

1.2 The Stable Matchings

Now consider a hypothetical match between man m and woman w. With perfectly transferable utility, the utility levels that they can achieve within this couple must add up to the joint utility $\tilde{\Phi}_{mw}$. We impose that any observed matching be *stable*. In this context, this means (see Chapter 5) that if we denote $\tilde{u}_m$ and $\tilde{v}_w$ the utility levels that any man m and any woman w obtain at an observed matching, the following inequalities hold:

$$\begin{aligned} &\tilde{u}_m \geqslant \tilde{\Phi}_{m0} \ \text{ for all } \ m \\ &\tilde{v}_w \geqslant \tilde{\Phi}_{0w} \ \text{ for all } \ w \\ &\tilde{u}_m + \tilde{v}_w \geqslant \tilde{\Phi}_{mw} \ \text{ for all } \ (m, w). \end{aligned}$$

The first set of inequalities imposes that a man cannot be left with a lower level of utility than he would get on his own. The second set of inequalities is the equivalent statement for women. Finally, the last line states that the sum of the utilities of a man m and woman w cannot be lower than the joint utility they would obtain by matching together. We call these $|M| + |W| + |M| \times |W|$ inequalities the *stability constraints*.

As explained in Chapter 5, the utilities attained at a stable matching minimize the sum of joint utilities

$$\sum_{m \in M} \tilde{u}_m + \sum_{w \in W} \tilde{v}_w \tag{1}$$

under the stability constraints. Moreover, the Lagrange multipliers of the stability constraints at the minimum $\tilde{\mu}_{m0} \geqslant 0$, $\tilde{\mu}_{0w} \geqslant 0$, and $\tilde{\mu}_{mw} \geqslant 0$ can be interpreted as a random matching, with $\tilde{\mu}_{m0}$ (resp. $\tilde{\mu}_{0w}$) representing the probability that man m (resp. woman w) is single and $\tilde{\mu}_{mw}$ the probability that they form a match. This random matching maximizes the sum of joint utilities

$$\sum_{mw} \tilde{\mu}_{mw} \tilde{\Phi}_{mw} + \sum_{m} \tilde{\mu}_{m0} \tilde{\Phi}_{m0} + \sum_{w} \tilde{\mu}_{0w} \tilde{\Phi}_{0w} \tag{2}$$

under the *feasibility constraints*

$$\sum_w \tilde{\mu}_{mw} + \tilde{\mu}_{m0} = 1 \text{ for all } m$$
$$\sum_m \tilde{\mu}_{mw} + \tilde{\mu}_{0w} = 1 \text{ for all } w.$$

The multipliers $\tilde{u}_m$ and $\tilde{v}_w$ of these two sets of constraints in turn solve the problem (1). This makes programs (1) and (2) dual versions of each other. The former is usually called the *dual* and the latter the *primal*. The dual has much fewer variables than the primal ($|M| + |W|$ rather than $|M| + |W| + |M| \times |W|$); on the other hand, the primal has much fewer constraints (again, $|M| + |W|$ rather than $|M| + |W| + |M| \times |W|$). These are important considerations when choosing which one to solve.

It is easy to see that the feasible sets of these two programs are non-empty. Since both programs are linear (with a linear objective function and linear constraints), the structure of their solution sets is quite simple. There is always a solution to the primal program (2) for which the matching probabilities are degenerate, in that they all equal 0 or 1. Generically, this is the unique solution. On the other hand, the utilities $\tilde{\boldsymbol{u}}$ and $\tilde{\boldsymbol{v}}$ cannot be uniquely determined: unless all men and women remain single, it is always possible to slightly perturb the sharing of joint utilities within couples without breaking the stability constraints.

1.3 The Data

I will always assume on this chapter that the analyst has data on "who matches whom". Since full types have an unobserved component, this data can only record matches by groups. A typical dataset will be sampled at the household levels. A household $h = 1, \ldots, H$ may consist of a single man of group x_h, a single woman of group y_h, or a married couple (x_h, y_h). We denote

$$\hat{\mu}_{xy} = \sum_{h=1}^{H} \mathbb{1}(x_h = x, y_h = y)$$

the number of observed matches between a man of group x and a woman of group y. Sometimes only data on observed pairs are available; unless specified otherwise, we will assume that the analyst also observes the numbers of singles $\hat{\mu}_{x0}$ and $\hat{\mu}_{0y}$. The numbers of men of group x (resp. women of group y) then are estimated by

$$\hat{n}_x = \sum_{y \in Y} \hat{\mu}_{xy} + \hat{\mu}_{x0}$$
$$\hat{m}_y = \sum_{x \in X} \hat{\mu}_{xy} + \hat{\mu}_{0y}.$$

The variance-covariance matrix of these estimators obtains easily. For instance,

$$\widehat{\text{cov}}(\hat{\mu}_{xy}, \hat{\mu}_{zt}) = \hat{\mu}_{xy} \left(\mathbb{1}(x = z)\mathbb{1}(y = t) - \frac{\hat{\mu}_{zt}}{H} \right).$$

These formulæ must be corrected if sampling weights are used.

Some datasets contain additional information: the salary paid by a firm to an employee, number of children in a marriage, etc. In Section 4, we will briefly discuss the case when data on transfers or other measurements is available.

1.4 Separability

In principle, given a specification of the joint utilities and the utilities when single $\tilde{\mathbf{\Phi}}$, it is possible to estimate the model using a brute-force approach. Suppose that we parameterize $\tilde{\Phi}_{mw} = F(x_m, y_w, \xi_{mw}; \boldsymbol{\theta})$, where F is a known function, $\boldsymbol{\theta}$ is a vector of unknown parameters, and ξ_{mw} has a distribution whose unknown parameters, if any, are subsumed in $\boldsymbol{\theta}$. For each value of $\boldsymbol{\theta}$, one could solve either the primal or the dual program to obtain the matching probabilities $\tilde{\mu}(x_m, y_w; \boldsymbol{\xi}, \boldsymbol{\theta})$, and then take their expectations over the matrix of random terms $\boldsymbol{\xi}$ to obtain matching patterns $\mu(x_m, y_w; \boldsymbol{\theta})$. All that is left is to choose $\boldsymbol{\theta}$ to minimize a distance between these matching patterns and the observed matching patterns $\hat{\mu}(x_m, y_w)$. Taking the Kullback-Leibler distance, for instance, would give the maximum-likelihood estimator.

This approach has three drawbacks. First, it can be very costly if the sets of potential partners are large; it requires solving for the equilibrium many times, and approximating high-dimensional integrals. Second, it gives very little insight on the source of identification of the parameters $\boldsymbol{\theta}$. Third, how are we to specify the distribution of the random matrix $\boldsymbol{\xi}$? The unobserved components of the full types of man m and woman w may a priori interact in complicated ways with the observed components and with each other. As we have known since Becker (1973) that interactions between characteristics drive matching patterns, this raises concerns that specification details may play a very large role.

For all of these reasons, most of the literature has adopted the *separability* assumption implicit in Choo and Siow (2006) and defined in Chiappori, Salanié, and Weiss (2017). This states that *conditionally on the groups of the partners*, full types do not interact in generating joint surplus. As an illustration, remember the example in the introduction where education is observed by the analyst and religion is not. Separability allows the education levels to affect the joint utility freely; it does not restrict the interaction between the education of the husband and the religion of the wife, or between the education of the husband and the religion of the wife; but it rules out any interaction between the religion of the two partners. It would clearly not be a credible assumption in societies where religious identities are very strong. On the other hand, it may be more appropriate in more secular settings. This example also shows that the separability assumption is less strong when the data contains more relevant covariates. Moreover, exploratory simulations reported in Chiappori, Nguyen, and Salanié (2019) suggest that fitting a separable model to a data-generating process that is moderately non-separable only creates small biases.

Assumption 2 states separability more formally. Part (i) is the crucial element, as discussed above. Equation (3) looks like an analysis-of-variance decomposition with a "missing" residual—the interaction that we ruled out. Part (vi) is a weak technical restriction.

Assumption 2 (Separability)**.** *There exist a* $(|X|, |Y|)$ *matrix* $\mathbf{\Phi}$ *and random terms* $\boldsymbol{\varepsilon}$ *and* $\boldsymbol{\eta}$ *such that*

(i) the joint utility from a match between a man m *of group* $x \in X$ *and a woman* w *of group* $y \in Y$ *is*

$$\tilde{\Phi}_{mw} = \Phi_{xy} + \varepsilon_{my} + \eta_{xw}. \tag{3}$$

We call the matrix $\mathbf{\Phi}$ *the* joint surplus[5].

[5] I follow the literature here. Strictly speaking, the joint surplus is $\tilde{\Phi}_{mw} - \tilde{\Phi}_{m0} - \tilde{\Phi}_{0w}$; Φ_{xy} only measures the part of the joint utility from a match that comes from the observed characacteristics of the partners.

(ii) *the utility of a single man m is $\tilde{\Phi}_{m0} = \varepsilon_{m0}$,*

(iii) *the utility of a single woman w is $\tilde{\Phi}_{0w} = \eta_{0w}$,*

(iv) *conditional on $m \in x$, the random vectors $\boldsymbol{\varepsilon}_m = (\varepsilon_{my})_{y \in Y_0}$ are independent across m; they have probability distribution $\mathbb{P}_x$,*

(v) *conditional on $w \in y$, the random vectors $\boldsymbol{\eta}_w = (\eta_{xw})_{x \in X_0}$ are independent across w; they have probability distribution $\mathbb{Q}_y$,*

(vi) *the random variables*

$$\max_{y \in Y_0} |\varepsilon_{my}| \quad \textit{and} \quad \max_{x \in X_0} |\eta_{xw}|$$

have finite expectations under $\mathbb{P}_x$ and $\mathbb{Q}_y$ respectively.

Note that Assumption 2 does not allocate the $\boldsymbol{\varepsilon}$ and $\boldsymbol{\eta}$ terms to either side of the market. One could interpret η_{xw}, for instance, as an idiosyncratic preference of woman w for men of group x, as a good fit between them in household production, as a common attraction of men in group x for woman w, or any combination of these and other motives. Since the model primitives are the *joint* utilities, these distinctions do not affect the solution.

Imposing the separability assumption will allow us to get transparent identifying equations that lead directly to simple estimation procedures. To see this, start from the problem facing a man m in group x. He could either stay single and obtain utility ε_{m0}, or match with a woman w and share the joint utility $\tilde{\Phi}_{mw}$ with her. Let $\tilde{v}_w$ denote the utility that woman w expects at a stable matching, so that man m can only obtain $\tilde{\Phi}_{mw} - \tilde{v}_w$ for himself in their match. To select his best option, this man will solve

$$\max \left(\varepsilon_{m0}, \max_{w} \left(\tilde{\Phi}_{mw} - \tilde{v}_w \right) \right);$$

the value of this program represents the utility $\tilde{u}_m$ he can expect at a stable matching. Using Assumption 2.(i), we rewrite

$$\begin{aligned} \tilde{u}_m &= \max \left(\varepsilon_{m0}, \max_{y \in Y} \max_{w \in y} \left(\Phi_{xy} + \varepsilon_{my} + \eta_{xw} - \tilde{v}_w \right) \right) \\ &= \max \left(\varepsilon_{m0}, \max_{y \in Y} \left(\Phi_{xy} + \varepsilon_{my} + \max_{w \in y} \left(\eta_{xw} - \tilde{v}_w \right) \right) \right). \end{aligned}$$

Now define $V_{xy} = \min_{w \in y}(\tilde{v}_w - \eta_{xw})$. We obtain

$$\begin{aligned}\tilde{u}_m &= \max\left(\varepsilon_{m0}, \max_{y \in Y}\left(\Phi_{xy} - V_{xy} + \varepsilon_{my}\right)\right) \\ &= \max_{y \in Y_0}\left(\Phi_{xy} - V_{xy} + \varepsilon_{my}\right),\end{aligned} \tag{4}$$

with the convention that $\Phi_{x0} = V_{x0} = 0$. The interpretation is that man m will choose to match with a woman whose group y attains the maximum in (4) (or remain single if the maximum is attained at $y = 0$.)

