École Doctorale Économie, Management, Mathématiques, Physique et Sciences Informatiques (EM2PSI)
Théorie Économique, Modélisation et Applications (THEMA) CNRS UMR 8184

THÈSE DE DOCTORAT

Pour l'obtention du grade de docteur de CY Cergy Paris Université

Spécialité: Sciences Économiques

Econometric Matching Models: Education, Inequality and Consumption

---

Présentée et soutenue publiquement par:

Ulysse LAWOGNI

le 29 Janvier 2021

Sous la direction de: Olivier DONNI

Jury de soutenance

Directeur de Thèse

Olivier DONNI, Professeur à CY Cergy Paris Université, France

Rapporteurs

Pierre-André CHIAPPORI, Professeur à Columbia University, New-York, USA
Arnaud DUPUY, Professeur à l'Université du Luxembourg, Luxembourg

Examinateurs

Nathalie PICARD, Présidente du Jury, Professeure à l'Université de Strasbourg, France
Pierre ANDRÉ, Maître de Conférences à CY Cergy Paris Université, France
Anne SOLAZ, Directrice de Recherches à l'Institut des Etudes Démographiques, France

Remerciements

Je tiens à remercier en premier lieu mon directeur de thèse Olivier Donni pour son encadrement, la disponibilité et la patience constantes dont il a fait preuve pour m'orienter et me diriger dans la bonne voie. Il m'a appris la persévérance dans la recherche, la précision des idées et la clarté. Ce fut un apprentissage difficile certes, mais mû par la passion pour ce sujet de thèse, ce me fut bien nécessaire pour pouvoir développer et exprimer mes idées clairement; cette école m'a rendu meilleur et je le remercie vraiment. Il m'a surtout permis d'atteindre un rêve que j'ai nourri très tôt dès le lycée et qui était celui de faire un doctorat. Je lui suis infiniment reconnaissant. Je tiens également à exprimer toute ma reconnaissance envers M. Pierre-André Chiappori pour qui j'ai une immense admiration, une grande référence pour moi en matière de recherche économique et qui m'honore si extrêmement en acceptant de lire mon manuscrit de thèse et en être rapporteur. Je le remercie vraiment pour cette considération. J'exprime tout autant ma gratitude envers M. Arnaud Dupuy qui a également accepté apprécier mon travail et être rapporteur de mon jury de thèse. J'en suis très honoré. Je lui adresse sincèrement mes remerciements et lui témoigne ma reconnaissance. Je ne pourrai assez le remercier. Il a été en effet surtout très crucial pour moi précisément grâce à son temps qu'il m'a accordé pour m'expliquer son modèle que je peinais à saisir. Je lui dois beaucoup car sans qu'il ne le sache, il m'a permis de passer à une vitesse supérieure dans ma thèse. Je le remercie vraiment. Je ne pourrais oublier de mentionner Mme Nathalie Picard qui a daigné être examinatrice de ma soutenance. Ma reconnaissance envers elle est grande; elle a montré très vite un intérêt pour mes recherches, et les différentes entrevues qu'elle m'a accordées m'ont permis de mieux comprendre nombre de points d'ombre de mes travaux qui m'échappaient et cela m'a donné progressivement confiance et assurance. Je la remercie très sincèrement. Mes remerciements s'adressent également à M. Pierre André pour avoir accepté être examinateur de ma soutenance. Les échanges que j'ai eus avec lui sur mes travaux m'ont été très constructifs. Ils m'ont permis d'avoir une idée sur la perception extérieure de mes recherches et d'affiner l'expression de mes idées de manière plus précise et intelligible. Je le remercie pour cela et pour sa disponibilité. Je tiens enfin à remercier Mme Anne Solaz, examinatrice de ma soutenance et qui m'honore également par sa présence. Son appréciation de ma thèse m'offre l'opportunité d'avoir un regard sous un autre angle sur mes travaux. Je la remercie du fond du coeur pour la considération qu'elle m'accorde.

J'adresse également mes remerciements à M. Gabriel Desgranges, directeur de l'Ecole Doctorale EM2PSI de CY Cergy Paris Université et M. Olivier Charlot, directeur du Laboratoire THEMA, cadres au sein desquels j'ai effectué cette thèse . Je remercie tous les membres du THEMA et en particulier Mme Laurence Jacquet co-membre de mon comité du suivi de thèse avec Mme Nathalie Picard. Je remercie M. Arnaud Lefranc. Je tiens à remercier tout le secrétariat du THEMA constitué de Lisa Collin, Cécile Moncelon, Imen Defferrard et Yolande Gustave dont l'accueil et l'écoute m'ont permis une intégration facile au THEMA. Ma collaboration avec Giovanni Lazzarini pour l'enseignement des Maths Appliquées à la gestion et les Fonctions Plusieurs Variables me manquera très certainement. Je tiens à lui témoigner toute mon amitié. Je n'oublie pas mes collègues de bureau, Salimeh, le dorénavant docteur Arnaud Daymard, mon cher ami Anderson, et la toujours joviale Gabrielle, notre bureau, notre Dream Team comme nous l'appelons, fut pour moi un cadre agréable de travail et d'amitié avec vous. Je veux exprimer également mon amitié à Abderhman; des bancs de l'Ecole d'Economie de Toulouse, nous nous sommes tous deux retrouvés doctorants de CY Cergy Paris Université. Je tiens également à remercier Jérémie Juste pour son aide sur le logiciel Knitr; je n'oublierai pas l'amitié et la gentillesse de Fatoumata qui n'a pas hésité à me prêter pendant l'interminable Grand Confinement son Fumio Ayashi d'économétrie que j'ai beaucoup lu et qui m'a été très utile. Je n'oublierai pas non plus Louis, Morritz, Samia, Flore, Aicha, Herman, Eloise, Patrick, le grand Ibrahim, Julien, Rozenn, Pauline, Erwan, le très travailleur Sangwon ni ceux dont le nom m'échappe mais que je n'oublie pas. Vous avez constitué pour moi dans votre ensemble un cadre d'amitiés et de travail durant ma thèse.

Mon départ, après mon baccalauréat, de Cotonou (Bénin) vers Paris pour la poursuite de mes études, a été possible grâce à un homme qui a cru en moi et qui m'a accordé sa bourse de La Fondation de France: cet homme c'est bien Odon Vallet. De mes Classes Prépa Maths Physique au Lycée Louis Le Grand jusqu'à l'Ecole d'Economie de Toulouse, il fut mon tuteur. Son caractère paternel encourageant l'excellence m'a permis de prendre conscience que seul le travail acharné garantit la constance dans l'excellence et la réussite et que le mérite de tout honneur devrait se démontrer. Je lui suis surtout très reconnaissant pour sa bonté et sa grande générosité sans limite. Je veux également exprimer ma gratitude et bien plus que mon amitié, ma fraternité envers Espéran Padonou. Je lui dois beaucoup et je ne pourrai oublier cela. J'adresse tout aussi mes remerciements envers sa maman Rita pour ses prières et la considération qu'elle m'accorde et qui me touche tant. Je remercie également le Père Armogathe pour son accueil chaleureux au foyer Bossuet à Paris et le suivi de mes études en Prépa à Louis Le Grand. Je veux exprimer une reconnaissance toute particulière à ma professeure de Mathématiques en Prépa MPSI à Louis Le Grand, Anne-Laure Biolley, virtuose des mathématiques dont la maîtrise et l'élégance dans cette science m'émerveillaient et grâce à qui j'ai développé et nourri une passion profonde pour cette discipline jusque, plus tard dans mes études économiques. Elle fut dans ma formation supérieure en Mathématiques, une professeure majeure sinon la meilleure que j'ai eu la chance d'avoir et que je n'oublierai jamais. Je remercie également Thierry Magnac et Hélène Bourguignon pour m'avoir donné la passion de l'économétrie et pour m'avoir initié à la recherche empirique en Economie en dirigeant mon mémoire de Master 2 à TSE. Ils ont fait preuve de patience avec moi, et surtout d'une grande pédagogie et d'une bienveillance à mon égard. Je leur exprime toute ma reconnaissance et mon amitié. Ce fut pour moi une opportunité qui me permit d'apprendre concrètement davantage en Econométrie et plus particulièrement en Micro-économétrie appliquée. Ce fut un tremplin très préparateur vers ma thèse.

Je tiens à remercier également un ami de longue date Henri de Parseval dont la constante amitié indéfectible a traversé toutes ces longues années. Il m'a soutenu dans des moments difficiles et il est un repère permanent pour moi. Je ne pourrais le remercier assez. Je remercie mon ami Alexandre pour son soutien et ses conseils pour mes objectifs professionnels, nos challenges sportifs et nos discussions sur des questions économiques et statistiques dont je mesure l'effet positif sur ma productivité. Je n'oublie pas La Première Compagnie de boursiers Odon Vallet de La Fondation de France; au milieu d'eux je me suis senti plus d'une fois dans mon Bénin lointain. Je n'oublie pas non plus ma bande d'amis dont les soirées gastronomiques et de jeux m'ont permis de sortir de temps en temps de ma thèse et de respirer; je veux nommer Ilénia, Ramon, Nicolas, Aicha, Grégory, Margaux et Sam. Nos soirées m'ont permis de découvrir davantage les gastronomies francaise et italienne mais aussi différents jeux de société. Je ne pourrai oublier les anniversaires surprises que vous m'aviez organisés et bien d'autres moments de partage qui ont renforcé notre belle amitié. Je veux remercier également Vicenza pour sa grande gentillesse et son accueil chaleureux durant mon séjour à Roccapiemonte à Salerno en Italie. Ces vacances me furent importantes pour découvrir la beauté de l'Italie et retrouver ensuite mon inspiration perdue.

Mes remerciements s'adressent également à Eugène et particulièrement à Germaine pour leur affection et le soutien sans faille et constant qu'ils m'ont manifesté tout au long de cette thèse et pour leurs prières et conseils. Ils m'ont permis de garder confiance et de persévérer. Je remercie également Frangilo, Yael et Honorine.

Je remercie mes parents Athanase et Marie-Franck dont l'amour et l'admiration que je leur porte me forcent de rendre fiers. Ma joie est la leur et mes peines leur souffrance. Leur foi en moi me fait avancer. Leur soutien et leur amour me fortifient dans mes faiblesses, me relèvent pour persévérer et poursuivre mes objectifs. Je veux leur témoigner tout mon amour pour eux. Je tiens à remercier également mon frère Yves pour ses précieux conseils et nos discussions scientifiques sur l'économie, les mathématiques et la physique, sans oublier mes soeurs Bérénice, Marjolaine, Gloria et Selma. Je remercie également mon oncle Paul et sa femme, mais aussi Cédric, Hélène, Paule-Annabelle, Pierre-Mathieu et Mathieu pour leur acceuil chaleureux et leur générosité; ils sont ma famille et grâce eux j'ai pu passer des réveillons inoubliables. Je témoigne également ma reconnaissance envers mes oncles Amos et Georges pour leur investissement inconditionnel dans mon éducation et mon instruction élémentaire et primaire. J'exprime ma fraternité et mon amitié envers mon cousin le plus illustre, le grand Eric pour nos débats entre économiste et démographe. Je remercie enfin tout particulièrement celle qui partage ma vie, Branda. Elle m'a appris à me surpasser, à être responsable et à être meilleur. Elle a été tout au long de ma thèse mon havre de paix.

A mes parents Marie-Franck et Athanase, je dédie ce travail.

Contents

1 General Introduction

Over the last decades, and this in all the industrialized countries, the family structure has undergone enormous changes. In the 1960s, the traditional family in France or the United States, for example, was still organized around a married couple and their children. The man was the only active person in the labor market, and the main, if not the only source of income of the household, while household chores were performed almost entirely by women. Today, the proportion of single people in the population has significantly increased, while the fertility rate has declined. The number of divorces is much more important than in the past while the number of marriages is declining. In married couples, the levels of education of husband and the wife are often very close, and that of the wife regularly exceeds that of the husband. Moreover, the wife's rate of participation in the labor market has increased significantly. And even if there is still a clear difference between men and women, the distribution of household chores within the household is much more egalitarian now than thirty or forty years ago.