For a woman w in group y, we get in the same way

$$\tilde{v}_w = \max_{x \in X_0}\left(\Phi_{xy} - U_{xy} + \eta_{xw}\right) \tag{5}$$

with $U_{xy} = \min_{m \in x}(\tilde{u}_m - \varepsilon_{my})$ and $\Phi_{0y} = U_{0y} = 0$.

The third part of the stability constraints implies that for any man $m \in x$ and woman $w \in y$,

$$(\tilde{u}_m - \varepsilon_{my}) + (\tilde{v}_w - \eta_{xw}) \geqslant \Phi_{xy};$$

taking the minimum over $m \in x$ and $w \in y$ gives $U_{xy} + V_{xy} \geqslant \Phi_{xy}$. If m and w do match at the stable matching, the utility possibility frontier $\tilde{u}_m + \tilde{v}_w = \tilde{\Phi}_{mw}$ gives $U_{xy} + V_{xy} = \Phi_{xy}$.

To summarize:

Theorem 1 (Discrete Choice Representation)**.** *Normalize utilities by $\Phi_{x0} = U_{x0} = V_{0y} = \Phi_{0y} = 0$ for all $x \in X$ and $y \in Y$.*

(i) There exist two $(|X|, |Y|)$ matrices of numbers $\boldsymbol{U}$ and $\boldsymbol{V}$ such that a man $m \in x$ (resp. a woman $w \in y$) remains single or marries a woman of group y (resp. a man of group x), depending on the maximizer in (4) (resp. in (5)).

(ii) $V_{xy} \geqslant \Phi_{xy} - U_{xy}$ for any pair (x, y) of groups.

(iii) Moreover, $V_{xy} = \Phi_{xy} - U_{xy}$ for any pair of groups such that $\mu_{xy} > 0$.

It is clear from (4) and (5) that separability implies some form of indifference: if woman $w \in y$ opts for a match with a man of group x, then the unobserved characteristics of this man do not matter to her. Separability does allow for a restricted form of "matching on unobservables", however: at a stable matching, such a woman is more likely to marry man $m \in x$ than man $m' \in x$ if man m has a stronger preference for women in group y—that is, $\varepsilon_{my} > \varepsilon_{m'y}$. To return to the education/religion example: if a religious woman marries an educated man at the stable matching, his religious beliefs do not affect her utility—but part of the reason is that this man will have a preference for religious wives.

Choo and Siow (2006) called the numbers U_{xy} and V_{xy} "systematic net gains". This is somewhat misleading as they do not represent the utility of any individual (or an average utility) in a stable matching; they are only useful notation. I will call them "systematic values" in what follows. On the other hand, the numbers $\tilde{u}_m$ and $\tilde{v}_w$ are utility values.

1.5 Identification

Our objective here is to identify the joint surplus $\mathbf{\Phi}$ and the distributions of the utilities $\tilde{\boldsymbol{u}}$ and $\tilde{\boldsymbol{v}}$ from an observed matching $\hat{\boldsymbol{\mu}}$, given (for now) assumed distributions $\mathbf{P}_x$ and $\mathbf{Q}_y$. Galichon and Salanié (2022) developed a series of results that achieve this goal.

Theorem 1 shows that separability reduces the dimensionality of the problem from $|M| + |W|$ (the utilities of all individuals at the stable matching) to the $|X| \times |Y|$ values of the U_{xy}. This is typically a much smaller number. Moreover, Theorem 1 will allow us to turn the linear programming program (2) into a differentiable, strictly convex minimization problem and to use convex duality to identify the parameters of the joint surplus.

To see this, let $\boldsymbol{U}$ be the matrix in Theorem 1, with $U_{x0} = 0$ for all groups x. Then the average value function of men in group x can be defined as

$$\tilde{G}_x(\boldsymbol{U}_{x\cdot}) = \frac{1}{n_x} \sum_{m \in x} \max_{y \in Y_0} (U_{xy} + \varepsilon_{my}).$$

Since it is the average of the maximum of a set of functions that are linear in the U_{xy}, it is a convex function. To simplify the exposition, I focus here in the "large market limit"[6].

[6] I describe the finite market version in Appendix B.

Assumption 3 (Large Market Limit)**.** *The population contains a continuum of individuals of each group. The notations n_x and m_y now refer to the masses of the groups.*

Moreover, the distributions $\mathbf{P}_x$ (resp. $\mathbf{Q}_y$) are atomless and have full support on $\mathbb{R}^{|Y|+1}$ (resp. $\mathbb{R}^{|X|+1}$).

In this large market limit, the average value becomes an expected value:

$$G_x(\boldsymbol{U}_{x\cdot}) = E_{\mathbf{P}_x} \max_{y \in Y_0}(U_{xy} + \varepsilon_{my})$$

and the function G_x is differentiable and strictly convex. A simple application of the envelope theorem shows that its partial derivatives

$$\mu^M_{y|x} \equiv \frac{\partial G_x}{\partial U_{xy}}(\boldsymbol{U}_{x\cdot}) \tag{6}$$

equal the choice probabilities for men of group x. This result[7] is of great interest for us, since these choice probabilities are simply the ratios of the matching patterns to the masses of the groups: $\mu^M_{y|x} = \mu_{xy}/n_x$. Even more importantly, this relationship can be inverted to express the unknown systematic values (U_{xy}) as a function of the observed matching patterns (μ_{xy}). This inversion relies on *convex duality*; Appendix A gives the results from convex analysis that we need here.

Given a set of non-negative numbers $\boldsymbol{\mu}^M_{\cdot|x} \equiv (\mu^M_{y|x})_{y \in Y}$ whose sum does not exceed one, we define the *Legendre-Fenchel transform* of the function G_x by

$$G^*_x(\boldsymbol{\mu}^M_{\cdot|x}) = \max_{U_{x0}=0, U_{x1}, \ldots, U_{x,|Y|}} \left(\sum_{y \in Y} \mu^M_{y|x} U_{xy} - G_x(\boldsymbol{U}_{x\cdot}) \right);$$

we assign it the value $+\infty$ if $\sum_{y \in Y} \mu^M_{y|x} > 1$ (which would imply a negative number of single men in group x). The function G^*_x is another maximum of linear functions, so that it is convex. Given Assumption 3, it is in fact strictly convex and differentiable. Using the envelope theorem again, its derivatives are

$$U_{xy} = \frac{\partial G^*_x}{\partial \mu^M_{y|x}}(\boldsymbol{\mu}^M_{\cdot|x}). \tag{7}$$

[7]It is called the Daly-Williams-Zachary theorem in the discrete choice literature.

Equations (6) and (7) are dual to each other: the former recovers choice probabilities $\boldsymbol{\mu}^M_{\cdot|x}$ from the systematic values $\boldsymbol{U}_{x\cdot}$ and the latter inverts the process to identify the systematic utility values from the choice probabilities, which are observed in the data by simple counting over households $h \in H$:

$$\hat{\mu}^M_{y|x} = \frac{\sum_{h=(m,w)\in H} \mathbb{1}(x_m = x, y_w = y)}{\sum_{h\in H} \mathbb{1}(x_m = x)}.$$

We can do exactly the same thing with the women's choice problem; substituting G_x with H_y and $\boldsymbol{U}$ with $\boldsymbol{V}$, we identify the systematic values $\boldsymbol{V}$ by

$$V_{xy} = \frac{\partial H^*_y}{\partial \mu^W_{x|y}}(\boldsymbol{\mu}^W_{\cdot|y}). \tag{8}$$

This already allows us to evaluate the utility attained by any individual, using Theorem 1. Moreover, it is easy to see that under Assumption 3, no cell (x, y) of the stable matching can be empty in the population. The last part of Theorem 1 shows that the matrix $\boldsymbol{\Phi}$ is simply the sum of $\boldsymbol{U}$ and $\boldsymbol{V}$ and is therefore identified:

$$\Phi_{xy} = \frac{\partial G^*_x}{\partial \mu^M_{y|x}}(\boldsymbol{\mu}^M_{\cdot|x}) + \frac{\partial H^*_y}{\partial \mu^W_{x|y}}(\boldsymbol{\mu}^W_{\cdot|y}). \tag{9}$$

In equilibrium, the choice probabilities of men and women must be consistent:

$$n_x \mu^M_{y|x} = m_y \mu^W_{x|y} = \mu_{xy}.$$

This allows us to define the *generalized entropy* of a matching:

$$\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{q}) = -\sum_{x\in X} n_x G^*_x(\boldsymbol{\mu}_{x\cdot}/n_x) - \sum_{y\in Y} m_y H^*_y(\boldsymbol{\mu}_{\cdot y}/m_y) \tag{10}$$

and to state our identification result as Theorem 2, which is proved more rigorously in Galichon and Salanié (2022).

Theorem 2 (Identifying the Joint Surplus). *Under Assumptions 1, 2, and 3, the stable matching $\boldsymbol{\mu}$ is unique; it solves the following globally convex program:*

$$\mathcal{W}(\boldsymbol{q}) = \max_{\boldsymbol{\mu}} \left(\sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} \mu_{xy}\Phi_{xy} + \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{q}) \right). \tag{11}$$

The solution is characterized by the first-order conditions

$$\Phi_{xy} = -\frac{\partial \mathcal{E}}{\partial \mu_{xy}}(\boldsymbol{\mu}, \boldsymbol{q}). \tag{12}$$

At the stable matching, a man $m \in x$ attains utility

$$\tilde{u}_m = \max\left(\varepsilon_{m0}, \max_{y \in Y}\left(U_{xy} + \varepsilon_{my}\right)\right), \quad \textit{with} \;\; U_{xy} = \frac{\partial G_x^*}{\partial \mu^M_{y|x}}(\boldsymbol{\mu}^M_{\cdot|x}). \tag{13}$$

Man m is single if the maximum is achieved by ε_{m0}, married to a woman $w \in y$ if the maximum is achieved at $y \in Y$. This woman attains utility

$$\tilde{v}_w = \max\left(\eta_{0w}, \max_{x \in X}\left(V_{xy} + \eta_{xw}\right)\right), \quad \textit{with} \;\; V_{xy} = \frac{\partial H_y^*}{\partial \mu^W_{x|y}}(\boldsymbol{\mu}^W_{\cdot|y})$$

and the maximum is attained at x.

The average utility of men of group x (resp. women of group y) ar the stable matching is $u_x \equiv G_x(\boldsymbol{U}_{x\cdot})$ (resp. $v_y \equiv H_y(\boldsymbol{v}_{\cdot y})$).

The quantity being maximized in (11) is the total joint utility generated by a matching $\boldsymbol{\mu}$. The generalized entropy is strictly concave since its components G_x^* and H_y^* are strictly convex; and it is differentiable like them. By construction, its value is minus infinity if the matching is not feasible: if we had say $\sum_{y \in Y} \mu_{xy} > n_x$, then the value of G_x^* would be plus infinity. If there were no unobserved heterogeneity, we would have $G_x(\boldsymbol{U}_{x\cdot}) = \max_{y \in Y_0} U_{xy}$; it is easy to see that G_x^* would be zero for all feasible matchings. The total joint utility would simply be the sum of the values of "matching on groups":

$$\sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy}.$$

The generalized entropy adds to it the value of "matching on unobservables". If men in group x have choice probabilities $\boldsymbol{\mu}^M_{\cdot|x}$ and utility values $\boldsymbol{U}_{x\cdot}$, then by definition

$$-G_x^*(\boldsymbol{\mu}^M_{y|x}) = G_x(\boldsymbol{U}_{x\cdot}) - \sum_{y \in Y} \mu^M_{y|x} U_{xy}$$

which is the difference between the expected utility of men of group x and what they would get if matching only on groups. Summing over men and women of all groups gives

$$\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{q}) = \left(\sum_{x \in X} n_x G_x(\boldsymbol{U}_{x\cdot}) + \sum_{y \in Y} m_y H_y(\boldsymbol{V}_{\cdot y}) \right) \\ - \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} \mu_{xy} \Phi_{xy}.$$

The term between parentheses is the total joint utility generated by the matching; the term on the second line is the joint utility generated by matching on groups only. The value of the generalized entropy comes from the fact that men and women select the group of their partners on the basis of their unobserved heterogeneity terms. If the $\boldsymbol{\varepsilon}$ and $\boldsymbol{\eta}$ have non-negative means, this contributes a positive term to the total joint utility. Appendix C gives formulæ for the generalized entropy functions in several useful models.

1.6 The Limits of Identification

Equation (12) shows that *for any choice of the distributions* $\mathbf{P}_x$ *and* $\mathbf{Q}_y$ *of the unobserved heterogeneity terms,* the joint surplus $\boldsymbol{\Phi}$ can be recovered from the derivatives of the generalized entropy, whose precise form depends on this choice of distributions. This is of course a major "if". At the end of the day, the data boils down to $|X| \times |Y|$ numbers: the marriage patterns $(\boldsymbol{\mu}_{xy})$. Therefore we can only hope to point-identify $|X| \times |Y|$ parameters. Some of these parameters can be used to specify the matrix $\boldsymbol{\Phi}$, and some others for the distributions $\mathbf{P}_x$ and $\mathbf{Q}_y$. As always, there are several ways to (partially) go around this limitation. Gualdani and Sinha (2023) show how the joint surplus can be partially identified if shape or symmetry restrictions are imposed on the distributions of the unobserved heterogeneity. Rather than adding assumptions, one can also add data, if available. A popular way of doing so is to pool data on several, isolated matching markets and to restrict the variation in $\boldsymbol{\Phi}$ across markets. We will see examples of this approach in Section 4.

1.7 The Logit Model

To make this concrete, I turn to the pioneering contribution of Choo and Siow (2006). They proposed a model that has become very popular in applications. It satisfies

Assumptions 1, 2, and 3. In addition, Choo and Siow chose distributions $\mathbf{P}_x$ and $\mathbf{Q}_y$ that lead to the usual multinomial logit formulae.

Assumption 4 (Multilogit Heterogeneity)**.** *For each x (resp. each y), the distribution $\mathbf{P}_x$ (resp. $\mathbf{Q}_y$) is the centered and standardized Gumbel (type I extreme value)* $(|Y|+1)-$ *(resp.* $(|X|+1)$*-) dimensional distribution.*

Given this distributional assumption, the expected value function takes the familiar "logexp" form:

$$G_x(\boldsymbol{U}_{x\cdot}) = \log\left(1 + \sum_{y\in Y} \exp(U_{xy})\right)$$

and it is easy to see that its Legendre-Fenchel transform is

$$G_x^*(\boldsymbol{\mu}^M_{\cdot|x}) = \sum_{y\in Y} \mu^M_{y|x} \log \mu^M_{y|x} + \mu^M_{0|x} \log \mu^M_{0|x}$$

where $\mu^M_{0|x} = 1 - \sum_{y\in Y} \mu^M_{y|x}$ is the probability that a man of group x is single. This is, of course, simply (minus) the entropy of the distribution $\boldsymbol{\mu}^M_{\cdot|x}$. This justifies the name Galichon and Salanié (2022) chose for the generalized entropy, which is in this specific model equals

$$\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{q}) = -2 \sum_{x\in X, y\in Y} \mu_{xy} \log \mu_{xy} + \sum_{x\in X} n_x \log n_x + \sum_{y\in Y} m_y \log m_y.$$

Equation (12) identifies the joint surplus from the very simple *Choo and Siow formula*:

$$\Phi_{xy} = \log \frac{\mu^2_{xy}}{\mu_{x0}\mu_{0y}}; \tag{14}$$

the utility attained by man $m \in x$ for instance is

$$\tilde{u}_m = \max_{y\in Y_0} \left(\log \frac{\mu_{xy}}{\mu_{x0}} + \varepsilon_{my} \right).$$

and he matches with a woman in the group y that achieves the maximum (or stays single if that y is 0). This woman gets utility

$$\tilde{v}_w = \max_{x\in X_0} \left(\log \frac{\mu_{xy}}{\mu_{0y}} + \eta_{xw} \right),$$

which she attains by choosing group x.

In this model, the expected utilities take a very simple form: men of group x get $u_x = -\log \mu^M_{0|x}$, so that expected utilities are ranked like the probabilities of marriage. This is a special feature of the multinomial logit: at the stable matching, the expected utility of a man of group x who is matched with a woman of group y also takes the value $-\log \mu^M_{0|x}$, no matter what the value of y is[8].

1.8 Comparative Statics

Theorem 2 easily yields some useful comparative statics result. I only mention a few here; the reader may turn to Galichon and Salanié (2017) and Galichon and Salanié (2022) for more.

First, all separable models exhibit constant returns to scale: if the numbers (or masses) of men and women in each group are multiplied by the same positive number k, then the stable matching patterns will be multiplied by k.

Applying the envelope theorem to (11) shows that the derivative of the total joint utility with respect to the number of men in a given group x is the expected utility $G_x(\boldsymbol{U}_{x\cdot})$ of men of group x, as one would expect. More interestingly, the value $\mathcal{W}$ of the total joint utility in program (11) is convex in $\boldsymbol{q}$, so that its Hessian is positive definite. The usual restrictions on symmetry and on the positivity of its minors translate directly into restrictions on the first derivatives of $\mathcal{W}$, which are the expected utilities of the different groups.

Specific separable models entail their own additional testable predictions. For instance, Graham (2013) showed that at any stable matching in the Choo and Siow (2006) model, have

$$\log \frac{\mu_{xy}\mu_{x'y'}}{\mu_{x'y}\mu_{xy'}} = \Phi_{xy} + \Phi_{x'y'} - \Phi_{xy'} - \Phi_{x'y} \tag{15}$$

for all (x, y, x', y').

Therefore changes in $\boldsymbol{\Phi}$ that leave the right-hand side unchanged should have zero effect on the left-hand side.

[8] In the notation of Section 1.2: $E(\tilde{u}_m | m \in x, \tilde{\mu}_{mw} > 0, w \in y) = E(\tilde{u}_m | m \in x, \tilde{\mu}_{m0} < 1)$ for all x and y. de Palma and Kilani (2007) shows (in the one-sided discrete choice model) that this only obtains with the multinomial logit.

1.9 Interpreting Matching Patterns

As a useful joint product, using structural models of matching allows us to understand the limits on ANOVA decompositions of matching patterns. These have often been used to interpret changes over time, or differences across groups or countries; and to infer changes in preferences for homogamy.

An ANOVA decomposition is akin to a regression with fixed effects. Say we observe discrete-valued characteristics x for men and y for women. We could run a regression of the number μ_{xy} of matches observed in each (x, y) cell as

$$\mu_{xy} = \sum_{k=1}^{|X|} b_k \mathbb{1}(x = x_k) + \sum_{l=1}^{|Y|} c_l \mathbb{1}(y = y_l) + \xi_{xy}.$$

By construction, the residual ξ_{xy} measures the contribution of the interactions of the characteristics of the partners to the likelihood that they match. Its interpretation is far from straightforward, however. Take the logit model as an example. Formula 14 gives

$$2 \log \mu_{xy} = \log \mu_{x0} + \log \mu_{0y} + \Phi_{xy}.$$

If we are interested in the *causes* of changes in matching patterns, that is in changes in the joint surplus $\mathbf{\Phi}$, the Choo and Siow model therefore suggests using an ANOVA decomposition of the *logarithm* of the matching patterns (or, as in Graham (2013) and Chiappori, Salanié, and Weiss (2017), using formula (15)). The more general point here is that ANOVA regressions cannot be interpreted in the absence of a structural model that justifies them.

Defining measures of changes in matching patterns is a related, controversial topic. Consider an apparently simple question: can we conclude from the observed matching patterns that assortative matching by education has increased in the US in recent decades? The answer may differ in the bottom and at the top of the range of education levels; it may also depend on how we group education levels. Suppose that we agree that there are only two education groups, and on the definitions of these groups. We could use the correlation between the educations of the partners; the value of a χ^2 test of independence; a log-likelihood ratio; or any of several other proposals in this literature. Even with this 2-by-2 case, they may not give the same answer, and some of them have counterintuitive properties. Siow (2015), Chiappori, Salanié, and Weiss

(2017), and Chiappori, Costa-Dias, and Meghir (2020) argue in favor of using the estimate of the joint utility $\mathbf{\Phi}$ that results from (14).

2 Computation and Inference in Separable Models

I show here how to solve for the equilibrium and to estimate the model in the one-to-one, bipartite, discrete model. Section 3 discusses extensions to other classes of separable models.

2.1 Computing the Stable Matching

For given $\mathbf{\Phi}$ and $\boldsymbol{q}$, one could solve the program (11) or the first-order conditions (12) to compute the stable matching patterns $\boldsymbol{\mu}$. There are much better ways, however. Galichon and Salanié (2022) proposed a version of the Iterative Projection Fitting Procedure (IPFP[9]).

The algorithm is best illustrated on the Choo and Siow (2006) model. Remember the adding up equation for men of group x:

$$\sum_{y \in Y} \mu_{xy} + \mu_{x0} = n_x.$$

Denote $a_x \equiv \sqrt{\mu_{x0}}$ and $b_y \equiv \sqrt{\mu_{0y}}$. Given (14), this becomes the quadratic equation

$$a_x^2 + a_x \sum_{y \in Y} b_y \exp(\Phi_{xy}/2) = n_x.$$

Start from an initial value $\boldsymbol{b}^{(0)}$. for the vector of b_y values. We can easily solve for a_x; in fact, we can do it in parallel for all values of x and get a vector $\boldsymbol{a}^{(1)}$ such that $(\boldsymbol{a}^{(1)}, \boldsymbol{b}^{(0)})$ and the implied matching patterns $\mu_{xy}^{(1)} = a_x^{(1)} b_y^{(0)} \exp(\Phi_{xy}/2)$ are feasible on the men's side (but not the women's). Using the adding up equations on the women's side:

$$b_y^2 + b_y \sum_{x \in X} a_x^{(1)} \exp(\Phi_{xy}/2) = m_y,$$

we solve for a new vector $\boldsymbol{b}^{(2)}$ and we now get feasibility on the women's side, but not on the men's side.

[9] This is also known as Sinkhorn's algorithm or RAS matrix completion in different contexts.

Galichon and Salanié (2022) show that this iterative process converges globally towards the stable matching. In practice, convergence is very fast. The algorithm can be seen as a tâtonnement on the "prices" a_x and b_y for the "goods" represented by the groups of men and women, with the particular feature that each iteration clears $|X|$ or $|Y|$ markets. It can be extended to all separable models, but each individual iteration may be more complicated than in the logit model[10]. Solvers for heteroskedastic or nested logit models are available in my `cupid_matching` Python package.

2.2 Estimating Separable Models

Some simple separable models can be estimated non-parametrically. This was the approach used by Choo and Siow (2006). They grouped individuals by age only and they applied equation (14). The limits of this method are obvious: the curse of dimensionality hits quickly since observable characteristics on both sides interact; and one may also want to allow for more flexible specifications than the underlying (parameter-free) multinomial logit.