One of the important changes we can focus on is the homogamy. Has it changed over the time ? and how ? These are the questions that retain our entire attention in the first chapter of this thesis. These questions are relevant for that actually, an increase of the homogamy would imply a reinforcement of the inequalities. In France, several studies, particularly in Sociology, have been carried out on this subject. The first study was carried out by Alain Girard (1964) for INED on the context and determinants of the choice of spouse. This study will be extended by Michel Bozon and Francois Heran (2006). Forsé and Chauvel (1995), Goux and Maurin (2003), Vanderschelden (2006) and Bouchet-Valat (2011, 2014) deal with the evolution of homogamy in France. Forsé and Chauvel (1995) used a measure derived from the odds ratio, the net diagonal index, and data from the survey "Emploi" 1989. They identified a decline in homogamy of social origin and stability of the homogamy of diploma in relative terms. A limit of this study is that the net diagonal index does not weight the odds ratios of the homogamy table in terms of the numbers they represent in the average population over the period. Using such a weighting, Bouchet-Valat (2011) observed a slight weakening of the degree of homogamy in the same survey. Using log-linear and log-multiplicative models, Vanderschelden (2006) studied homogamy of education and socio-professional homogamy, using the survey "Etude de l'Histoire Familiale" 1999. She measured an irregular decline in educational homogamy, but no change in the level of socio-professional homogamy (both in relative terms). But regarding educational homogamy, only the first unions that had not experienced separation before the survey were taken into account in her study. Using different models based on odds ratios and using data from "Employment" surveys carried out between 1969 and 2011, Bouchet-Valat (2014) shows that the homogamy of diploma and social class has clearly decreased, both in absolute and relative terms. Theses models are fundamentally statistical and do not have an economic approach of marriage; this will be the main point of difference of our study from theirs. To do this, we need a framework of modeling marriage.

The interest of economists for marriage and its modeling is not new since it dates back at least to the work of Becker (1973, 1974). Economic models of marriage that have been developed along this line are based on the idea that each person obtains some (positive or negative) utility from marriage. These models differ depending on the hypothesis made on the transferability of the utility. They rely on a concept of stable equilibrium defined so that:
(a) there is no married person who would prefer to be single,
(b) there are no pairs of people, married or not, who would prefer forming a couple together.
These models are said "frictionless" (that is, matching instantly reach the equilibrium without the mechanism leading to this equilibrium being specified).

Recent researches led to give empirical content to those models. Using a very simple model of marriage, known as with (perfectly) transferable utility, Choo and Siow (2006a, b) assume that the utility derived from marriage breaks down into two parts: the first is deterministic and takes a finite number of values based on observable characteristics (as, for example, the diploma, religion, income, expressed in the form of discrete variables) and the second consists of a continuous random term representing what is not observable by the analyst. They then show that the net gain (compared to the status of single) generated by marriage for a person of a given sex s having a certain characteristic c was equal, on average, to the logarithm of the ratio of the number of singles with sex s and characteristics c. By simply comparing the proportion of marriage of different groups of individuals, it becomes possible to measure the value for these groups of individuals to have certain intellectual, physiological or physical traits. It is also possible to assess the effect on the net gain of marriage of changes in the number of men or women with different characteristics. In this perspective, Choo and Siow (2006a,b) presented empirical results using data from the U.S Census and the Canadian Census.

The hypothesis of (perfectly) transferable utility means that the utility of an individual may be transferred to another individual at a constant rate. To fix ideas (but the interpretation of transferability is more general), individuals can exchange money or services that represent changes in the level of utility. Naturally, this hypothesis implies restrictions that are important. Whatever it may be, this model of marriage may provide answers to some important questions. For example, Chiappori, et alii (2014) showed by using U.S Census data that, since education for women also had an effect on marriage opportunities, the global rate of return of education has proved to be larger for women than for men. These empirical models of marriage market are generalized by Galichon and Salanié (2012).

In the first chapter of this work, we will present these models and reproduce some empirical results presented with French data. Considering the Choo and Siow model, we empirically measure the marriage surplus and the educational endogamy on the global metropolitan french marriage market and on its different geographic regions. We use French data census from 1962 to 2011. In the first section, we describe the general theoretical model of Galichon and Salanié (2012) inspired by the Choo and Siow model (2006). This is followed by the inference approaches: nonparametric and parametric. In the third part, we present the empirical results. We particularly observe that the joint surplus has increased on the period 1962-1982 and has decreased significantly on the period 1982-2011. More precisely, the intensity of increase of the joint surplus on the period 1962-1982 has been higher for very educated couples than for the less educated ; and the decrease of the joint surplus on the period 1982-2011 has been higher for less educated couples than for the very educated couples. About the endogamy, we remark that the endogamy presents a significant increase on the period 1962-1999 and a weak decrease on the period 1999-2011. On the period 1962-1982, the intensity of the increase of endogamy has been higher for very educated couples than for less educated couples. But on the period 1982-1990, the reinforcement of the endogamy within less educated couples became higher than within very educated couples. And finally, the decrease of endogamy on the period 1999-2011 has been higher for less educated couples than for the very educated ones. Our results are conform to the existence a trend with a form of an inverted U for the evolution of the educational homogamy in France, widely observed for The U.S and Canada. In that way we confirm the conclusion of the studies of Schwart and Mare (2005) and Hou and Myles (2008) on the evolution of the educational homogamy in U.S and Canada in the French case. But these results are in contrast with the global decrease observed by sociological studies in France. In our work, the educational homogamy has decreased only on the period 1999-2011 after its significant increase on the period 1962-1999. The divergence of our results from sociological studies may come essentially come from the difference of the approaches adopted to answer to the common question of educational homogamy in France.

The increase of the endogamy mainly on the period 1962-1999 should have implied a reinforcement of inequalities and particularly of the income inequality between wealthy and poor households. One may then be interested on the link between marriage and income inequality. Indeed, Greenwood et alii (2014) find that, if mating patterns in the U.S. had remained the same as in 1960, the Gini coefficient would have been significantly lower in 2005. This leads us to investigate on the impact of marriage on the income inequality. This will be the fundamental aim of Chapter II of our work.

The socio-demographic changes observed during the last fifty years has, rather unsurprisingly, had consequences on the intra-household distribution of consumption, so that growth in income inequality actually hides more ambiguous trends in individual consumption. Lise and Seitz (2011) use structural methods to recover individual consumption from survey data and show that, in the U.K., inequalities in terms of individual consumption have increased more slowly than income inequalities. The shrinking wage gap among spouses explains a more equal distribution of total consumption within the household. Bargain et alii (2019) confirm this result with another identification strategy and stress that progress in education and in living standards are the key drivers of the increase in women's consumption.Since the proportion of single persons and couples has changed significantly in recent years, and the costs and benefits of forming a couple have certainly not remained constant, it is difficult to correctly assess the observed increase in inequality. In this second chapter, we assume that individual utility is perfectly transferable and, for single individuals, is equal to observable income, i.e., we use a particularly convenient cardinalization of utility. It means that utility, like money, can be redistributed between individuals at a constant rate. This formulation is convenient because it gives a way to compare utility of individuals living alone and living in a couple. To identify individual utility, we then use an empirical marriage model à la Choo and Siow (2006) who start from the assumption that the surplus derived from marriage breaks down into two parts: the first is deterministic and depends on a finite number of observable individual characteristics and the second consists of a continuous random term, with an extreme value distribution, representing what is not observable by the analyst. More precisely, in order to take into account the continuity of individual incomes, we propose a continuous extension of the Choo and Siow 's model. We consider the extension of this model proposed by Dupuy and Galichon (2014) who account for continuous observable individual characteristics (instead of finite number of characteristics). Our contribution is two-fold. In the initial formulation of the Dupuy and Galichon's (2014) marriage model, single individuals are ignored, i.e., all individuals are assumed to be married. Firstly, we propose in the light of Dupuy and Galichon's model, an alternative approach to the case of presence of single individuals and that is empirically tractable. We show that the individual utilities can be identified from observed marriages and propose a Maximum Likelihood methodology to estimate the parameters of the model. The integration of singles is crucial because our main objective is to compare inequality among single individuals and married individuals. Secondly, we estimate this model using data from the PSID over the period 1968-2001. We show a reduction of income inequality by marriage and this inequality index has slightly increased on the period of the study. We have also measured the conversion rate between marriage surplus and income. We interpreted it as the "monetary value of marriage". It has increased on the period 1968-1971, decreased on the period 1971-1985, then slightly increased on the period 1985-1990 before decreasing on the period 1990-1999. It seems then to be on an increasing dynamics from 1999 to 2001.

The continuous matching model we have presented in Chapter II is slightly different from the Dupuy and Galichon's model (2014). In Chapter III, we try to show firstly a correspondence between the two approaches and secondly the advantage we try to bring by our formalization to find analytically the equilibrium of the matching market. To do so, we consider a bipartite matching market with transferable utility. We firstly show the equivalence between the Dupuy-Galichon model and the model presented in Chapter II, in the case of nearly full matching and secondly we investigate in that particular state of the market, on analytical solution of the matching market equilibrium i.e the optimal matching and the individual surpluses gained by the agents. Dupuy and Galichon (2014) show that the individual surpluses depend on two potentials function $a(x)$ and $b(y)$ that can be retrieved up to a constant that we called The Dupuy-Galichon Constant. Bojilov and Galichon (2015) with the setting of Dupuy and Galichon, show that the assignment problem can be solved analytically in quadratic specification with gaussian distributions of the observable characteristics. And actually, they provide closed forms of the optimal matching and identify clearly the affinity matrix of their specification. But they did not provide analytical solution for the individual surpluses. We show that the model proposed in Chapter II provides theoretical relations for these individual surpluses and we show that in the case of nearly full matching, with a quadratic specification, with gaussian distributions of the observable characteristics, we find analytically the optimal matching distribution, and the affinity matrix and we retrieve the Theorem 2 of Bojilov and Galichon (2016), but furthermore, we find also analytical expressions for the individual surpluses. We also investigate the functions $a(x)$ and $b(y)$ and we found their expression and then we also the expression of the Dupuy-Galichon Constant up to which the functions $a(x)$ and $b(y)$ are determined and we prove that this constant is actually exogenous and is related to single individuals. we also show that the model proposed in Chapter II remains available when the matching tends to be scarce and we provide the analytical solution of this extreme state of the market. We present two ways to estimate parametrically the model, by maximum likelihood eased by the fact that its is completely expressed analytically, and the estimation by moments based on the identification of the affinity matrix.