A parametric model needs to specify the distributions of the unobserved heterogeneity ($\mathbf{P}_x$ and $\mathbf{Q}_y$) and the joint utility $\mathbf{\Phi}$. Let us denote $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ the corresponding vectors of parameters. Our goal is to recover estimates of $\boldsymbol{\lambda} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$.

2.2.1 Maximum Likelihood Estimation

Suppose that we have a way of computing the stable matching patterns $\boldsymbol{\mu}^{\boldsymbol{\lambda}}$ for any parameter vector $\boldsymbol{\lambda}$, with IPFP for instance. Given a sample of households as described in Section 1.3, the log-likelihood function is[11]

$$\sum_{h=1}^{H} \log p_h^{\boldsymbol{\lambda}}$$

where $p_h^{\boldsymbol{\lambda}}$ is the probability that household h is of the observed type, given a model with parameters $\boldsymbol{\lambda}$. This is easily computed from the stable matching. The only

[10] See Galichon and Salanié (2022) for details.

[11] Technically, it is the log-likelihood function conditional on the observed group numbers (the $\hat{n}_x$ and $\hat{m}_y$, which are used to compute the stable matching).

thing to note is that the number of households at the stable matching depends on the number of marriages, which varies with $\boldsymbol{\lambda}$. Denoting it by $N^{\boldsymbol{\lambda}}$, we obtain

- $p_h^{\boldsymbol{\lambda}} = \mu_{x0}^{\boldsymbol{\lambda}}/N^{\boldsymbol{\lambda}}$ if household h is a single man of group x
- if it is a single woman of group y, then $p_h^{\boldsymbol{\lambda}} = \mu_{0y}^{\boldsymbol{\lambda}}/N^{\boldsymbol{\lambda}}$
- if it is a married couple with groups x and y, then $p_h^{\boldsymbol{\lambda}} = \mu_{xy}^{\boldsymbol{\lambda}}/N^{\boldsymbol{\lambda}}$.

Maximizing the log-likelihood function gives as usual consistent, asymptotically normal, and asymptotically efficient estimates[12]. It has two drawbacks. First, it requires evaluating the stable matching patterns $\boldsymbol{\mu}^{\boldsymbol{\lambda}}$ for many values of $\boldsymbol{\lambda}$. Second, it relies on numerical optimization of a function that may have a complicated shape.

I now turn to two alternative estimation methods that are coded in my `cupid_matching` package.

2.2.2 Minimum Distance Estimation

It seems natural to start from the identifying equation (12):

$$\boldsymbol{\Phi}^{\boldsymbol{\beta}_0} + \frac{\partial \mathcal{E}^{\boldsymbol{\alpha}_0}}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{q}) = \mathbf{0},$$

where we made explicit the dependence of the $\boldsymbol{\Phi}$ matrix on $\boldsymbol{\beta}_0$ and that of the generalized entropy $\mathcal{E}$ on $\boldsymbol{\alpha}_0$ (via the distributions $\mathbf{P}_x$ and $\mathbf{Q}_y$). This can be seem as a *mixed hypothesis*:

$$\exists \boldsymbol{\lambda}_0 = (\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0), \;\; \boldsymbol{D}^{\boldsymbol{\lambda}_0}(\boldsymbol{\mu}, \boldsymbol{q}) \equiv \boldsymbol{\Phi}^{\boldsymbol{\beta}_0} + \frac{\partial \mathcal{E}^{\boldsymbol{\alpha}_0}}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{q}) = \mathbf{0},$$

where we stacked all $X \times Y$ conditions in (12) in a vector $\boldsymbol{D}^{\boldsymbol{\lambda}_0}$. Our assumptions imply that this hypothesis holds in the population. Since we have consistent estimates of both $\boldsymbol{q}$ and $\boldsymbol{\mu}$, we can plug them in[13] and fit the equations

$$\boldsymbol{D}^{\boldsymbol{\lambda}_0}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{q}}) \equiv \boldsymbol{\Phi}^{\boldsymbol{\beta}} + \frac{\partial \mathcal{E}^{\boldsymbol{\alpha}}}{\partial \boldsymbol{\mu}}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{q}}) = \mathbf{0}$$

[12]Note that $N^{\boldsymbol{\lambda}}$ is a real number not an integer, as it is the sum of the numbers of simulated couples and of simulated singles, which are themselves real numbers.

[13]Recall that since all separable matching models exhibit constant returns to scale, the derivatives of the generalized entropy are homogeneous of degree 0 in $(\boldsymbol{\mu}, \boldsymbol{q})$. This implies that the mixed hypothesis is scale-invariant.

to obtain estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. This *minimum distance estimator* is consistent and asymptotically normal.

In practice, we can both test our mixed hypothesis and estimate $\boldsymbol{\lambda}_0$ by minimizing $\|\boldsymbol{D}^{\boldsymbol{\lambda}}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{q}})\|^2_{\boldsymbol{S}}$ for some positive definite $(|X| \times |Y|, |X| \times |Y|)$ matrix $\boldsymbol{S}$. By the general theory of minimum distance estimators, we know that this yields a consistent estimator of $\boldsymbol{\lambda}$ if the model is well-specified and $\boldsymbol{\lambda}_0$ is point-identified. Moreover, if we choose $\boldsymbol{S} = \hat{\boldsymbol{\Omega}}^{-1}$ where $\hat{\boldsymbol{\Omega}}$ consistently estimates the asymptotic variance of $\boldsymbol{D}^{\boldsymbol{\lambda}_0}(\hat{\boldsymbol{\mu}}, \boldsymbol{q})$, the minimum distance estimator will reach its efficiency bound[14]. Finally, under this choice of $\boldsymbol{S}$ the minimized value of the squared norm follows a χ^2 under the mixed hypothesis[15]. This gives us a straightforward specification test.

A major attraction of the minimum distance estimator is that it does not require solving for the stable matching patterns for even one value of the parameters: it only relies on the *observed* matching patterns. The objective function to be minimized may still be complicated, however. Things simplify dramatically if the parameters $\boldsymbol{\beta}$ enter the joint surplus $\boldsymbol{\Phi}$ linearly and the derivatives of the generalized entropy are linear in $\boldsymbol{\alpha}$. Then the mixed hypothesis is linear in $\boldsymbol{\lambda}$ and the objective function is quadratic. In fact, the minimum distance estimator can be implemented using only two OLS regressions! Galichon and Salanié (2023a) show that a large subclass of distributions $\mathbf{P}_x$ and $\mathbf{Q}_y$ can be parameterized in such a way that the derivatives of the generalized entropy are linear. The linearity of the joint surplus is much more restrictive; on the other hand, it can be seen as a flexible expansion on basis functions as in Galichon and Salanié (2022).

In many applications, the data has zeroes in some low-probability matching cells (x, y) due to sampling variations. Since the derivative of the generalizesd entropy at zero is infinite in common specifications, as our discussion of the logit model illustrates, the data must be adjusted before the minimum distance estimator can be used. A simple fix is to add a small number $\delta > 0$ to the number of observations in each cell, and to adjust the denominator accordingly so that proportions still sum to one. Alternative imputation procedures could also be used.

[14] Given an initial consistent estimate of $\boldsymbol{\lambda}_0$ and the estimator $\hat{\boldsymbol{\Sigma}}$ of the variance-covariance matrix, such a consistent estimate $\hat{\boldsymbol{\Omega}}$ can be obtained by applying the delta method.

[15] The number of degrees of freedom of the test is $|X| \times |Y|$ minus the dimension of $\boldsymbol{\lambda}_0$.

2.2.3 Estimating the Logit Model

The Choo and Siow (2006) model of Section 1.7 has parameter-free distributions $\mathbf{P}_x$ and $\mathbf{Q}_y$. The minimum distance estimator relies on the analog of (14):

$$\Phi^{\boldsymbol{\beta}}_{xy} = \log \frac{\hat{\mu}^2_{xy}}{\hat{\mu}_{x0}\hat{\mu}_{0y}}.$$

If in addition $\Phi^{\boldsymbol{\beta}}_{xy} \equiv \boldsymbol{\phi}_{xy} \cdot \boldsymbol{\beta}$ for some set of basis functions $\boldsymbol{\beta}$, then the minimum distance estimator can be implemented as a generalized least squares estimator. Galichon and Salanié (2023a) describe another possibility: they show that this model can be estimated via a Poisson regression with two-way fixed effects. Their procedure also gives estimates of the expected utilities of each group of individuals at the stable matching. An appealing feature of the Poisson estimator is that it maximizes a function that is linear in the matching patterns, so that zero cells do not create any difficulty.

3 Extensions of Separable Models

The separable model is not confined to the bipartite one to-one matching model of Section 1.1. A variety of extensions have been considered.

3.1 Matches Only

Sometimes the analyst can only observe realized matches, and not who stays single. Mergers between firms are an example, as it may be hard to define the universe of potential partners. This only requires small changes to the analysis: the maximum for G_x must be taken over $y \in Y$ only, and the maximum in G^*_x must be taken under $\sum_{y\in Y} \mu^M_{y|x} = 1$. Take the logit model as an example. Now $G_x(\boldsymbol{U}_{x\cdot}) = \log \sum_{y\in Y} \exp(U_{xy})$ and for $\sum_{y\in Y} \mu^M_{y|x} = 1$, we have

$$G^*_x(\boldsymbol{\mu}^M_{\cdot|x}) = \sum_{y\in Y} \mu^M_{y|x} \log \mu^M_{y|x}.$$

The difference is that this is attained for all $\boldsymbol{U}_{x\cdot} = \log \boldsymbol{\mu}^M_{\cdot|x} + a_x$, for an arbitrary choice of scalar a_x. As a consequence, Equation (12) becomes

$$\Phi_{xy} = \log \frac{\mu_{xy}^2}{n_x m_y} + a_x + b_y \tag{16}$$

and we only identify $\boldsymbol{\Phi}$ up to this additive sum.

This partial identification result holds for all separable models: it is a simple consequence of the fact that without singles, the utility scale for each group is only determined up to location[16]. As a result, only the components of the $\boldsymbol{\Phi}$ matrix that interact x and y can be identified in the absence of data on singles. Inference can proceed as in Section 2.2 once the locations of the utility scales are chosen.

3.2 Dynamic Matching

Choo (2015) extended Choo and Siow (2006) model to a multiperiod setting. He assumed that divorce was exogenous divorce, with partners committed to the value of transfers as long as they stay matched. He showed that in this specific model, the Bellman equations for the value of marriage imply a very tractable "dynamic marriage function" that can be taken to the data at very little computational cost.

3.3 Unipartite Matching

In the traditional marriage market, the two partners in a couple have different genders. Workteams may also consist of employees with different skills and specializations. One may of course want to model matches between partners who can be chosen without restrictions in aa population. Gender, or specialization, then becomes an observable characteristic which affects the joint utility in any match. This is the *unipartite* model, also known as the *roommate* model.

In the separable unipartite model with a finite number of groups, a given individual i has a full type and an observable group $x \in X$; we denote n_x the number of individuals in group x. If this individual matches with a partner j in group x', their joint utility is

$$\Phi_{xx'} + \varepsilon_{ix'} + \varepsilon_{jx},$$

[16] In the model with singles, it was quite natural to normalize $\Phi_{x0} = \Phi_{0y} = 0$.

where $\mathbf{\Phi}$ is a *symmetric* matrix. This formula is symmetric with respect to the transformation $(i, x) \leftrightarrow (j, x')$: in mathematical terms, we are forming a pair rather than a couple. This symmetry requirement turns out to be demanding; it may destroy the possibility of a stable matching. However, this is not a concern in large markets, which are the main focus of our analysis[17].

In the unipartite model, the decomposition in Theorem 1 becomes

$$\Phi_{xx'} = U_{xx'} + U_{x'x},$$

where $U_{xx'}$ denotes the systematic value for an individual of group x in a match with a partner of group x'. Note that the matrix $\boldsymbol{U}$ is not symmetric in general: if group x is more scarce and desirable on average than group x', it will tend to get a larger share of the joint utility in a match.

In the unipartite model, we define $\mu_{xx'}$ as the number of matches with one partner of type x and one of type x'. This implies in particular that $n_x = 2\mu_{xx} + \sum_{x' \neq x \in X} \mu_{xx'} + \mu_{x0}$, since there are two partners of type x in each (x, x) match. Given these definitions, we have just as in the bipartite model

$$U_{xx'} = \frac{\partial G_x^*}{\partial \mu_{x'|x}}(\boldsymbol{\mu}_{\cdot|x}).$$

The generalized entropy now takes the following form:

$$\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{n}) = -\sum_{x=1}^{X} n_x G_x^*(\mu_{x\cdot}/n_x)$$

and (12) becomes

$$\Phi_{xx'} = -\frac{\partial \mathcal{E}}{\partial \mu_{xx'}}(\boldsymbol{\mu}, \boldsymbol{n}).$$

which can be used as the basis for a minimum-distance estimator.

3.4 Beyond One-to-One Matching

Studies of matching on the labor markets require a many-to-one matching model in which a firm can employ a variable number of workers as in Kelso and Crawford (1982).

[17]See Chiappori, Galichon, and Salanié (2019).

For simplicity, assume that all workers occupy interchangeable functions within the firm. Consider a firm i of type x that employs a team of workers S; and let the number of workers in S who have group y be n^S_y. The vector $\boldsymbol{n}^S$ is the *group profile* of the team S. A separable specification of the joint utility of the firm owner and the workers may be

$$\tilde{\Phi}(i, S) = \Phi(x, \boldsymbol{n}^S) + \varepsilon_i(\boldsymbol{n}^S) + \sum_{j \in S} \eta_j(x).$$

This allows for unobserved shocks to the production of this team in this firm, as well as to the disutility of labor of each worker; but it rules out their interactions.

One can show (see Galichon and Salanié (2023a) or Corblet (2022)) that there exist functions $U(x, \boldsymbol{n})$ and $V(x, y)$ such that

$$U(x, \boldsymbol{n}) + \sum_{y=1}^{Y} n_y V(x, y) \geqslant \Phi(x, \boldsymbol{n})$$

with equality if some firm of type x employs a team $\boldsymbol{n}$ of workers. From this, Legendre-Fenchel transforms of the expected utilities can be defined. Note that G^*_x is defined over the set of probabilities that a firm in a given group employs a team with a given profile, $\mu(\boldsymbol{n}|x)$, while H^*_y is defined over probabilities that workers in this group work in the different groups of firms, $\mu(y|x)$. These probabilities are linked, however. Since a firm with a team of type profile $\boldsymbol{n}$ employs n_y workers, the number of workers of type y employed by all firms of type x must be

$$\mu(x, y) = \sum_{\boldsymbol{n} \in \mathbb{N}^Y} n_y \mu(x, \boldsymbol{n}).$$

This allows us to redefine H^*_y as a function $\bar{H}_y$ of the matching patterns $\mu(x, \boldsymbol{n})$ and to define the generalized entropy $\mathcal{E}$. It is easy to check that just as in the one-to-one model,

$$\Phi(x, \boldsymbol{n}) = -\frac{\partial \mathcal{E}}{\partial \mu(x, \boldsymbol{n})}. \tag{17}$$

Our minimum distance estimator can be directly applied to (17). The only difficulty is that with so many possible values of the type profile $\boldsymbol{n}$, many observed matching cells will be zero.

3.5 Continuous Observables

The most important extension may be that to continuous observed characteristics (which we called "groups" earlier). The data may have relevant covariates like income, sales, etc that are best modeled as continuous. First consider the case when all observable characteristics x and y are continuous (possibly multidimensional). Simply taking the limit of discrete models will not do: expected utilities will diverge to infinity. Building on earlier work by Dagsvik (1994), Dupuy and Galichon (2014) extended the Choo and Siow "logit" model to allow for continuous characteristics. To do so, they suppose that each individual can only match with a denumerable set of potential partners; and that the probability that they meet any of these partners is governed by a Poisson process. More precisely, consider a man with characteristics x. Under separability, there exists a function U such that this man will obtain a utility level $U_{xy} + \varepsilon_{my}$ if he matches with a woman with characteristics y. Now assume that the set of potential partners of this man generates draws (y_k, ε_k) for $k \in \mathbb{N}$, and that these draws are generated by a Poisson process that is uniform over y and has exponential density over ε. Given a similar assumption on the women's side, it can be shown that the matching patterns are the continuous analog of the discrete Choo and Siow (2006) model: the density of the matches between men of characteristics x and women of characteristics y is

$$\mu^2(x, y) = \mu(x, 0)\mu(0, y) \exp(\Phi(x, y)),$$

where we changed the notation from subscripts to parenthesized arguments to emphasize that they are continuous.

Dupuy and Galichon (2014) show that the model can be solved with IPFP and that the function $\mathbf{\Phi}$ can be estimated essentially in the same way as the discrete model. They emphasize the quadratic case in which $\Phi(x, y) = x'Ay$ for some unknown "affinity matrix" A. Then the affinity matrix is overidentified by the equality

$$A = \frac{\partial^2 \log \mu}{\partial x \partial y}(x, y) \text{ for all } x, y.$$

Moreover, one can test the rank of A to check whether the model may have a more sparse representation.

Bojilov and Galichon (2016) specialize the model further by assuming that the vector of continuous characteristics x is jointly Gaussian, as is the vector y. They also rule out singles. Then the stable matching takes a very simple form: men of characteristics x match with women of characteristics drawn from $N(Tx, V)$, where T and V depend on A and on the variances of x and y. This is one example when regressing the characteristics of one partner on that of the other recovers useful information on the joint surplus. More generally, Chiappori, Oreffice, and Quintana-Domeque (2012) studied "index-based" models, in which scalar combinations $I(x)$ and $J(y)$ of the continuous characteristics x and y have all payoff-relevant information. This requires that $\Phi(x, y)$ be a function of $I(x)$ and $J(y)$ only, and that the unobserved heterogeneity terms be independent of x and y conditionally on $I(x)$ and $J(y)$. They show that at any stable matching, the conditional distribution of $I(x)$ conditional on y only depends only on the value of $J(y)$, and conversely. Unfortunately, this is only moderately useful: the index model is a non-generic case in which the stability requirement only determines the matching between values of $I(x)$ and $J(y)$ and *not* between x and y. Therefore one cannot learn the form of the index $J(y)$ by regressing the characteristics x of men on the characteristics y of their partners[18]. Guadalupe, Rappoport, Salanié, and Thomas (2021, Appendix A.1) show that as a special but useful case, it is possible to test for this index property and to recover the indices I and J in the Dupuy and Galichon (2014) model.

Our datasets often contain both discrete and continuous variables. Suppose for instance that men have discrete characteristics a and continuous characteristics x, while women have discrete characteristics b and continuous characteristics y. Under the natural extension of the Poisson process described above, the formula for the density of matches becomes

$$\mu_{ab}(x, y) = \sqrt{\mu_a(x, 0)\mu_b(0, y)} \exp(\Phi_{ab}(x, y)/2).$$

Guadalupe, Rappoport, Salanié, and Thomas (2021) show how inference can proceed in this more general case.

[18] See Chiappori, Oreffice, and Quintana-Domeque (2020) for more information.

4 Applications of Separable Models

The marriage market has proved to be a very fruitful area for applications of separable models. I will only mention a few papers here[19].

Choo and Siow (2006) proposed the logit model and applied it to study the change of marriage patterns in the US after the Roe vs Wade 1973 ruling on abortion. Before 1973, some states already had abortion rules that were at least as permissive as the Supreme Court ruling; Choo and Siow (2006) call them the *non-reform states* as the rulling had no direct effect there. In the other, *reform states*, Roe vs Wade relaxed the regulations on abortion. Formula (14) suggests a simple differences-of-differences estimator. The relative changes in the log-marriage patterns across 1973 directly identify the relative changes in the joint surplus:

$$(\Delta\Phi_{xy})_R - (\Delta\Phi_{xy})_{NR} = \left(\Delta \log \frac{\mu_{xy}^2}{\mu_{x0}\mu_{0y}}\right)_R - \left(\Delta \log \frac{\mu_{xy}^2}{\mu_{x0}\mu_{0y}}\right)_{NR}.$$

Now consider what Choo and Siow (2006) call the "gains from marriage"—the expected utilities of the various groups on the marriage market. Since they coincide with minus the logarithm of the probability of staying single, they are easily identified, with for instance

$$(\Delta u_x)_R - (\Delta u_x)_{NR} = (\Delta \log \mu_{x0}/n_x)_{NR} - (\Delta \log \mu_{x0}/n_x)_R\,.$$

Choo and Siow (2006) found that Roe vs Wade reduced the gains from marriage for young adults, and that it may explain up to 20% of the fall in marriage rates in the 1970s. The simplest explanation is that the availability of abortion allowed some young adults to delay marriage or to forgo it altogether.

In the first application of separable matching models with continuous observed types, Dupuy and Galichon (2014) used Dutch data that has information not only on the usual sociodemographics, but also on anthropometric measurements and on the "Big Five" personality traits. Their estimates of the affinity matrix A stress the importance of diagonal terms (education, then BMI), and also of some off-diagonal interactions (e.g., a positive one between the emotional stability of husbands and the conscientiousness of women). Their tests indicate that the affinity matrix is full-rank: marital sorting cannot be explained by a small number of indices.

[19] I refer the reader to Chiappori (2020) and Chiappori and Salanié (2023) for more information.

Several recent studies have incorporated same-sex marriage in their analysis of the marriage market. One could model the marriage market as unipartite, with gender becoming just one of the observable characteristics that contribute to the joint surplus. Ciscato, Galichon, and Goussé (2020) model the same-sex marriage market and the other-sex marriage market as separate instead: the former as unipartite, and the latter as bipartite. They find that while sorting by ethnicity is not as strong for same-sex couples, sorting by education is stronger, and age matters less.

The marriage market, like others, faces formal and informal trade barriers. Most marriages are between citizens of the same country, for instance. Ahn (2023) studies cross-border marriages across Asian countries—rich Taiwan importing brides from poorer Vietnam. She shows that when obstacles to these marriages change, the resulting shifts in the equilibrium "prices" (the division of the surplus) affects the marriage outcomes and intra-household allocations of all individuals, whether they participate in cross-border marriages or not.

Chiappori, Salanié, and Weiss (2017) studied trends in the marriage patterns of US cohorts born between 1943 and 1972. They tested several variants of the Choo and Siow (2006) model against the data, allowing for arbitrary changes in the value of marruage for each education group over time. The basic model is masssively rejected; the data can only be rationalized by an increase in the preferences for assortative matching on education, as measured by the double differences in the esttimated joint surplus. This results in a increase in the "marital college premium" for white women: as home production and in particular investing in children's human capital has become more important for educated parents, assortative matching by education has increased, to the benefit of women.

Chiappori, Costa-Dias, and Meghir (2018) analyze the British marriage market using a structural model based on a three-stage game: individuals independently invest in their human capital; then they match on the marriage market; finally, the resulting households consume and supply labor. They use their estimates to evaluate the impact of cheaper access to higher education.