In Chapter IV, we examine consumption and marriage decisions of individuals in a unified framework. To do so, we assume that the marriage market is stable according to Gale and Shapley (1962)'s definition and that the decision of the couple can be described by the collective approach, i.e, each individual has its own preference and the outcome of the decision process is Pareto efficient. The collective approach has been introduced to study the behavior of multi-individual households by Chiappori (1988, 1992) and, since then, has rapidly gained in popularity. Bourguignon, Browning and Chiappori (2009) provide the main theoretical results for the case of constant prices. They show that the collective approach generates testable restrictions and that, if individual preferences are egoistic and consumed goods are private, the intrahousehold distribution of resources can be recovered. Browning and Chiappori (1998) incorporate prices. Blundell, Chiappori, and Meghir (2005) and Donni (2012) extend the collective model to public goods. Chiappori and Ekeland (2007, 2009) provide a complete characterization of the collective model. Cherchye, De Rock, and Vermeulen (2011) and Cherchye et al. (2015) propose an alternative approach to the identification of the intrahousehold distribution of resources based on the theory of revealed preferences. Empirical applications have shown that the intra-household distribution of resources depends on variables such as individual wages, prices, or distribution factors. See, for example, Bargain and Donni (2012), Bourguignon, Browning, and Chiappori (1994), Browning, Chiappori, and Lewbel (2013), Cherchye, De Rock, and Vermeulen (2012), Couprie, Peluso, and Trannoy (2010), Lewbel and Pendakur (2008), Lise and Seitz (2011) and Dunbar, Lewbel, and Pendakur (2013). However, the precise mechanism behind the intra-household distribution of resources is not clearly identified. As pointed out by Browning, Chiappori, and Weiss (2014, p. 122), the collective approach remains "agnostic" about the specific intra-household bargaining process that generates Pareto efficient outcomes. One of the rare structural models that attempt to explain how resources are shared among household members is proposed by Cherchye, Demuynck, De Rock, and Vermeulen (2017). They integrate the assumption of a stable marriage market with the collective consumption model to analyze the choice behavior of households. They then use a revealed preferences approach and identify the sharing rule of household resources, which is thus determined by the state of the marriage market. The work presented in Chapter IV is related to Cherchye, Demuynck, De Rock, and Vermeulen

(2017)'s contribution but differs from it in at least two important aspects. Instead of using a revealed preferences approach, we construct a model of marriage market with transferable utility inspired from that of Choo and Siow (2006) and its sequels. In addition, we do not use information on household consumption but we identify consumption from the observation of matching patterns. More precisely, our approach is based on Dupuy and Galichon (2014) and its variation proposed in Chapter II. The individual utility is represented as the sum of a deterministic part and a stochastic part. The deterministic part corresponds to the consumption-based utility while the stochastic part can be seen as a "sympathy shock". From traditional results in the marriage market literature, the deterministic part is identified from the observation of matching patterns. The basic idea of the model is then simple. If we assume that the deterministic part, i.e., the consumption-based utility function of each individual, has a parametric form compatible with transferable utility, namely, a Generalized Quasi-Linear form (Bergstrom and Cornes, 1983), that depends on the consumption of a private and a public good, then the parameters of the consumption-based utility function can be recovered. Once the parameters are recovered, the consumption of the private and the public good can be identified. The parametric specification of individual-based utility functions has at least three applications. (i) If single individuals have the same deterministic preferences as married individuals, then the consumption-based utility function of single individuals are identified from that of married individuals. This potentially allows comparing the changes in welfare of single individuals as well as that married individuals. To make a comparison with the results in Chapter II, the unidentified constant that represents the exchange rate between utility and money can be identified. (ii) The changes in endogamy observed during the second half of the 20th century can be structurally interpreted as a change in the individual taste for public goods together with a change in the distribution of individual incomes. (iii) From a more theoretical perspective, the identification of individual consumption suggests that observed matching patterns contain information on household consumption. It is a first step toward the unification of collective models and marriage models. The theoretical model is illustrated by an empirical application using data from the PSID for the period 1968-2001. We find that, unsurprisingly, men and women have different parameters for the consumption-based utility functions. To be more precise, women have a higher taste for the public good than men. Consequently, the average gain from marriage is higher for men than for women. We also find that married individuals have a higher level of consumption and a higher level of surplus than singles. In addition, inequality is higher among single individuals than among married individuals. The study thus confirms the conclusion we drew in Chapter II, i.e., marriage reduces inequality in terms of consumption as well as surplus. Over the period that we examine, we observe a general increase in (consumption and surplus) inequality, with analogous dynamics to the incomes dynamics. Finally, we compute a parameter measuring the exchange rate between marriage surplus and money(which can be interpreted as the monetary value of marriage). This parameters varies over the examined period: it increases between 1968 and 1970, drastically decreases between 1970 and 1983, and it has been stable between 1983 and 2001. The method used to compute this value is an alternative to what we suggested in Chapter II. Although, the two measures give us very similar results in terms of dynamics over the time of the study. The Chapter III is structured as follows. In Section 1, we describe the marriage market and define the stability conditions in the sense of Gale and Shapley (1960). In Section 2, we identify the individual consumption-based utility of individual within households in function of the primitives of the model. In Section 3, we consider the case of the parametric Generalized Quasi-Linear utility function and we show that the parameters of the function can be retrieved. In Section 4, we present estimates of the model on the PSID data on the period 1968-2001.

2 Marriage Surplus and Educational Endogamy in France

Abstract

This article deals with the fundamental question of educational homogamy in France. One interest in this interrogation find its root in the fact that the global education level has highly increased during the last decades and particularly women'. This demographic changes may have influenced individuals well-being in some way. This article aims to highlight some of the eventual effects particularly within households in France.

2.1 Introduction

The family structure has undergone enormous changes in all the industrialized countries Over the last 50 years. The number of divorces is much more important than in the past while the number of marriages is declining. In married couples, the levels of education of husband and the wife are often very close, and that of the wife regularly exceeds that of the husband. Moreover, the wife's rate of participation in the labor market has increased significantly. And even if there is still a clear difference between men and women, the distribution of household chores within the household is much more egalitarian now than thirty or forty years ago. One of the important changes we can focus on is the homogamy more especially the educational homogamy. An increase of this would imply a reinforcement of the inequalities. Several studies have been done in different industrialized countries to investigate on the evolution of the homogamy, mainly in Sociology. In that way, we can mention some of those studies. Mare (1991) has studied the variations of the educational homogamy in U.S. on the period 1940-1987 and found an increase until the 70s. Schwartz and Mare (2005) using U.S data as well and covering a long period, confirmed a reinforcement of the educational homogamy until the 1990s and a decrease then. Similar results are found with Canadian data (Hou and Myles, 2008). More generally, a study of Ultee and Luijkx (1990) on 23 industrialized countries revealed a weakening of the educational over the time with a trend with a form of an inverted U, i.e an increase followed by a decrease over the time. That suggests that the educational homogamy has first increased with the increase of the education rate, and then has decreased. This is confirmed by Smits, Ultee and Lammers (1998). In Europe, we observe different evolutions of the educational homogamy: for instance, a decrease in U.K, Sweden, and Norway (Halpin and Chan (2003), Birkelund ans Heldal (2003), Henz and Jonsson (2003)) and a decrease in France and The Netherlands too (Ultee and Luijkx, 1990) and in parallel, a global increase of the educational homogamy is observed in Italy (Bernardi, 2003) and Spain (Esteve and Cortina, 2006). In France, several sociological studies, have been carried out on this subject. The first study was carried out by Alain Girard (1964) for INED on the context and determinants of the choice of spouse. This study will be extended by Michel Bozon and Francois Heran (2006). Forsé and Chauvel (1995), Goux and Maurin (2003), Vanderschelden (2006) and Bouchet-Valat (2011, 2014) deal with the evolution of homogamy in France. Forsé and Chauvel (1995) used a measure derived from the odds ratio, the net diagonal index, and data from the survey "Emploi" 1989. They identified a decline in homogamy of social origin and stability of the homogamy of diploma in relative terms. A limit of this study is that the net diagonal index does not weight the odds ratios of the homogamy table in terms of the numbers they represent in the average population over the period. Using such a weighting, Bouchet-Valat (2011) observed a slight weakening of the degree of homogamy in the same survey. Using log-linear and log-multiplicative models, Vanderschelden (2006) studied homogamy of education and socio-professional homogamy, using the survey "Etude de l'Histoire Familiale" 1999. She measured an irregular decline in educational homogamy, but no change in the level of socio-professional homogamy (both in relative terms). But regarding educational homogamy, only the first unions that had not experienced separation before the survey were taken into account in her study. Using different models based on odds ratios and using data from "Employment" surveys carried out between 1969 and 2011, Bouchet-Valat (2014) shows that the homogamy of diploma and social class has clearly decreased, both in absolute and relative terms. Theses models are fundamentally statistical and do not have an economic approach of marriage; this will be the main point of difference of our study from theirs.

The interest of economists for marriage and its modeling is not new since it dates back at least to the work of Becker (1973, 1974). Economic models of marriage that have been developed along this line are based on the idea that each person obtains some (positive or negative) utility from marriage. These models differ depending on the hypothesis made on the transferability of the utility. They rely on a concept of stable equilibrium defined so that:
(a) there is no married person who would prefer to be single,
(b) there are no pairs of people, married or not, who would prefer forming a couple together.
These models are said "frictionless" (that is, matching instantly reach the equilibrium without the mechanism leading to this equilibrium being specified).

Recent research led to give empirical content to those models. Using a very simple model of marriage, known as with (perfectly) transferable utility, Choo and Siow (2006a, b) assume that the utility derived from marriage breaks down into two parts: the first is deterministic and takes a finite number of values based on observable characteristics (as, for example, the diploma, religion, income, expressed in the form of discrete variables) and the second consists of a continuous random term representing what is not observable by the analyst. They then show that the net gain (compared to the status of single) generated by marriage for a person of a given sex s having a certain characteristic c was equal, on average, to the logarithm of the ratio of the number of singles with sex s and characteristics c. By simply comparing the proportion of marriage of different groups of individuals, it becomes possible to measure the value for these groups of individuals to have certain intellectual, physiological or physical traits. It is also possible to assess the effect on the net gain of marriage of changes in the number of men or women with different characteristics. In this perspective, Choo and Siow (2006a,b) presented empirical results using data from the U.S Census and the Canadian Census.

The hypothesis of (perfectly) transferable utility means that the utility of an individual may be transferred to another individual at a constant rate. To fix ideas (but the interpretation of transferability is more general), individuals can exchange money or services that represent changes in the level of utility. Naturally, this hypothesis implies restrictions that are important. Whatever it may be, this model of marriage may provide answers to some important questions. For example, Chiappori and alii (2014) showed by using U.S Census data that, since education for women also had an effect on marriage opportunities, the global rate of return of education has proved to be larger for women than for men. These empirical models of marriage market are generalized by Galichon and Salanié (2012).

In this article, we will present these models and reproduce some empirical results presented with French data. It may be interesting to look at how the surplus generated by marriage has changed over time. The reduction in the number of marriages in France suggests that this surplus has declined. However, it is possible to go further and to identify groups of people and the types of marriage that have been most affected by changes. For example, Chiappori et alii (2014) argue that the increase in the correlation between spouses' income generally observed in recent years is not due to a amplification in the taste for assortative matchings. This increase is a mechanical consequence of the rise in the general level of education of women: a larger number of graduated men marry a larger number of graduated women because the latter are more numerous than before.

In this paper considering the Choo and Siow model, we empirically measure the marriage surplus and the educational endogamy on the global metropolitan french marriage market and on its different geographic regions. We use French data census from 1962 to 2011. In the first subsection, we describe the general theoretical model of Galichon and Salanièè (2012) inspired by the Choo and Siow model (2006). This is followed by the inference approaches: nonparametric and parametric. In the third part, we present the empirical results. We particularly observe that the joint surplus has increased on the period 1962-1982 and has decreased significantly on the period 1982-2011. More precisely, the intensity of increase of the joint surplus on the period 1962-1982 has been higher for very educated couples than for the less educated ; and the decrease of the joint surplus on the period 1982-2011 has been higher for less educated couples than for the very educated couples. About the endogamy, we remark that the endogamy presents a significant increase on the period 1962-1999 and a decrease on the period 1999-2011. On the period 1962-1982, the intensity of the increase of endogamy has been higher for very educated couples than for less educated couples. But on the period 1982-1990, the reinforcement of the endogamy within less educated couples became higher than within very educated couples. And finally, the decrease of endogamy on the period 1999-2011 has been higher for less educated couples than for the very educated ones. Our results are conform to the existence a trend with a form of an inverted U for the evolution of the educational homogamy in France, widely observed for The U.S and Canada. In that way we confirm the conclusion of studies of Schwart and Mare (2005) and Hou and Myles (2008) on the evolution of the educational homogamy in U.S and Canada in the French case. But these results are in contrast with the global decrease observed by sociological studies in France. In our work, the educational homogamy has decreased only on the period 1999-2011 after it has increased on the period 1962-1999.

2.2 Theoretical Framework

At the heart of this theory is a two - sided assignment model with transferable utilities where agents on both sides of the market (men and women) are characterized by a set of attributes that is only partially observed by the econometrician. Each agent aims to match a member of the opposite sex in order to maximize his own utilities. This model is particularly interesting because under certain conditions (mainly restrictions on the form of the surplus function), it is possible to identify and estimate the characteristics of agent preferences. We suppose the set of observable attributes is discrete.