Koh (2023) uses the Choo and Siow (2006) model to evaluate the effects of the change in the racial mix of the US since 1980 on the gains from marriage of various groups. Her estimates suggest that college-educated Black men benefited from increased intermarriage with college-educated White women. On the other hand, Black women did not gain as much.

While it is hard to get even indirect proxies of transfers between partners on the marriage market, the situation is usually more favorable on the labor market. Data on *individual* transfers can prove quite helpful to identify the parameters[20].

Applications of matching to the labor market often use the search framework (see Section 6). Still, there has been much work recently on separable models. Lindenlaub (2017) models matching between workers and jobs who differ in two dimensions: manual and cognitive. She shows that the US labor market has moved towards stronger worker-job complementarities in cognitive skills, which explains part of the increased wage polarization. Her paper also defines a new index of multidimensional sorting. Corblet (2022) uses a many-to-one variant of the Choo and Siow model to examine how workers in different age and education groups sort across firms in different industries on the Portuguese labor market, where the supply of high-school and university graduates was low until the 1990s and has increased markedly since then. Her estimates suggest that increases in the relative productivity of graduates have not been enough to compensate for the large increase in their numbers.

Traditional labor markets only allow firms to hire a worker's bundle of skills: those that are valuable to the firm and those that are less valuable to it. Choné and Kramarz (2022) argue that technological change, remote work, and the gig economy will make it easier for worker to "unbundle" their skills. They show that the resulting shifts in labor market sorting benefit "generalist" workers and lead firms to specialize more.

Calvo, Lindenlaub, and Reynoso (2023) link sorting in the marriage market and on the labor market. When partners form a couple, they choose an intra-household allocation of consumptions and labor supplies; then those who work match with firms in the labor market. Marriage choices thus depend on expectations of later firm-worker sorting. Using German data, Calvo, Lindenlaub, and Reynoso (2023) find that the observed drop in the gender pay gap is not due to technical change in the labor market, but rather to changes in both the complementarity and the allocation of the time allocated to home production hours.

[20] On the other hand, data on average transfers between groups—say the average salary of a foreman in the car industry—are much less useful; see Salanié (2015).

Finally, it is worth noting that the literature on trade that follows Eaton and Kortum (2002) partly relies on a stochastic specification that is similar to a multiplicative version of the Choo and Siow (2006) model[21].

4.1 A Caveat

One point is not always appreciated in the literature: getting reasonably precise estimates may require large datasets when the matching cells have very different sizes. To take a simple illustration, consider a Choo and Siow (2006) model with no singles (that is, we only observe matches) and only two groups: $X = Y = 2$. Then we know from Section 3.1 that we can only identify one parameter, the double log-difference

$$\log \frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}.$$

Call it p_0. The natural estimator on a sample of n matches is the empirical analog

$$\hat{p}_n = \log \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

where n_{xy} is the observed number of (x, y) matches. It is easy to see that its asymptotic variance is

$$\operatorname{Var} \hat{p}_n = \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}.$$

Since $x \to 1/x$ is convex and $n_{11} + n_{12} + n_{21} + n_{22} = n$, the minimum variance is achieved when all $n_{xy} = n/4$; then $\operatorname{Var} \hat{p}_n = 4/n$. Any small cell size deteriorates the performance of the estimator, as

$$\operatorname{Var} \hat{p}_n \geqslant \frac{1}{\min_{x,y=1,2} n_{xy}}.$$

The simulations reported in Galichon and Salanié (2023a) show that it can be a serious problem in applications. They fit an 8-parameter logit model to the data used by Choo and Siow (2006); then they use the estimated parameter values as the data-generating process for a Monte Carlo study of the performance of the minimum-distance and Poisson estimators. They found that both estimators perform similarly; they exhibit sizable biases for samples as large as 300,000 observations.

[21] See Costinot and Vogel (2015) for a survey.

The Choo and Siow (2006) dataset has unusually large variation in the size of the matching cells: the P90-P10 ratio is larger than 100. The estimator is likely to perform better in other applications. As an example, Galichon and Salanié (2023a) obtain very good results in another simulation when they "shrink" parameter values so that the sizes of the matching cells are less dispersed. Still, this is an important caveat to keep in mind; it may be prudent for instance to combine groups whose sizes are small.

5 Non-separable Models

Separable models restrict the distributions of the unobserved heterogeneity terms in order to identify the joint surplus. In contrast, most of semiparametric econometrics has emphasized relaxing assumptions on the distribution of error terms. In a series of papers (Fox, 2010; Fox, Yang, and Hsu, 2018; Fox, 2018), Jeremy Fox has proposed methods that follow this alternative path.

The "rank-order property" is a central concept in these papers. To understand it, first consider the usual single-agent discrete choice models. Such models have a 0-1 variable y determined by whether some scalar index $x\beta_0$ is large enough:

$$y = 1 \text{ iff } x\beta_0 > \varepsilon.$$

If the distribution of ε is independent of x, then clearly $\Pr(y = 1|x)$ is an increasing function of $x\beta_0$. This idea goes back to Manski (1975). It can be used to estimate β_0 up to an increasing transformation: first obtain a flexible estimate $\hat{p}(x)$ of $\Pr(y = 1|x)$, then trace the level curves of $x \to \hat{p}(x)$. They coincide with the level curves of the unknown function $x\beta_0$. Moreover, $x\beta_0 > x'\beta_0$ if and only if $\hat{p}(x) > \hat{p}(x')$. This can be written as the set of inequalities

$$(x\beta_0 - x'\beta_0)(\hat{p}(x) - \hat{p}(x')) \geqslant 0 \text{ for all } x, x'.$$

This yields several consistent estimators. One can for instance take a set of pairs of observations (i, j) and find the values of β that minimize the number of violations of the set of inequalities

$$(x_i\beta_0 - x_j\beta_0)(y_i - y_j) \geqslant 0.$$

Manski's maximum score estimator is another variant, under slightly different hypotheses.

The form of the Choo and Siow formula (14) suggests a natural extension of this monotonicity property: we could posit that

$$\begin{aligned} &\Phi(x,y) + \Phi(x',y') - \Phi(x,y') - \Phi(x',y) > 0 \\ \text{iff} \quad &\log \mu(x,y) + \log \mu(x',y') - \log \mu(x,y') - \log \mu(x',y) > 0. \end{aligned} \tag{18}$$

Just as in the one-sided model, this would give a consistent estimator of the function Φ, up to additive terms $a(x) + b(y)$ and to a scalar multiple. This *rank-order property* clearly holds in the Choo and Siow (2006) model, since the double difference on Φ is exactly twice the double difference on $\log \mu$. Graham (2013) showed that it holds in all separable models where the distributions of heterogeneity terms are iid for each gender; that is, $\mathbf{P}_x$ does not depend on x and $\mathbf{Q}_y$ does not depend on y. His result was extended by Fox (2018) to distributions that are only exchangeable. These two results give a rigorous justification for the use of a maximum score approach under these assumptions.

Another way of using monotonicity is to compare matchings across markets. Suppose that the joint utility from a match is

$$\tilde{\Phi}_{mw} = \Phi_{xy} + \xi_{mw}.$$

If we could observe many markets in which both the joint surplus $\boldsymbol{\Phi}$ and the numbers of men and women in each group $\boldsymbol{q}$ are the same, then differences across matchings across markets would be entirely due to different draws of the $\boldsymbol{\xi}$ terms. If all matchings are stable in their respective markets, they must maximize the sum of joint utilities on each. More precisely, suppose that a matching $\mathcal{M}$ (a list of couples $h = (m,w)$) is stable when the unobservables $\boldsymbol{\xi}$, and another matching $\mathcal{M}'$ is stable when the unobservables take different values $\boldsymbol{\xi}'$. Then we must have

$$\sum_{h=(m,w)\in\mathcal{M}} (\Phi(x_m, y_w) + \xi_{mw}) > \sum_{h=(m,w)\in\mathcal{M}'} (\Phi(x_m, y_w) + \xi_{mw}).$$

Fox (2010) suggested dropping the unobservables from this equation. This results in a set of inequalities that can be used in a maximum-score estimator. It can only be an approximation in general. Somewhat miraculously, it is exact in the Choo and Siow

(2006) model: given (16), maximizing the sum of the joint surpluses is equivalent to maximizing the loglikelihood.

Finally, Fox, Yang, and Hsu (2018) take the opposite approach to Choo and Siow: they restrict the specification of the joint surplus $\mathbf{\Phi}$ in order to identify the distribution of the unobservables. The simplest (and most restrictive) result in Fox, Yang, and Hsu (2018) assumes that the joint surplus Φ_{xy} is observed for every possible match, that it varies continuously, and that the $\boldsymbol{\xi}$ terms are distributed independently of it. They show that if the analyst observes many markets with different joint surplus matrices $\mathbf{\Phi}$ but the same distribution for $\boldsymbol{\xi}$, then the distribution of the complementarities across these unobservables is identified.

One advantage of the inequalities-based approaches is that when the underlying assumptions are satisfied, they hold with little modification from one-to-one to many-to-many matching problems. In fact, they have found their main applications in the industrial organization, marketing strategy fields, where firms interact within often complex networks of relationships. In addition, the analyst can choose to base estimation on a subset of what they think are the most relevant or informative inequalities.

In an early example, Fox and Bajari (2013) applied the maximum-score estimation to the FCC spectrum auctions. Since bidders value complementarities across licenses, this can be seen as a many-to-one matching game. In this context, the inequalities are meant to reflect pairwise stability: in equilibrium, a bidder will not want to swap one of their licenses for one they don't own, given prevailing prices. Fox and Bajari (2013) found that while auctioning licenses was much more efficient than the previous lottery system, auctioning larger blocks of licenses would be even better.

Fox (2018) uses data on the sales of car parts to carmakers in the US. This is a many-to-many example, as a given seller may trade with several buyers, and vice versa. Fox (2018) uses his estimates to evaluate a counterfactual in which General Motors was made to divest Opel, and to test whether the entry of Asian carmakers led suppliers to improve the quality of their products.

6 Matching with Frictions

Introducing frictions in the matching process is another way of reconciling the data and the theory. Just as in the labor market, this can be done in a search model. Meetings between potential parners occur randomly; when a meeting takes place, each partner must choose whether to propose a match and a division of the joint utility, or continuing to search. The cost of an additional search is usually taken to be the utility that is lost while waiting for another meeting. In the simplest version of the model, matches are destroyed randomly; richer models incorporate utility shocks that prompt partners to dissolve their match.

6.1 Identification

It can be hard to distinguish the predictions of matching models with unobserved heterogeneity and those of matching models with frictions. To be more precise, take the pioneering Shimer and Smith (2000) model. They describe a one-to-one, unipartite matching market in which types $x \in [0,1]$ meet randomly. If a type x meets a type y, each individual then decides whether to wait for a better deal or to agree to form a match with an agreed division of the surplus. The match lasts until it is randomly dissolved. Shimer and Smith (2000) give conditions under which there is a unique steady state equilibrium. Denote $\mu(x,0)$ the numbers of singles and $\mu(x,y)$ the matching patterns in this steady state. We know from Section 1.5 that we could fit to this data a variety of separable models (for instance the Choo and Siow (2006) model) and rationalize the data without any reference to frictions.

On the other hand, even the logit model can generate matching patterns that cannot be rationalized by some classes of models with frictions, even though they have more parameters. Take a type x in the Shimer and Smith (2000) model. When unmatched, she will meet all other types with a probability proportional to their frequency in the unmatched population. This applies in particular to all types that belong to x's "matching set" (that is, types that are acceptable to her.) Take two such types y and y'; then it can be shown that

$$\frac{\mu(x,y)}{\mu(x,y')} = \frac{\mu(0,y)}{\mu(0,y')}.$$

This equality imposes testable restrictions on the observed matching patterns that are only compatible with some separable models. They are specific to the assumptions in Shimer and Smith; models of directed search, for instance, would generate different restrictions. The key point here is that models with frictions really add value in settings in which matches are observed over several periods and/or transfers can be observed. On labor markets for instance, wage transitions and the dynamics of employment offer rich information to identify all components of the model[22].