2.2.1 Marriage Market Description

We consider a bipartite matching market with transferable utility. We start with two sets $\mathcal{H}$ and $\mathcal{F}$, whom we will refer to from now on as set of "men" and set of "women". Let P and Q be two measures defined respectively on $\mathcal{H}$ and $\mathcal{F}$. Any individual $m \in \mathcal{H}$ can be matched with an individual $w \in \mathcal{F}$ or remain single, and conversely. Remaining single can be considered as a match to a null agent of the other side of the market. So, we add two dummy populations of null agents, denoted 0 to $\mathcal{H}$ and $\mathcal{F}$ respectively in order to take in to account single individuals. $\mathcal{P}$ represents the total population.

$$|\mathcal{H}| + |\mathcal{F}| = |\mathcal{P}|$$

We suppose that there exists two random variables X and Y defined respectively from $\mathcal{H}$ to a discrete finite space $\mathcal{X}$ and from $\mathcal{F}$ to a discrete finite space $\mathcal{Y}$.

We will adopt the Galichon and Salanié (2012) framework for its simplicity. Let $\mathbf{M}$ and $\mathbf{N}$ be the vectors of the numbers of men and women.

$$\mathbf{M} = (M_x)_{x\in\mathcal{X}} \text{ and } \mathbf{N} = (N_y)_{y\in\mathcal{Y}}$$

where M_x represents the number of men of type x and N_y the number of women of type y. In the case the population is infinite we can consider the probabilities $(m_x)_{x\in\mathcal{X}}$ and $(n_y)_{y\in\mathcal{Y}}$ such that:

$$\forall x \in \mathcal{X},\ \forall y \in \mathcal{Y},\ m_x = Pr(X = x) \text{ and } n_y = Pr(Y = y)$$

In other terms, m_x is the proportion of men of type x in the population of men and n_y is the proportion of women of type y in the population. In the light of Galichon and Salanié (2012) we introduce the following notions:matching, feasible matching and aggregated matching matrix.

Definition 2.1 (Matching)**.** *A matching is an application $\tilde{\mu}$ defined on the product space $\mathcal{H} \times \mathcal{F}$ to $\{0,1\}$ such that, for all $(m,w) \in \mathcal{H} \times \mathcal{F}$, $\tilde{\mu}(m,w) = 1$ if man m and woman w are matched and $\tilde{\mu} = 0$ if they are not matched. Formally we have:*

$$\tilde{\mu} : \begin{array}{ccc} \mathcal{H} \times \mathcal{F} & \longrightarrow & \{0,1\} \\ (m,w) & \longmapsto & \boldsymbol{1}\{\ m \text{ matched to } w\ \} \end{array}$$

Intuitively, matching must satisfy some constraints to be feasible. A matching is feasible if and only if any individual is matched with at most one individual of the opposite side.

Definition 2.2 (Feasible Matching)**.** *A feasible matching is a matching such that:*

$$\forall (m,w) \in \mathcal{H} \times \mathcal{F}, \tilde{\mu}(m,w) = 1 \Rightarrow \forall m' \neq m,\ \forall w' \neq w, \tilde{\mu}(m',w) = \tilde{\mu}(m,w') = 0$$

We introduce now the aggregated matching matrix.

Definition 2.3 (Aggregated Matching Matrix). *Let $\tilde{\mu}$ be a matching. The aggregated matching matrix μ associated to $\tilde{\mu}$ is defined as:*

$$\forall\,(x,y) \in \mathcal{X} \times \mathcal{Y},\ \mu_{xy} = \sum_{m,w} \mathbb{1}(x_m = x, y_w = y)\tilde{\mu}(m,w)$$

The number μ_{xy} represents the number of realized marriages of type (x,y). In the case of feasible matching, we have:

$$\forall x \in \mathcal{X},\ \sum_{y \in \mathcal{Y}} \mu_{xy} \leq M_x;\ \forall y \in \mathcal{Y},\ \sum_{x \in \mathcal{X}} \mu_{xy} \leq N_y$$

For any type x of men, the number of singles men of this type is the difference between the total number of men of type x and the total number of matched men of type x.

$$\forall x \in \mathcal{X},\ \mu_{x0} = M_x - \sum_{y \in \mathcal{Y}} \mu_{xy},\ \forall y \in \mathcal{Y},\ \mu_{0y} = N_y - \sum_{x \in \mathcal{X}} \mu_{xy}$$

For notation, we define:

$$\mathcal{X}_0 = \mathcal{X} \cup \{0\},\ \mathcal{Y}_0 = \mathcal{Y} \cup \{0\}$$

2.2.2 Matching Surplus

Matching Surplus And Heterogeneity The introduction of the concept of the matching surplus is essential to give an economic basis to matching. In fact, matching can be justified by the hypothesis of the existence of a surplus generated by any matching between two individuals. This surplus is a joint surplus for the two matched partners and they share it into two individual surpluses that represent their respective utilities from the current matching. The joint surplus is a primitive of the matching and its sharing is endogenous to the problem. So the main information we need to determinate equilibrium is the joint surpluses of any feasible matching.

To model the utility from marriage, we use the separability assumption suggested by Choo and Siow (2006) and formalized by Galichon and Salanié (2012) and Chiappori, Salanié and Weiss (2014).

Assumption 2.1 (Separability Assumption)**.** *The joint surplus of the union of a man m of type x with a woman w of type y is:*

$$s_{mw} = S_{xy} + \varepsilon_{my} + \eta_{xw}$$

where S_{xy} is a deterministic function defined on the cartesian space $\mathcal{X} \times \mathcal{Y}$

The term S_{xy} represents the gross systematic joint surplus after transfers. By normalization, we assume respectively for a single man m of type x and for a single woman w of type y:

$$S_{x0} = S_{0y} = 0, \text{ and } s_{m0} = \varepsilon_{m0},\ s_{0w} = \eta_{0w}$$

We can normalize the surplus of singles to 0 because only the net surplus can identified. The following assumption will specify the distribution of the random terms.

Assumption 2.2 (Unobserved heterogeneity)**.** *The random terms ε_{my} and η_{xw} in the joint surplus of the union a man m of type x with a woman w of type y are supposed to have a Gumbel extreme values distribution conditional on X and Y.*

$$\forall (m, w, x, y) \in \mathcal{H} \times \mathcal{F} \times \mathcal{X}_0 \times \mathcal{Y}_0, \varepsilon_{my} \sim Gumbel(-\gamma, 1), \text{ and } \eta_{xw} \sim Gumbel(-\gamma, 1).$$

As Galichon and Salanié (2012) showed it, this hypothesis of specification is not necessary. In fact, a more general hypothesis is to assume that the random terms ε_{my} and η_{xw} have respectively a distribution P_X and Q_Y conditional on $X$and Y. But, for simplification, we stay in the logit framework used by Choo and Siow (2006).

Equilibrium We assume that the equilibrium is defined in the sense of Gale and Shapley (1962).

Definition 2.4 (Stable Matching)**.** *A matching is stable if:*
(i) no matched individual would rather be single, and
(ii) no pair of individuals would rather be matched together than remain in their current situation.

This equilibrium is frictionless. All the agents are satisfied of their choices. The primitive of the problem is the joint surplus s_{mw}. It is the generated surplus by the union of a man m and a woman w. This quantity is shared between the two partners into two parts u_{mw} and v_{mw} that represent respectively the individual utility of man m and the individual utility of woman w from their union. From Shapley and Shubik (1971), we know that the solution of the problem is given by a feasible matching $\tilde{\mu}$ that maximizes the social surplus that is the sum of the surplus of all of the individuals.

The utilities u_m and v_w are respectively the utility a man m obtains at the stable state of the market and the utility a woman obtains at the stable state of the market. Single individuals are self matched; a single man m gets ε_m^0 and a single woman w gets η_w^0. At the equilibrium, partners are reciprocally the best choice of their spouse, and single individuals at the equilibrium haven't found any partner that brings to them a greater utility than their own utility. The following proposition showed by Galichon and Salanié (2012) characterizes the equilibrium.

Proposition 2.1 (Characterization of the sharing of the joint surplus)**.** *At the equilibrium, the respective individual utilities $u_m(x)$ and $v_w(y)$ of a man m of type x and of a woman w of type y matched to each other are such that:*

$$u_m(x) = \max\left(\varepsilon_{m0}, U_{xy} + \varepsilon_{my}\right)$$

and

$$v_w(y) = \max\left(\eta_{0w}, V_{xy} + \eta_{xw}\right)$$

where U_{xy} and V_{xy} are two functions such that $U_{xy} + V_{xy} = S_{xy}$.

Proof. Let m be a man. We denote by u_m his utility at the equilibrium.

$$u_m \geq \varepsilon_{m0}$$

In the case m does not get matched, we have $u_m = \varepsilon_{m0}$. For any woman w, with utility v_w at the equilibrium, the utility of the man m from a union with w is:

$$s_{mw} - v_w$$

As u_m is the utility of man m at the equilibrium, we have:

$$u_m \geq s_{mw} - v_w$$

In the case, m gets matched with w, we have:

$$u_m = s_{mw} - v_w$$

We can then conclude:

$$u_m = \max\left(\varepsilon_{m0}, \max_w \left(s_{mw} - v_w\right)\right)$$

By exploiting the separability assumption, we have at the equilibrium:

$$u_m = \max\left(u_m^0, \max_w \left(s_{mw} - v_w\right)\right) = \max\left(u_m^0, \max_w \left(S_{x,y(w)} + \varepsilon_{m,y(w)} + \eta_{xw} - v_w\right)\right)$$

This leads to:

$$u_m = \max\left(u_m^0, \max_y \max_{w|y_w=y} \left(S_{xy} + \varepsilon_{my} - (-\eta_{xw} + v_w)\right)\right)$$

so

$$u_m = \max\left(u_m^0, \max_y \left(S_{xy} + \varepsilon_{my} - \min_{w|y_w=y} \left(v_w - \eta_{xw}\right)\right)\right)$$

We can denote by:

$$V_{xy} = \min_{w|y_w=y} \left(v_w - \eta_{xw}\right)$$

Then we get:

$$u_m = \max\left(u_m^0, \max_y \left(S_{xy} - V_{xy} + \varepsilon_{my}\right)\right) = \max\left(u_m^0, \max_y \left(U_{xy} + \varepsilon_{my}\right)\right)$$

■

The proposition shows that the total surplus is shared into two parts between the partners. These parts are the deterministic parts of the individual utilities of the partners and sympathy shocks constitute the stochastic parts. The utility obtained at the equilibrium by an individual is the highest utility he can get from a union or of being single.

2.2.3 Identification of individual utilities

For any pair (m, w) in $(\mathcal{H} \times \mathcal{F})$ of two individuals, their matching requires the existence of a joint surplus s_{mw} that will be parted into two utilities u_m and v_w for each other. These utilities are the gains of the partners after transfers. Let $\tilde{\mu}$ be the optimal feasible matching. For any man m and woman w, the sum of their utilities u_m and v_w they can get on the market, is higher than or equal to their joint surplus s_{mw}. In fact u_m and v_w are respectively higher than or equal to the individual parts of utility the man m and the woman w can get from the sharing of s_{mw}. There will an equality $u_m + v_w = s_{mw}$ if and only if m and w are matched.

Under the assumption of the Gumbel distribution of the random terms ε_{my} conditionally to x and of η_{xw} conditionally to y, we can derive the distribution of the equilibrium utilities u_m and v_w. As shown by Galichon and Salanié (2012), the following proposition states the distribution of u_m and of v_w at the equilibrium. It is an important step in the identification of the individual surpluses.

Proposition 2.2. *Assume a stable matching. Under the above assumptions, at the equilibrium, the utility $u_m(x)$ of a man m of type x and the utility $v_w(y)$ of a woman w of type y are such that:*

$$u_m(x) \sim \textit{Gumbel}\left(-\gamma + \ln\left(1 + \sum_{y \in \mathcal{Y}} e^{U(x,y)}\right), 1\right)$$

and

$$v_w(y) \sim \textit{Gumbel}\left(-\gamma + \ln\left(1 + \sum_{x \in \mathcal{X}} e^{V(x,y)}\right), 1\right)$$

where γ is the Euler-Mascherano constant.