In such multiperiod models, three new elements must be specified: how meetings occur; the degree of commitment to the division of the surplus; and the process by which matches dissolve (divorces in marriages, quits and layoffs in employment relationships). Take marriages as an illustration. Potential partners used to be found in a very narrow geographic area, and they are still driven by social and work groups. Divorces are legal in most countries, which by definition limits the commitment abilities of partners. They are governed by a wide variety of legal frameworks that determine exit options. Moreover, divorces clearly do not happen randomly. Random shocks may affect the joint utility of existing matches. Better still, shocks to wages for instance change exit options and therefore the bargaining power of partners. This begs the question of the ability of partners to commit to the division of the surplus. Full commitment makes little sense when matches are allowed to dissolve. In the no-commitment version, partners may renegotiate at any point in time. Several recent papers have adopted the limited commitment hypothesis in Mazzocco (2007). In this framework, partners cannot commit not to dissolve their match; the division of surplus is renegotiated only when the participation constraint of one partner becomes binding. The only way to keep that partner in the match may be to violate the participation constraint of another partner; then renegotiation fails and the match dissolves.

6.2 Applications

Greenwood, Guner, Kocharkov, and Santos (2016) set up a search-and-matching model in which individuals choose their education, marry or not, work and invest in home production, and may divorce or remarry. They fit their model on US data

[22]See Hagedorn, Law, and Manovskii (2017).

from 1960 and simulate it for 2005, using the new prices and wages. They find that cheaper durable goods have reduced labor inputs into home production and allowed women to work in the market; the lower gender wage gap played a smaller role. As individuals became more independent, they searched more; marriage declined and divorce increased.

Ciscato (2021) also studied 50 years of US data with a model of education, marriage, work, and divorce. He puts more emphasis in complementarities on home production, and on renegotiation of the division of surplus when wage shocks hit individuals. His analysis suggests that while changes in relative wages explain part of the rise in education and the decline in marriage, complementarities in home production are needed to explain the relative increase in the number of college-educated women and the stability on their marriage rate.

Shephard (2019) considers an overlapping generations model that allows different cohorts to inter-marry. He uses this framework to analyze the evolution of age at first marriage or of the marital agegap. Among other findings, the significant increase in women's relative earnings since the 1980s has reduced the marital age gap.

Lindenlaub and Postel-Vinay (2023) model the matching between workers and jobs with multiple characteristics on the labor market. They find that sorting is positive between those worker-job characteristics that have the stronger complementarities, and negative in other dimensions. Lindenlaub and Postel-Vinay (2021) relies on the job ladder idea to identify the observable characteristics of workers and jobs that generate joint surplus. The intuition is that since matches maximize joint surplus, workers tend to move to jobs that offer a higher joint surplus. This revealed preference argument allows them to use observed transitions to derive a sparse proxy for the joint surplus from a match.

Concluding Remarks

Most of the literature has modelled matches as affecting only the utilities of the partners. Social norms like the local marriage rate could influence the preferences of individuals; peer effects can also play a role. Mourifié (2019) and Mourifié and Siow (2021) introduced marriage matching functions that go in this direction, without a specific structural underpinning. More research is needed, especially for applications to industrial organization—firms's mergers for instance clearly affect other firms.

Finally, there are tight analogies between matching with perfect transferable utility and hedonic models, as pointed out by Chiappori, McCann, and Nesheim (2010). More generally, they can both be reinterpreted as network flow problems[23]; this may open the way to new insights and methods.

References

Ahn, S. Y. (2023): "Matching Across Markets: An Economic Analysis of Cross-Border Marriage," forthcoming at the *Journal of Labor Economics*.

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

——— (1974): "A theory of marriage, part II," *Journal of Political Economy*, 82, S11–S26.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Berry, S., and A. Pakes (2007): "The Pure Characteristics Demand Model," *International Economic Review*, 48, 1193–1225.

Bojilov, R., and A. Galichon (2016): "Matching in closed-form: equilibrium, identification, and comparative statics," *Economic Theory*, 61, 587–609.

Calvo, P., I. Lindenlaub, and A. Reynoso (2023): "Marriage Market and Labor Market Sorting," forthcoming in the *Review of Economic Studies*.

Chiappori, P. (2020): "The Theory and Empirics of the Marriage Market," *Annual Review of Economics*, 12(1), 547–578.

Chiappori, P., M. Costa-Dias, and C. Meghir (2018): "The Marriage Market, Labor Supply, and Education Choice," *Journal of Political Economy*, 126, S26–S72.

[23]See Galichon and Salanié (2023b).

——— (2020): "Changes In Assortative Matching: Theory And Evidence For The US," Discussion paper, Cowles Foundation Discussion Paper No. 2226.

Chiappori, P., A. Galichon, and B. Salanié (2019): "On Human Capital and Team Stability," *Journal of Human Capital*, 13, 236–259.

Chiappori, P., 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., D. L. Nguyen, and B. Salanié (2019): "Matching with Random Components: Simulations," Columbia University mimeo.

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

——— (2020): "Erratum: Fatter Attraction: Anthropometric and Socioeconomic Characteristics in the Marriage Market," *Journal of Political Economy*, 128, 4673–5.

Chiappori, P., and B. Salanié (2023): "Mating Markets," in *Handbook of the Economics of the Family, vol. 1.* North-Holland.

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

Choné, P., and F. Kramarz (2022): "Matching Workers' Skills and Firms' Technologies: From Bundling to Unbundling," mimeo Crest.

Choo, E. (2015): "Dynamic Marriage Matching: An Empirical Framework," *Econometrica*, 83, 1373–1423.

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

Ciscato, E. (2021): "The changing wage distribution and the decline of marriage," KU Leuven.

Ciscato, E., A. Galichon, and M. Goussé (2020): "Like Attract Like: A Structural Comparison of Homogamy Across Same-Sex and Different-Sex Households," *Journal of Political Economy*, 128, 740–781.

Corblet, P. (2022): "Education Expansion, Sorting, and the Decreasing Education Wage Premium," Discussion paper, Sciences Po Paris, mimeo.

Costinot, A., and J. Vogel (2015): "Beyond Ricardo: Assignment Models in International Trade," *Annual Review of Economics*, 7, 31–62.

Dagsvik, J. (1994): "Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes," *Econometrica*, 62, 1179–1205.

de Palma, A., and K. Kilani (2007): "Invariance of Conditional Maximum Utility," *Journal of Economic Theory*, 132, 137–146.

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

Eaton, J., and S. Kortum (2002): "Technology, Geography, And Trade," *Econometrica*, 70, 1741–1779.

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

——— (2018): "Estimating Matching Games with Transfers," *Quantitative Economics*, 8, 1–38.

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

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

Galichon, A., and B. Salanié (2017): "The Econometrics and Some Properties of Separable Matching Models," *American Economic Review Papers and Proceedings*, 107, 251–255.

Galichon, A., and B. Salanié (2022): "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," *The Review of Economic Studies*, 89, 2600–2629.

——— (2023a): "Estimating Separable Matching Models," *Journal of Applied Econometrics*, forthcoming.

——— (2023b): "Hedonic and Matching Models," mimeo Columbia.

Graham, B. (2013): "Uniqueness, Comparative Static, And Computational Methods for an Empirical One-to-one Transferable Utility Matching Model," *Structural Econometric Models*, 31, 153–181.

Greenwood, J., N. Guner, G. Kocharkov, and C. Santos (2016): "Technology and the Changing Family: A Unified Model of Marriage, Divorce, Educational Attainment and Married Female Labor-Force Participation," *American Economic Journal: Macroeconomics*, 8, 1–41.

Guadalupe, M., V. Rappoport, B. Salanié, and C. Thomas (2021): "The Perfect Match: Assortative Matching in Mergers and Acquisitions," Columbia mimeo.

Gualdani, C., and S. Sinha (2023): "Partial identification in matching models for the marriage market," *Journal of Political Economy*, 131, 1109–1171.

Hagedorn, M., T. Law, and I. Manovskii (2017): "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*, 85, 29–65.

Kelso, A. S., and V. P. Crawford (1982): "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, pp. 1483–1504.

Koh, Y. K. (2023): "Evolution of Gains from Racial Desegregation in the US Marriage Market," Columbia mimeo.

Lewbel, A. (2000): "Semiparametric Qualitative Response Model Estimation with Unknown Heteroskedasticity and Instrumental Variables," *Journal of Econometrics*, 97, 145–177.

Lindenlaub, I. (2017): "Sorting multidimensional types: theory and application," *Review of Economic Studies*, 84, 718—-789.

Lindenlaub, I., and F. Postel-Vinay (2021): "The Worker-Job Surplus," mimeo.

——— (2023): "Multi-Dimensional Sorting Under Random Search," forthcoming in the *Journal of Political Economy*.

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

Matzkin, R. (2007): "Heterogenous choice," in *Advances in Economics and Econometrics, Theory and Applications. Ninth World Congress of the Econometric Society*, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge University Press.

Mazzocco, M. (2007): "Household Intertemporal Behavior: A Collective Characterization and a Test of Commitment," *Review of Economic Studies*, 74, 857–895.

Mourifié, I. (2019): "A Marriage Matching Function with Flexible Spillover and Substitution Patterns," *Economic Theory*, 67, 421–461.

Mourifié, I., and A. Siow (2021): "The Cobb Douglas Marriage Matching function: Marriage Matching with Peer and Scale Effects," *Journal of Labor Economics*, 39 (S1), S239–S274.

Salanié, B. (2015): "Identification in Separable Matching with Observed Transfers," Columbia University mimeo.

——— (2023): "`cupid-matching`: a Python package to solve and estimate separable matching models," https://pypi.org/project/cupid-matching/.

Shephard, A. (2019): "Marriage market dynamics, gender, and the age gap," U Penn mimeo.

Shimer, R., and L. Smith (2000): "Assortative matching and Search," *Econometrica*, 68, 343–369.

Siow, A. (2015): "Testing Becker's Theory of Positive Assortative Matching," *Journal of Labor Economics*, 33, 409–441.

A Some Convex Analysis

Consider a convex function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ that is not identically $+\infty$. If φ is differentiable at x, we denote its *gradient* at x as the vector of partial derivatives, that is $\nabla\varphi(x) = (\partial\varphi(x)/\partial x_1, \ldots, \partial\varphi(x)/\partial x_n)$. It is the only vector $y \in \mathbb{R}^n$ such that

$$\varphi(\tilde{x}) \geqslant \varphi(x) + y^\top(\tilde{x} - x) \quad \forall \tilde{x} \in \mathbb{R}^n. \tag{19}$$

This motivates the definition of the *subdifferential* $\partial\varphi(x)$ of φ at x as the set of vectors $y \in \mathbb{R}^n$ such that relation (19) holds. Equivalently, $y \in \partial\varphi(x)$ holds if and only if

$$x \cdot y - \varphi(x) = \max_{\tilde{x}} \{\tilde{x} \cdot y - \varphi(\tilde{x})\} \tag{20}$$

(where the equality sign obtains by taking $\tilde{x} = x$ in (19)).

The right-hand side of (20) plays a special role in convex analysis: it is the *Legendre-Fenchel transform* of φ at y and is usually denoted $\varphi^*(y)$. By construction,

$$\varphi(x) + \varphi^*(y) \geqslant x \cdot y.$$

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

$$\partial\varphi(x) = \arg\max_y \{x \cdot y - \varphi^*(y)\}.$$

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

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

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

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

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

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

If φ is differentiable at x and φ^* is differentiable at y, the last two items obviously become $y = \nabla\varphi(x)$ and $x = \nabla\varphi^*(y)$.