Proof. Let's assume F_{u_m} the distribution function of u_m. We have:

$$\forall\, c \in \mathbb{R},$$

$$\begin{aligned} F_{u_m}(c) &= P(U_{xy} + \varepsilon_{my} \leq c, \forall y \in \mathcal{Y}_0) \\ &= \prod_{y \in \mathcal{Y}_0} F_{\varepsilon_{my}}(c - U_{xy}) \\ &= \exp\left(-\exp\left(-\left[c + \gamma - \ln\left(1 + \sum_{y \in \mathcal{Y}} e^{U_{xy}}\right)\right]\right)\right) \end{aligned}$$

This is the cumulative distribution function of the Gumbel distribution stated in the proposition. ■

We adopt the notation of Galichon and Salanié (2012). Let $G_x(U_{x.})$ be the average utility of man m conditionally his type is x, where: $U_{x.} = (U_{xy})_{y \in \mathcal{Y}}$. We have:

$$G_x(U_{x.}) = E(u_m | x_m = x).$$

We deduce from the distribution of u_m:

$$G_x(U_{x.}) = \ln(1 + \sum_{y \in \mathcal{Y}} e^{U_{xy}})$$

Analogously, we denote by $H_y(V_{.y})$ the average utility of woman with type y.

$$H_y(V_{.y}) = \ln(1 + \sum_{y \in \mathcal{X}} e^{V_{xy}})$$

Following Galichon & Salanié (2015), we know that the optimal matching aims to maximize the total social surplus that is:

$$\mathcal{W} = \sum_{x \in \mathcal{X}} M_x G_x(U_{x.}) + \sum_{y \in \mathcal{Y}} N_y H_y(V_{.y})$$

under the constraint $U(x, y) + V(x, y) = S(x, y)$. This result is a consequence of the Gale-Shapley equilibrium. In fact at the equilibrium, any deviation from this position will diminish the total surplus.

The following proposition gives the identification formulas for the different systematic surpluses.

Proposition 2.3. *At the equilibrium, we have:*

$$\forall (x, y) \in \mathcal{X} \times \mathcal{Y}, U_{xy} + V_{xy} = S_{xy}$$

and the optimal aggregated matching verifies

$$\frac{\mu_{xy}}{\mu_{x0}} = e^{U_{xy}},\ \frac{\mu_{xy}}{\mu_{0y}} = e^{V_{xy}},\ \frac{\mu_{xy}^2}{\mu_{x0} \cdot \mu_{0y}} = e^{S_{xy}}$$

Proof. As pointed by Galichon and Salanié (2012), we have from the envelope theorem:

$$\frac{\partial G_x(U_{x.})}{\partial U_{xy}} = \frac{\mu_{xy}}{M_x} \Leftrightarrow \frac{\mu_{xy}}{M_x} = \frac{e^{U_{xy}}}{1 + \sum_{y \in \mathcal{Y}} e^{U_{xy}}}$$
$$\Leftrightarrow \frac{\frac{\mu_{xy}}{\mu_{x0}}}{1 + \sum_{y \in \mathcal{Y}} \frac{\mu_{xy}}{\mu_{x0}}} = \frac{e^{U_{xy}}}{1 + \sum_{y \in \mathcal{Y}} e^{U_{xy}}}$$
$$\Leftrightarrow \frac{\mu_{xy}}{\mu_{x0}} = e^{U_{xy}}$$

Analogically, we have:

$$\frac{\mu_{xy}}{\mu_{0y}} = e^{V_{xy}}$$

By using the fact that:

$$U_{xy} + V_{xy} = S_{xy}$$

we obtain:

$$\frac{\mu_{xy}^2}{\mu_{x0} \cdot \mu_{0y}} = e^{S_{xy}}$$

■

2.3 Inference

In this subsection we will describe two approaches to estimate this model: nonparametric and parametric.

2.3.1 Nonparametric estimation

We estimate the systematic joint surplus by:

$$\hat{S}_{xy} = \ln\left(\frac{\hat{\mu}_{xy}^2}{\hat{\mu}_{x0}.\hat{\mu}_{0y}}\right)$$

where $\hat{\mu}$ is the observed matching matrix.

The support spaces $\mathcal{X}$ and $\mathcal{Y}$ are discrete. We smooth this estimator by the Nadaraya-Waston Kernel estimator on the interval of the variable age:

$$\hat{S}^{NW}(s,t) = \frac{\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} K\left(\frac{s-x}{h_1};\frac{t-y}{h_2}\right) * \ln\left(\frac{\hat{\mu}_{xy}^2}{\hat{\mu}_{x0}*\hat{\mu}_{0y}}\right)}{\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} K\left(\frac{s-x}{h_1};\frac{t-y}{h_2}\right)}$$

where the function K is the Epanechnikov Kernel.

2.3.2 Parametric Inference

Maximum Likelihood estimation

We specify the systematic joint surplus S_{xy} by a parametric function of the observable characteristics X and Y.

We can estimate the parameter vector by the maximum likelihood method. The likelihood is:

$$L(\theta) = \prod_{x\in\mathcal{X},y\in\mathcal{Y}_0}\left(\frac{e^{U_{xy}^{\theta}}}{\sum_{y\in\mathcal{Y}_0} e^{U_{xy}^{\theta}}}\right)^{\hat{\mu}_{xy}} * \prod_{x\in\mathcal{X}_0,y\in\mathcal{Y}}\left(\frac{e^{V_{xy}^{\theta}}}{\sum_{x\in\mathcal{X}_0} e^{V_{xy}^{\theta}}}\right)^{\hat{\mu}_{xy}}$$

where $\hat{\mu}_{xy}$ is the observed number of unions of type (x,y). This is the product of the conditional probabilities of being married to a particular type of partner or to remain single with respect the type of the individuals. In fact the conditional probability a man of type x gets married to a woman of type y is:

$$\frac{e^{U_{xy}^{\theta}}}{\sum_{y\in\mathcal{Y}_0} e^{U_{xy}^{\theta}}}$$

and the probability he remains single is:

$$\frac{1}{\sum_{y\in\mathcal{Y}_0} e^{U^\theta_{xy}}}$$

The probability a woman of type y gets married to a man of type x is:

$$\frac{e^{V^\theta_{xy}}}{\sum_{x\in\mathcal{X}_0} e^{V^\theta_{xy}}}$$

and the probability she remains single is:

$$\frac{1}{\sum_{x\in\mathcal{X}_0} e^{V^\theta_{xy}}}$$

The main problem is that we don't know the expression of U and V. We will express the likelihood in the function of the theoretical matching and of the observed matching and by an algorithm used by Galichon and Salanié (2012) we will compute it. At the equilibrium, we have:

$$\frac{\mu^\theta_{xy}}{M_x} = \frac{e^{U^\theta_{xy}}}{\sum_{y\in\mathcal{Y}_0} e^{U^\theta_{xy}}} \text{ and } \frac{\mu^\lambda_{xy}}{N_y} = \frac{e^{V^\theta_{xy}}}{\sum_{y\in\mathcal{X}_0} e^{V^\theta_{xy}}}$$

μ^θ is the theoretical matching matrix. Therefore, we obtain:

$$L\left(\theta\right) = \prod_{x\in\mathcal{X},y\in\mathcal{Y}_0} \left(\frac{\mu^\theta_{xy}}{M_x}\right)^{\hat{\mu}_{xy}} * \prod_{x\in\mathcal{X}_0,y\in\mathcal{Y}} \left(\frac{\mu^\theta_{xy}}{N_y}\right)^{\hat{\mu}_{xy}}$$

The log likelihood is:

$$\log L(\theta) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}_0} \hat{\mu}_{xy}\, \log\left(\frac{\mu^\theta_{xy}}{M_x}\right) + \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}_0} \hat{\mu}_{xy}\, \log\left(\frac{\mu^\theta_{xy}}{N_y}\right)$$

We use the Iterated Projection Fitting Procedure proposed by Galichon and Salanié (2012) to compute the likelihood for each θ chosen in sufficiently large range. The estimator estimator is given by:

$$\hat{\theta}^{MLE} = Arg\max_\theta \log L(\theta)$$

2.4 The Data

In this paper, we use data from the French Census (*Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.1 [dataset]. Minneapolis, MN: IPUMS, 2018. https://doi.org/10.18128/D020.V7.1*), in particular the waves 1962, 1968, 1975, 1982, 1990, 1999, 2006 and 2011 to estimate the marriage surplus and its evolution over the time. Each observation in the data is individual. It precises the year of the interview, the household i.e the family unit in which the individual lives, her sex, her age, her relation with the head of the household, her matrimonial status, her educational attainment, her employment status, her work place, and her region of residence in the last census. We don't observe the current region of residence. We have also these information for her spouse in the case the individual is in union. But the spouse may not live the household. This information reveals the relationship between the various individuals of each household and allows us to reconstruct "couples". The data allow us to do spatial analysis because it contains the indication of the region of work; we will use this variable for the place of residence.

2.4.1 Variables

Our analysis variable is mainly Education. Educational attainment is measured from the respondent's reported highest level of education achieved. There are four levels:

1- No schooling or less then primary

2- Primary completed

3- Secondary completed

4- University completed

2.4.2 Identification of couples

We consider as a couple a man and a woman who live in the same household and reported as head of the household and spouse (partner) of the head. We construct four data-frames for each wave: one for married men, one for married women, one for single men and one for single women. We create in the initial data set, the variable "match" that indicates under our criteria if the individual is married or not. We keep only persons aged 17 or over.

match=1 if the individual is married and 0 if not.

We then, extract married men to construct the data for married men. We do so for married women, single men and single women. We verify that the number of married women and married men are same and they household identifiers match.

2.5 Empirical results

2.5.1 Nonparametric estimation

Individual Surpluses and Joint Surplus from marriage in France

For each level of education i.e 5 years of education, 13 years of education, 16 years of education and 19 years of education, and for men and women, we estimate the individual surplus of men and women in function of the education of their partner. Note that we obtain continuous curves on the interval [5,19] of the variable age because the nonparametric estimator is a smoothing function defined on that interval. The nonparametric estimation of the marriage surpluses for men shows that for the different levels of education of men, the choice of partner is very linked to the educative distance between the partners. In fact, we remark that very educated men (level 4) maximize their individual surplus with very educated women. Their surplus curve is increasing. The less educated men (leve1) have a decreasing surplus curve: in the case the sympathy shock is dwarf, their utility is maximized with less educated women. For intermediary levels (level 2 and 3) the curves are parabolic and concave; they obtain a maximal surplus with women who have same education level as them. Their surplus curve is increasing until the maximum then is decreasing. Their highest surplus is brought by union with women who have same education level as them. We can also remark that the level of education seems to not influence the maximum of marriage surplus men can get from their optimal union. The maximum of surplus curves for the different levels of education are approximately the same. We illustrate our analysis with the following graph that represents the individual surplus of men in France in 1999 for each level of education in function of the education of the partner.

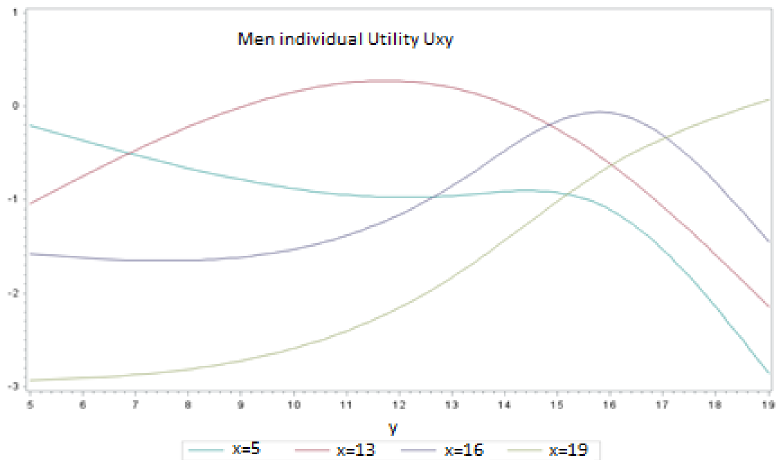


Figure 1: Men individual surplus in France in 1999

The surplus curves of women are very similar to men' curves. High educated women have high surplus with high educated men. Women with intermediary level (level 2 and 3) have their maximum surplus from union with men who have same education level as them. One difference between men and women in the partner choice is particular in the case women are less educated (level 1). These women seem indifferent to men of level 2 and 3. Their surplus is maximal with men of level 1 and it significantly decreases with high educated men (level 4). Another difference we can notice is the level of the maximal marriage surpluses. We remark that very educated women, level 3 and 4, have maximal surplus higher than the maximal marriage surplus of less educated women, level 1 and 2. The following graph represents the individual surplus of women in France in 1999 for each level of education in function of the education of the partner.

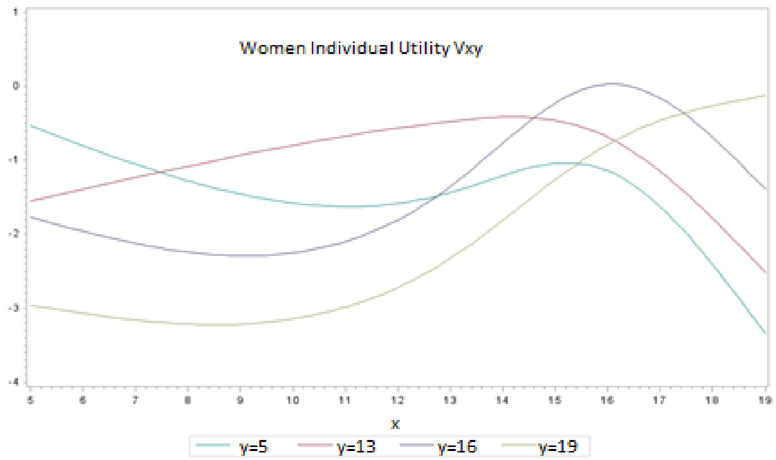


Figure 2: Women individual surplus in France in 1999

The estimation of the joint surplus shows that it is maximal in the subsection plan $x - y = 0$. Individuals match with people who have the same education level as them. The joint surplus is negative. This can be explained by the fact that there is less married people than singles. The following graph shows the curve of the joint surplus in function of the education of the two partners.

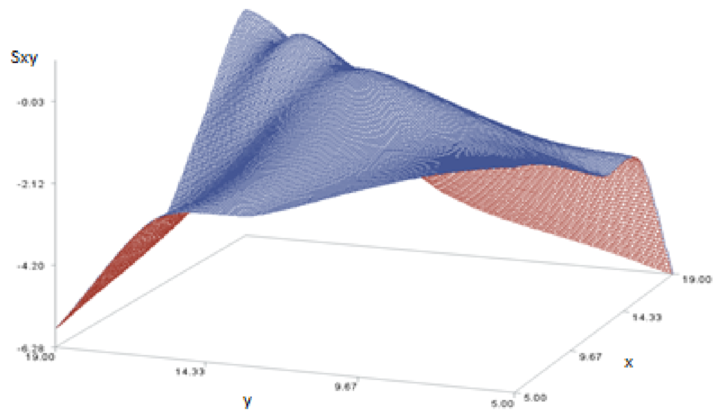


Figure 3: Joint surplus in France in 1999

Evolution of marriage surplus on the global french market

It can be interesting to look at the evolution of marriage surpluses over time in France. We compute the averaged values of U, V and S for each year from 1968 to 2011.

$$\bar{U} = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{xy}U_{xy} \text{ and } \bar{V} = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{xy}V_{xy}$$

and,

$$\bar{S} = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{xy}S_{xy} \text{ with } p_{xy} = \frac{\mu_{xy}}{\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\mu_{xy}}$$

The following graph represents the evolution of the marriage surplus of men, of women and the joint surplus in France from 1962 to 2011. We have also represented the surplus gap between men and women.

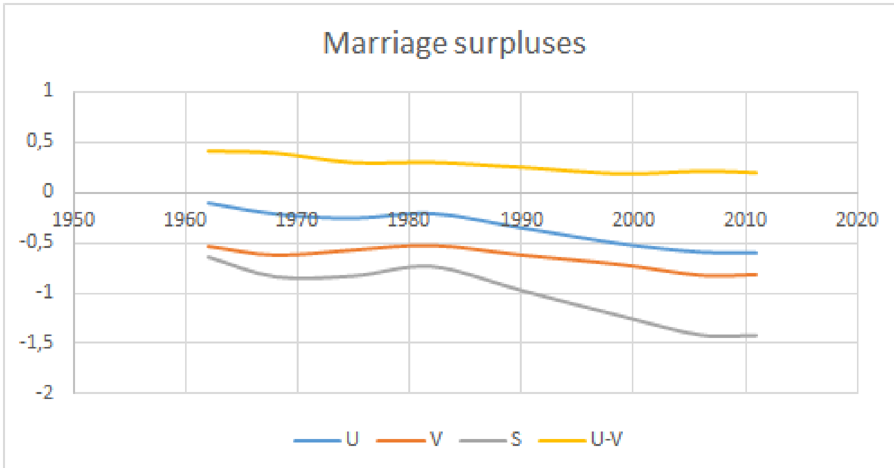


Figure 4: Global evolution of marriage surpluses

The marriage surplus has been relatively stable on the period 1962-1968 with an increase on 1968-1982. Then it has decreased from 1982 to 2011. We can conclude that there has been a negative shock on the marriage surplus in 1982. Marriage in terms of cohabitation brings less surplus today than by the past. The drop of the men surplus is higher than the drop of the women surplus. We also notice the reduction of the surplus gap between men and women. Men take the greater part of the joint surplus but this difference have significantly decreased and the marriage tends to be more egalitarian now. The shock of marriage surplus has impacted more men then women.

Regional french markets

In this subsection, we focus on the regional marriage markets. Regions are considered as local local marriage markets. The different french regions we have studied are: Ile -de-France, Champagne, Picardie, Haute-Normandie, Centre, Basse Normandie, Bourgogne, Nord-Pas-Calais, Lorraine, Alsace, Franche Comté, Pays de la Loire, Bretagne, Poitou Charentes, Aquitaine, Midi-Pyrénées, Limousin, Rhône-Alpes, Auvergne, Languedoc-Roussillon, Provence Côte d'Azur, Corse. The estimation of the marriage surplus in these regions reveals that in term of evolution, the regional marriage surpluses for all the regions have had the same dynamic of evolution everywhere as the general dynamic marriage surplus has had on the global french market: that is , an increase on the period 1968-1982 and a decrease on the period 1982-2011. On the period 1962-1968 regional surpluses have decreased. Today the regional marriage surpluses are generally negative. In 1999, only Pays de la Loire (0.26), Haute Normandie (0.15), Basse Normandie (0.11), Poitou Charentes (0.11), Nord Pas de Calais (0.07), Bretagne (0.05) and Picardie (0.048) still had positive marriage surplus. The lowest marriage surpluses in 1999 were in Ile-deFrance (-0.68), Languedoc Roussillon (-0.39), Provence Cte d'Azur (-0.53) and Corse (-0.75).For the other regions, see annex C.

Regional marriage surpluses have known two significant shocks: a positive shock in 1968 and a negative shock in 1982. These shocks have impacted all the regions but not with the same intensity. During the period 1962-1968, the highest impact of the positive shock is observed in Pays de la Loire and Aquitaine. In fact on that period, the marriage surplus has in increased in these regions by 0.4 points. The lowest impacts are observed in Basse Normandie, Alsace and Lorraine with an increase of the marriage surplus by 0.2 points. Particular fact is the decrease of the marriage surplus in this period in Languedoc-Roussillon and Provence Cte d'Azur; they are the only regions where the positive shock of marriage has not been observed.

The negative shock occurred in 1982 and its impact is observed everywhere. The highest losses of surplus are observed in Haute Normandie (-0.6), Ile-de-France (-0.61), Poitou Charentes (-0.61), Picardie (-0.66), Nord Pas de Calais (-0.71) and Languedoc Roussillon (-0.71);

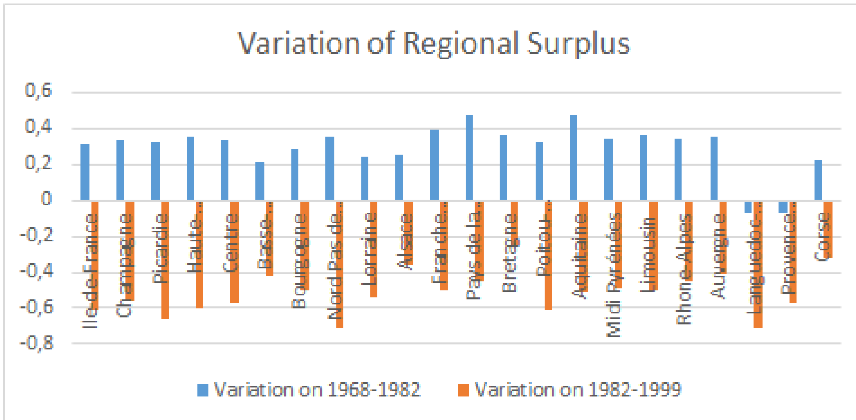


Figure 5: Variation of Regional Surplus

2.5.2 Education and Endogamy

Effect of the education on the marriage surplus

We evaluate the averaged impact of x on U, V and S by α_U, α_V, α_S on the population of married people. The averaged impact of y on U, V and S is evaluated by β_U, β_V and β_S on the population of married people. We measure these effects and we analyze their evolution over the time.

$$\alpha_U = E_{X,Y}\frac{\Delta U_{XY}}{\Delta X},\ \alpha_V = E_{X,Y}\frac{\Delta V_{XY}}{\Delta X},\ \alpha_S = E_{X,Y}\frac{\Delta S_{XY}}{\Delta X}$$

We can remark that:

$$\alpha_S = \alpha_U + \alpha_V$$

We have similar expressions for β.

$$\beta_U = E_{X,Y}\frac{U_{XY}}{\Delta Y},\ \beta_V = E_{X,Y}\frac{\Delta V_{XY}}{\Delta Y},\ \beta_S = E_{X,Y}\frac{\Delta S_{XY}}{\Delta Y}$$

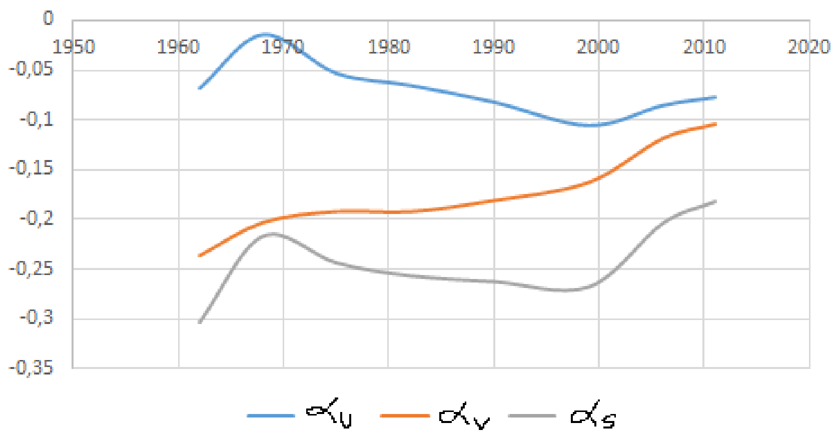


Figure 6: Impact of men education on marriage surplus

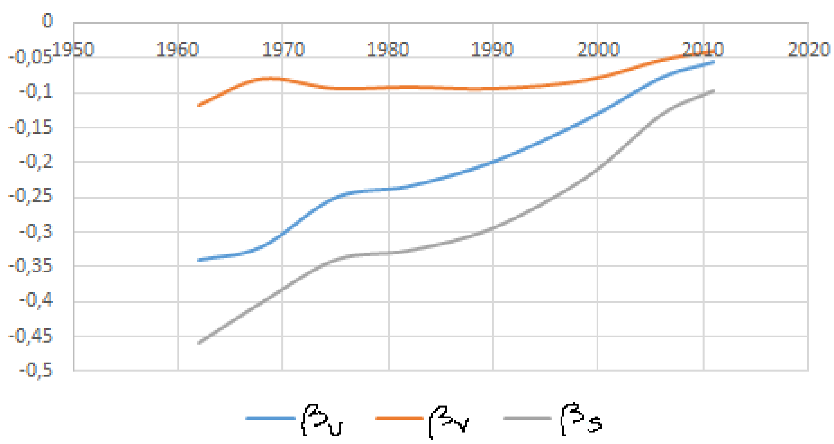


Figure 7: Impact of women education on marriage surplus

The averaged effect of the level of education of men and women on the marriage surpluses is negative. This can be explained by the fact the optimal choices of the partners on the subsection plane $x = y$; the education levels of the partners are very close. Any deviation from the subsection line intersubsection of the surplus curve with the plane of equation $(x = y)$ of the surface of the surplus has a negative impact on the surplus. The sign of the impact of x and y is logical. The intensity of the impact of the men's education is higher on the women' surplus than on the men's surplus. The intensity of the impact of the men's education on the women's surplus has decreased over time where as it has increased on the men's surplus. The intensity of the impact of the men's education on the joint surplus has globally decreased. In others words, considering two stable couples of type (x, x) and $(x + \Delta x, x)$, the joint surplus of (x, x) is greater than the joint surplus of $(x + \Delta x, x)$, but this difference in surplus is weaker today than by the past. For women, things don't work necessarily same as men. The intensity of the impact of women' education on the joint surplus and on the men's surplus has significantly decreased over time where as this intensity has relatively been stable on their own surplus. The impact of women's education on the marriage surplus is lower today than by the past.