B Finite Markets

Assumption 3 may not be a good approximation in smaller matching markets. Most tools presented in Sections 1 and 2 adapt without difficulty, however. Take the function $\tilde{G}_x$ defined in Section 1.5:

$$\tilde{G}_x(\boldsymbol{U}_{x\cdot}) = \frac{1}{n_x} \sum_{m \in x} \max_{y \in Y_0} (U_{xy} + \varepsilon_{my}).$$

As the maximum of linear functions, it is convex, and it has a subgradient everywhere. This subgradient consists of the derivative almost everywhere; the exceptions are vectors $\boldsymbol{U}_{x\cdot}$ where $U_{xy} + \varepsilon_{my} = U_{xt} + \varepsilon_{mt}$ for some $(y, t) \in Y^2$ and $m \in x$.

The Legendre-Fenchel transform of $\tilde{G}_x$ is defined by

$$\tilde{G}^*_x(\boldsymbol{\mu}_{\cdot|x}) = \max_{U_{x0}=0, U_1, \ldots, U_{|Y|}} \left(\sum_{y \in Y} \mu_{xy} U_{xy} - \tilde{G}_x(\boldsymbol{U}_{x\cdot}) \right).$$

It is convex, as always; and since $\tilde{G}_x$ was convex, the properties of the transform (see Appendix A) imply that

$$\boldsymbol{U}_{x\cdot} \in \partial \tilde{G}^*_x(\boldsymbol{\mu}_{\cdot|x}) \text{ iff } \boldsymbol{\mu}_{\cdot|x} \in \partial \tilde{G}_x(\boldsymbol{U}_{x\cdot}).$$

Proceeding similarly for women, we define the generalized entropy as in the text:

$$-\tilde{\mathcal{E}}(\boldsymbol{\mu}; \boldsymbol{q}) = \sum_{x \in X} n_x \tilde{G}^*_x(\boldsymbol{\mu}_{x\cdot}/n_x) + \sum_{y \in Y} m_y \tilde{H}^*_y(\boldsymbol{\mu}_{\cdot y}/m_y).$$

Since $\boldsymbol{\mu} \to -\tilde{\mathcal{E}}(\boldsymbol{\mu}; \boldsymbol{q})$ is convex, we denote its subgradient as

$$\partial \left(-\tilde{\mathcal{E}}\right)(\boldsymbol{\mu}; \boldsymbol{q}) \equiv \left\{ \sum_{x \in X} n_x \boldsymbol{a}_x + \sum_{y \in Y} m_y \boldsymbol{b}_y \mid \boldsymbol{a}_x \in \partial \tilde{G}^*_x(\boldsymbol{\mu}_{x\cdot}/n_x),\ \boldsymbol{b}_y \in \partial \tilde{H}^*_y(\boldsymbol{\mu}_{\cdot y}/m_y) \right\}.$$

It is a convex, closed subset of $\mathbf{R}^{|X|+|Y|}$; as such, its projection on the xy axis is a closed interval $[\underline{e}_{xy}, \bar{e}_{xy}]$. Remember from Theorem 1 that $U_{xy} + V_{xy} \geqslant \Phi_{xy}$. This gives an upper bound $\Phi_{xy} \geqslant \bar{e}_{xy}$. If $\mu_{xy} > 0$, then $U_{xy} + V_{xy} = \Phi_{xy}$, so that we also have a lower bound $\Phi_{xy} \geqslant \underline{e}_{xy}$. Moreover, if $\tilde{G}^*_x$ is differentiable at $\boldsymbol{\mu}_{\cdot|x}$ and $\tilde{H}^*_y$ is differentiable at $\boldsymbol{\mu}_{\cdot|y}$, we have $\underline{e}_{xy} = \bar{e}_{xy}$.

C Some Generalized Entropies

We give here the forms of the generalized entropy functions for a few common specifications.

C.1 The Choo and Siow Model with Heteroskedasticity

Let us start with an easy extension of the Choo and Siow (2006) logit model: the distributions $\mathbb{P}_x$ and $\mathbb{Q}_y$ are type I extreme values iid vectors with unknown scale factors σ_x and τ_y respectively. The homoskedastic model obtains when all σ_x and τ_y equal one; a gender-heteroskedastic model would have all σ_x equal to one and all τ_y equal to an unknown τ.

The generalized entropy of this model is

$$\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{q}) = -\sum_{x \in X} \sigma_x \sum_{y \in Y_0} \mu_{xy} \log \frac{\mu_{xy}}{n_x} - \sum_{y \in Y} \tau_y \sum_{x \in X_0} \mu_{xy} \log \frac{\mu_{xy}}{m_y}.$$

Note that if we collect the scale factors in $\boldsymbol{\alpha} = (\boldsymbol{\sigma}, \boldsymbol{\tau})$, the generalized entropy and its derivatives in $\boldsymbol{\mu}$ are linear in $\boldsymbol{\alpha}$.

C.1.1 Nested Logit

Consider a two-layer nested logit model. Take men of type x first. Alternative 0 (singlehood) is obviously special; we put it alone in its nest. Each other nest $n \in \mathcal{N}_x$ contains alternatives $y \in \mathcal{Y}_n$. The correlation of alternatives within nest n is proxied by $1 - (\rho^x_n)^2$ (with $\rho^x_0 = 1$ for the nest made of alternative 0). Similarly, for women of type y, alternative 0 is in a nest by itself with parameter $\delta^y_0 = 1$ and alternatives $x \in \mathcal{X}_{n'}$ are in a nest $n \in \mathcal{N}'_y$ with parameter $\delta^y_{n'}$. We collect the parameters $\boldsymbol{\rho}$ and $\boldsymbol{\delta}$ into $\boldsymbol{\alpha}$.

The formulæ in Example 2.1 of Galichon and Salanié (2022) imply that if y is in nest $n \in \mathcal{N}_x$ and x is in nest $n' \in \mathcal{N}_y$, then

$$\begin{aligned}\frac{\partial \mathcal{E}^{\boldsymbol{\alpha}}}{\partial \mu_{xy}}(\boldsymbol{\mu}, \boldsymbol{q}) = &-\rho_n^x \log \frac{\mu_{xy}}{\mu_{x0}} - (1-\rho_n^x) \log \frac{\mu_{xn}}{\mu_{x0}} \\ &- \delta_{n'}^y \log \frac{\mu_{xy}}{\mu_{0y}} - (1-\delta_{n'}^y) \log \frac{\mu_{n'y}}{\mu_{0y}},\end{aligned} \tag{21}$$

where we defined $\mu_{xn} = \sum_{t \in \mathcal{Y}_n} \mu_{xt}$ and $\mu_{n'y} = \sum_{z \in \mathcal{X}_{n'}} \mu_{zy}$. Once again, this is linear in the parameters $\boldsymbol{\alpha}$; it remains linear if we impose constraints on the nests (for instance, that $\mathcal{N}_x$ is the same for all types x) and/or linear constraints on the $\boldsymbol{\rho}$ parameters (for instance, that ρ_n^x only depends on n). Here too the second derivatives, while more complicated than in the multinomial logit, can be computed in closed form.

C.1.2 Mixed Logit

Let us now describe a random coefficient logit model. Consider a man i of type x, endowed with preferences $\boldsymbol{e}_i$ over a set of d observable characteristics $\boldsymbol{Z}$ of potential partners. We add an idiosyncratic shock $\boldsymbol{\xi}_i$ that is distributed as a standard iid type I extreme value vector over $\mathbb{R}^{Y+1}$, independently of $\boldsymbol{e}_i$, and a scale factor $s > 0$:

$$\varepsilon_{iy} = \sum_{k=1}^{d} Z_{yk} e_{ik} + s \xi_{iy}$$

or in matrix form: $\boldsymbol{\varepsilon} = \boldsymbol{Z}\boldsymbol{e} + s\boldsymbol{\xi}$. This specification is standard in empirical IO[24].

Let individual preferences $\boldsymbol{e}$ of men of type x have distribution $\mathbb{P}_x^e$. We will seek to estimate the parameters $\boldsymbol{\beta}$ of the joint surplus, the scale factor s, and the parameters of the distributions $\mathbb{P}_x^e$. We collect s and the parameters of $\mathbb{P}_x^e$ in a vector $\boldsymbol{\alpha}$.

To compute the derivative of the generalized entropy function, we only need to replace the max operator in the definition of the function G_x with a regularized "softmax" that accounts for the integration over the shocks $\boldsymbol{e}$:

$$G_x(\boldsymbol{U}; \boldsymbol{\alpha}) = \int s \log \sum_{y=0,1,\ldots,Y} \exp\left(\frac{U_y + (\boldsymbol{Z}\boldsymbol{e})_y}{s}\right) d\mathbb{P}_x^e(\boldsymbol{e}).$$

[24] In Berry, Levinsohn, and Pakes (1995), the covariates in $\boldsymbol{Z}$ stand for the observed characteristics of the products; the $\boldsymbol{e}$ are individual valuations of these characteristics; and the $\boldsymbol{\xi}$ are idiosyncratic shocks.

This gives

$$G_x^*(\boldsymbol{\nu}; \boldsymbol{\alpha}) = \max_{\boldsymbol{U} \in \mathbb{R}^{Y+1}} \left[\sum_{y \in \mathcal{Y}_0} \nu_y U_y - \int s \log \sum_{y=0,1,\dots,Y} \exp\left(\frac{U_y + (\boldsymbol{Z}\boldsymbol{e})_y}{s} \right) d\mathbb{P}_x^e(\boldsymbol{e}) \right].$$

By the envelope theorem, the derivative of $G_x^*(\boldsymbol{\nu}; \boldsymbol{\alpha})$ with respect to $\boldsymbol{\nu}$ is the vector $\boldsymbol{U}$ that solves the system

$$\nu_y = \int \frac{\exp((U_y + (\boldsymbol{Z}\boldsymbol{e})_y)/s)}{\sum_{t=0,1,\dots,Y} \exp((U_t + (\boldsymbol{Z}\boldsymbol{e})_t)/s)} d\mathbb{P}_x^e(\boldsymbol{e}) \;\; \forall y = 1, \dots, Y.$$

This is exactly isomorphic to the inversion problem in Berry, Levinsohn, and Pakes (1995), with the unknown $\boldsymbol{U}$ standing for the product effects and $\boldsymbol{\nu}$ playing the role of the product market shares[25]. After replacing $\boldsymbol{\nu}$ with the observed $\boldsymbol{\mu}_{x\cdot}/n_x$, the system can be solved by any of the algorithms that are standard in this literature. The solution gives row x of the matrix $\boldsymbol{U}$. Proceeding in the same way for other types of men, and solving for $\boldsymbol{V}$ for women, gives the derivatives of the generalized entropy function:

$$\frac{\partial \mathcal{E}^{\boldsymbol{\alpha}}}{\partial \mu_{xy}}(\boldsymbol{\mu}, \boldsymbol{q}) = -\frac{\partial G_x^*}{\partial \nu_{xy}} \left(\frac{\boldsymbol{\mu}_{x\cdot}}{n_x} \right) - \frac{\partial H_y^*}{\partial \nu_{xy}} \left(\frac{\boldsymbol{\mu}_{\cdot y}}{m_y} \right) = -U_{xy} - V_{xy} = -\Phi_{xy}.$$

Note that except in trivial cases, the derivatives of the generalized entropy are *not* linear in $\boldsymbol{\alpha}$.

[25]The limit case $s = 0$ yields the pure characteristics model of Berry and Pakes (2007).