Endogamy

The endogamy is characterized by the assortative mating. We estimate it by the cross derivative of the joint surplus. In fact, as shown by Becker (1974), the positivity (resp. negativity) of the cross derivative of the joint surplus indicates that a positive (resp. negative) correlation between the characteristics of the partners is optimal.

$$s_{x,y} := \frac{\Delta^2 S_{x,y}}{\Delta x \Delta y} = ([S_{x+1,y+1} - S_{x+1,y}] - [S_{x,y+1} - S_{x,y}]) * \frac{1}{\Delta x * \Delta y}$$

If $s_{xy} > 0$ then there is a positive assortative mating and negative if $s_{xy} < 0$. We then compute the averaged value of s.

$$E_{X,Y} \frac{\Delta^2 S_{x,y}}{\Delta x \Delta y} = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \left[\left(\frac{\Delta^2 S_{x,y}}{\Delta x \Delta y} \right) * \frac{\mu_{xy}}{\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \mu_{xy}} \right]$$

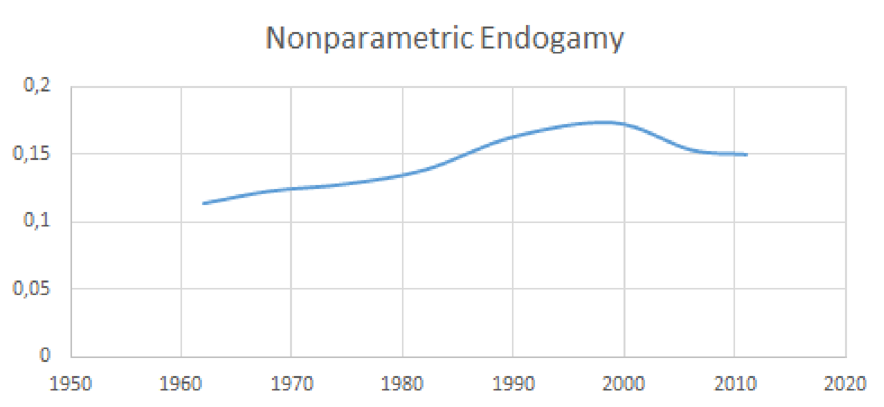


Figure 8: Evolution of Endogamy (Non-parametric)

We can remark that in France, the endogamy is positive in France and has increased on the period 1962-1999 and then has decreased from 1999 to 2011. We have tried to analyze this evolution depending on the different geographic french regions, on the period 1962-1999.

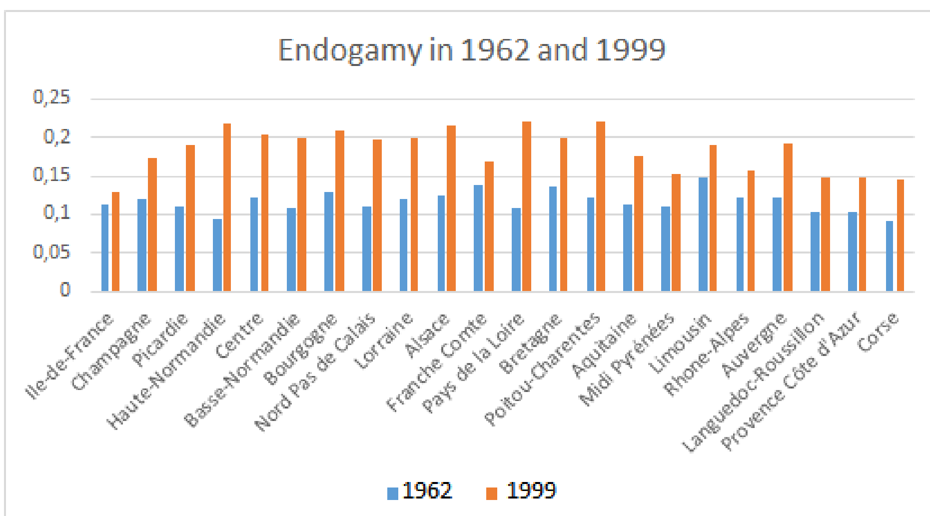


Figure 9: Regional Endogamy in 1962 and in 1999

The parisian region has the lowest level of endogamy and the lowest progression rate on the period 1962-1999. The endogamy has more increased in the North West of France than in the South. Pays de la Loire, Basse Normandie and Haute Normandie are the regions where endogamy has more increased on the period 1962-1999.

2.5.3 Parametric estimation

We complete the nonparametric analysis by a parametric analysis to prove the consistency of our results. To do this, we specify the joint surplus as follows:

$$S_{xy} = \lambda x^{\alpha} y^{\beta} + \sum_{k=1}^{6} \rho_k (x-y)^k$$

where x and y represent respectively the numbers of education years of the men, and of the women. For partners who have the same degree, their joint surplus is explained only by their their numbers of education years x and y.

The parameter λ is simply the joint surplus of a couple of type $(x=1, y=1)$.

$$\lambda = S_{11}$$

The parameters α and β represent respectively the elasticity of the joint surplus S_{xy} with respect to x and the elasticity of the joint surplus S_{xy} with respect to y, in the set of couples such that $x = y$, where the partners have the same level of education. In fact, if $x = y$, then $S_{xy} = \lambda x^{\alpha} y^{\beta}$ and

$$\frac{xdS}{Sdx} = \alpha, \text{ and } \frac{ydS}{Sdy} = \beta$$

This is equivalent to

$$\frac{dS}{Sdx} = \frac{\alpha}{x}, \text{ and } \frac{dS}{Sdy} = \frac{\beta}{y}$$

So when we consider a couple of type $x = y$, an increase by 1 in x makes vary the joint surplus at a rate of $sgn(\lambda)\frac{\alpha}{x}$ and similarly, an increase by 1 in y makes vary the joint surplus at a rate of $sgn(\lambda)\frac{\alpha}{y}$. The effect of x and y are inversely proportional to the level of education of the partner. So the intensity of the impact of the education on the joint surplus is higher for couples less educated than for very educated couples. But this effect depends on the sign of the product $\lambda\alpha$ for x and $\lambda\beta$ for y.

An increase in x and y of a couple of type $x = y$ from (x, x) to $(x + 1, x + 1)$ impacts the joint surplus at a rate of $sgn(\lambda)\frac{\alpha+\beta}{x}$.

2.5.4 Education and Endogamy

Impact of the education on the marriage surplus The parameter λ is negative and has significantly decreased over time, so does S_{11}. This is coherent to the general decrease of the joint surplus observed in the nonparametric estimation. The parameters α has been positive on the period 1962-1975 whereas β has been negative. The fact that λ is negative induces that the conditional effect of men education on the joint surplus with respect to $x = y$ has been negative on the period 1962-1975 and the conditional effect of women education on the joint surplus with respect to $x = y$ has been positive. These two parameters change suddenly sign in 1982 and λ also so that the effects have preserved their sign. The intensity of these effects has decreased nearly to zero on the period 1982-2011. They are both positive today. The following tables show the estimated values of α, β, λ and the estimated conditional effects of x and y on the joint surplus with respect to $x = y$, and their evolution over time.

Year	λ	α	β	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
1962	-0.0018 (6.35×10^{-7})	2.9724 (2.47×10^{-3})	-0.6030 (3.9×10^{-3})	1.394 (1.19×10^{-2})	-1.7184 (2.59×10^{-2})	-0.0597 (9.57×10^{-3})	0.281 (1.09×10^{-2})	0.006 (9.73×10^{-4})	- 0.0164 (6.34×10^{-4})
1968	-0.0038 (5.09×10^{-5})	2.768 (9.97×10^{-3})	-0.9435 (9.35×10^{-3})	1.126 (7.36×10^{-3})	-1.59 (8.21×10^{-3})	-0.0496 (3.63×10^{-2})	0.25052 (2.74×10^{-3})	0.0039 (3.94×10^{-4})	-0.0148 (2.34×10^{-4})
1975	-0.0001 (1.05×10^{-7})	3.802 (6.48×10^{-3})	-0.8059 (1.04×10^{-2})	0.702 (5.14×10^{-3})	-1.743 (5.27×10^{-3})	-0.0271 (2.95×10^{-3})	0.316 (2.17×10^{-3})	0.0014 (3.08×10^{-4})	-0.0208 (1.82×10^{-4})
1982	0.0202 (2.48×10^{-4})	-2.567 (3.71×10^{-2})	3.068 (2.09×10^{-2})	0.641 (5.52×10^{-3})	-2.264 (5.90×10^{-3})	-0.0473 (2.80×10^{-3})	0.4381 (2.21×10^{-3})	0.0062 (2.87×10^{-4})	-0.0311 (1.84×10^{-4})
1990	-0.3371 (9.64×10^{-3})	-0.7 (2.19×10^{-2})	-0.4964 (1.19×10^{-2})	0.487 (1.14×10^{-2})	-1.0129 (5.10×10^{-3})	-0.0649 (9.80×10^{-3})	0.0421 (6.10×10^{-3})	0.00421 (3.39×10^{-3})	-0.033 (1.88×10^{-3})
1999	-0.7914 (5.0×10^{-3})	-0.1582 (2.8×10^{-3})	-0.2985 (5.97×10^{-3})	0.149 (9.80×10^{-3})	-2.371 (9.60×10^{-3})	-0.0285 (4.90×10^{-3})	0.537 (4.00×10^{-3})	0.00048 (4.51×10^{-4})	-0.0381 (3.32×10^{-4})
2006	-2.663 (3.09×10^{-3})	-0.3970 (1.05×10^{-3})	-0.2988 (1.20×10^{-3})	-0.0141 (2.29×10^{-3})	-2.446 (2.21×10^{-3})	0.0062 (9.92×10^{-4})	0.5574 (8.2710^{-3})	-0.0019 (9.64×10^{-5})	-0.0397 (6.83×10^{-5})
2011	-4.137 (2.52×10^{-3})	-0.5088 (1.25×10^{-3})	-0.3090 (1.79×10^{-3})	-0.102 (1.34×10^{-3})	-2.588 (1.71×10^{-3})	0.00958 (9.69×10^{-4})	0.600 (5.75×10^{-4})	-0.0022 (1.04×10^{-4})	-0.043 (4.85×10^{-5})

Figure 10: Estimation

The following figures give respectively $\frac{\Delta S}{S\Delta x}$ and $\frac{\Delta S}{S\Delta y}$ with $\Delta x = \Delta y = 1$ and $x = y$.

Years	$x = 5$	$x = 13$	$x = 16$	$x = 19$
1962	-0.59448	-0.22865	-0.18578	-0.15644
1968	-0.55378	-0.21299	-0.17306	-0.14573
1975	-0.76052	-0.29251	-0.23766	-0.20014
1982	-0.51358	-0.19753	-0.16049	-0.13515
1990	-0.01829	-0.00703	-0.00571	-0.00481
1999	0.03164	0.012169	0.009888	0.008326
2006	0.0794	0.030538	0.024813	0.020895
2011	0.10176	0.039138	0.0318	0.026779

Figure 11: Impact of men' education on the joint surplus

Years	$y = 5$	$y = 13$	$y = 16$	$y = 19$
1962	0.1206	0.046385	0.037688	0.031737
1968	0.1887	0.072577	0.058969	0.049658
1975	0.16118	0.061992	0.050369	0.042416
1982	0.61374	0.236054	0.191794	0.161511
1990	0.00288	0.001108	0.0009	0.000758
1999	0.0597	0.022962	0.018656	0.015711
2006	0.05976	0.022985	0.018675	0.015726
2011	0.0618	0.023769	0.019313	0.016263

Figure 12: Impact of women' education on the joint surplus

In the period 1962-1990, for a couple of type (x, x), an additional year of education of the man impact negatively the joint surplus. The consequent reduction can be very high, about 76% in 1975 for couples $(x = 5, y = 5)$, whereas it is about 20% for couples $(x = 19, y = 19)$. This effect of x turns positive by 1999 until 2011. Today, for couples of type $(x = 16, y = 16)$, an increase in x by 1 brings 3.18% of the initial surplus as benefit. The change in the sign of the impact of the men' education on the joint surplus, negative by the past and positive today, can be explained by the reduction of the gap in education between men and women. By the past men were too much more educated than women so that an increase in their education increases that gap and affects more negatively the woman than it affects positively the man. This leads to a negative global effect on the joint surplus. Nowadays, the women education is close to the men education and even sometimes surpasses the men education. And increase in the men education reduces then actually the gap between him and his partner and they all both gain in surplus from this increase. For women, the effect has always been positive. Its highest intensity has been obtained in 1982. In this year, for a couple of type $(x = 5, y = 5)$, an increase in y by 1 could bring 61% of the initial surplus as benefit. But today, the intensity is very few. In 2011, for a couple $(x = 16, y = 16)$, when y by 1, the joint surplus increases 1.9%. Today the conditional effects of x and y on the joint surplus are very few but the impact of x is slightly higher than the impact of y. As we have shown it, $sgn(\lambda)\frac{\alpha+\beta}{x}$ is the variation rate of the joint surplus from (x, x) to $(x+1, x+1)$.It is the joint effect of x and y. The following graph shows the evolution of $\alpha+\beta$.

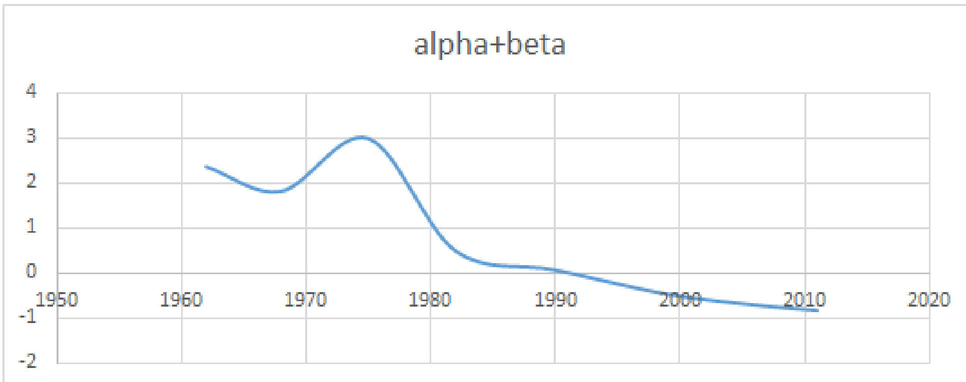


Figure 13: Evolution of $\alpha+\beta$

$\alpha+\beta$ is positive on the period 1962-1990, so $sgn(\lambda)\frac{\alpha+\beta}{x}$ is negative on this period. Less educated couples with same education level were happier then very educated couples. The conditional joint effect of x and y on the joint surplus turns positive by 1990 until 2011.

Endogamy We measure the endogamy that we denote by $E_{x,y}$ by the double cross derivative of $S_{x,y}$ with respect to x and y.

$$E_{x,y} = \frac{\partial^2 S_{x,y}}{\partial x \partial y} = \lambda\alpha\beta x^{\alpha-1} y^{\beta-1} - \sum_{k=2}^{6} \rho_k k(k-1)(x-y)^{k-2}$$

For simplification, we will consider couples of same education. The endogamy takes the form:

$$\tilde{E}_x = \lambda\alpha\beta x^{\alpha+\beta-2}.$$

The analysis shows that the endogamy has globally increased on the period 1962-1999 with a little depreciation of this dynamic in 1968, and a global decrease after 1999. The intensity of the evolution of the endogamy is evaluated by $|\frac{d\tilde{E}_x}{dt}|$. Its analysis reveals that, on the period 1962-1982, the intensity of the increase of endogamy has been higher for very educated couples than for less educated couples. But on the period 1982-1990, the reinforcement of the endogamy within less educated couples became higher than within very educated couples. And finally, the decrease of endogamy on the period 1999-2011 has been higher for less educated couples than for the very educated. The parametric estimation of the averaged endogamy shows the positivity of the assortative mating in France, its increase on the period 1962-1999 and its decrease on the period 1999-2011. This trend is exactly what reveals the nonparametric estimation. As represented in the following graph, the two estimations are very near all over the time.

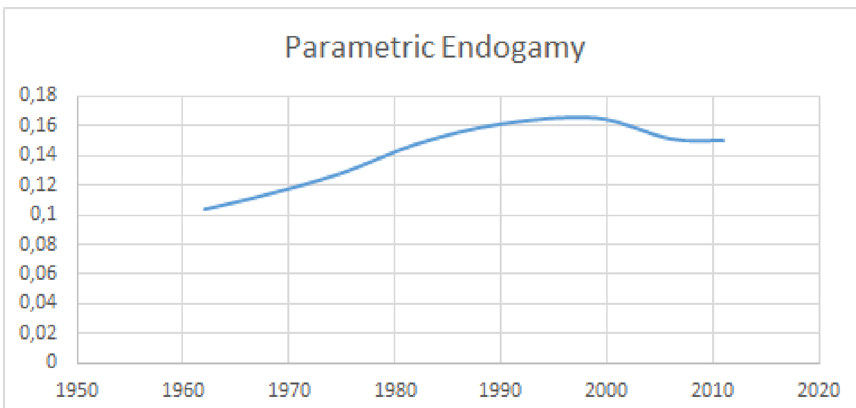


Figure 14: Evolution of the parametric endogamy

Our results are conform to the existence a trend with a form of an inverted U for the evolution of the educational homogamy in France, widely observed for The U.S and Canada. In that way we confirm the conclusion of studies of Schwart and Mare (2005) and Hou and Myles (2008) on the evolution of the educational homogamy in U.S and Canada in the French case. But these results are in contrast with the global decrease observed by sociological studies in France. In our work, the educational homogamy has decreased only on the period 1999-2011 after it has increased on the period 1962-1999.

Years	Endogamy
1962	0.1047 (3.66×10^{-4})
1968	0.1145 (4.11×10^{-4})
1975	0.1286 (3.89×10^{-4})
1982	0.1425 (4.77×10^{-4})
1990	0.161 (4.1×10^{-4})
1999	0.1621 (5.27×10^{-4})
2006	0.1463 (4.0×10^{-5})
2011	0.1497 (1.27×10^{-4})

Figure 15: Estimation of the averaged endogamy

Evolution of the joint surplus The evolution of the joint surplus is linked to the evolution of the different parameters on which it depends. To simplify the functional analysis, we will focus on couples in which partners have the same level of education. This is justified by the fact that the education levels of the partners are very close at the equilibrium and the education endogamy is positive in France. Let $\tilde{S}_x$ be the joint surplus of a couple of type (x, x) takes the form.

$$S_{x,x} = \tilde{S}_x(\lambda, \gamma) = \lambda x^{\gamma}$$

The parameters λ and γ depend on the time t.

$$\lambda = \lambda(t) \text{ and } \gamma = \gamma(t)$$

We have:

$$\frac{d\tilde{S}_x}{dt} = x^{\gamma}\left(\frac{d\lambda}{dt} + \lambda \ln(x)\frac{d\gamma}{dt}\right)$$

The parameter λ is generally negative and γ has decreased over time so: $\frac{d\gamma}{dt} \leq 0$. So we have:

$$\frac{d\tilde{S}_x}{dt} \geq 0 \Leftrightarrow x \geq \exp\left(-\frac{1}{\lambda}\frac{d\lambda}{d\gamma}\right)$$

We denote x_s by:

$$x_s = \exp\left(-\frac{1}{\lambda}\frac{d\lambda}{d\gamma}\right)$$

We have:

$$\frac{d\tilde{S}_x}{dt} \geq 0 \Leftrightarrow x_s \leq x$$

The data show that on the period 1962-1982, $x \geq x_s$. So the surplus has increased on the period 1962-1982.

$$\frac{d\tilde{S}_x}{dt} = x^\gamma \left(\frac{d\lambda}{dt} + \lambda \ln(x) \frac{d\gamma}{dt}\right)$$

The function $x \mapsto \frac{d\tilde{S}_x}{dt}$ is increasing in x and positive on the period 1962-1982, so $|\frac{d\tilde{S}_x}{dt}|$ is increasing in x. We deduce that the increase of the joint surplus has been higher for very educated couples then for less educated couples.

On the period 1982-2011, we have $x \leq x_s$. So the surplus has decreased on the period 1982-2011. The function $x \mapsto \frac{d\tilde{S}_x}{dt}$ is increasing in x and negative on the period 1982-2011, so $|\frac{d\tilde{S}_x}{dt}|$ is decreasing in x. We deduce that the decrease of the joint surplus has been higher for less educated couples then for very educated couples.

The following table shows the estimation of the averaged joint surplus and its evolution over time in France. The parametric estimation of the averaged joint surplus in France shows an increase of the joint surplus on the period 1962-1982 and a decrease on the period 1982-2011 as noticed with the nonparametric estimation. In the nonparametric estimation, on the period 1962-1982, the joint surplus seems to be relatively constant before it decreases then by 1982.

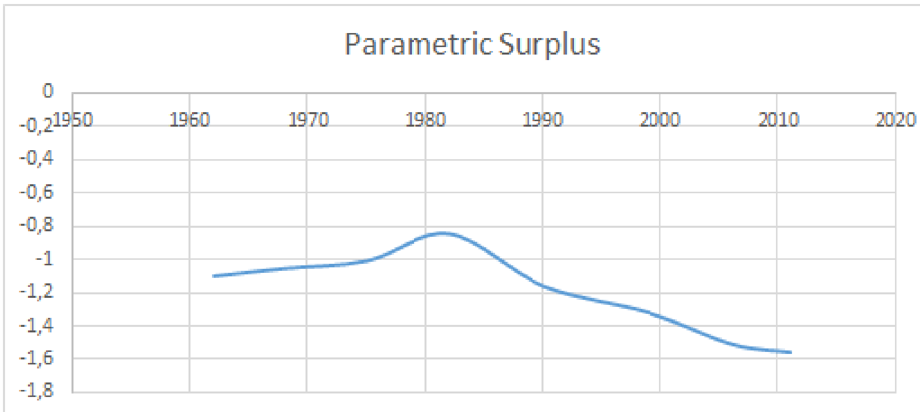


Figure 16: Evolution of the parametric joint surplus

The curve shows a significant decrease on the period 1982-2011.

Years	Averaged Joint Surplus
1962	-1.102 (1.8×10^{-3})
1968	-1.056 (2.5×10^{-3})
1975	-1.011 (1.95×10^{-3})
1982	-0.849 (2.33×10^{-3})
1990	-1.1625 (4.1×10^{-3})
1999	-1.322 (3.8×10^{-3})
2006	-1.51 (1.022×10^{-3})
2011	-1.557 (1.063×10^{-3})

Figure 17: Estimation of the averaged joint surplus

2.6 Conclusion

Marriage generates a marital joint surplus. This joint surplus is parted into two individual utilities between the spouses. These surpluses depend on the levels of education of the spouses and particularly on the distance between their levels. Less educated men maximize their utility with less educated women and reversely, and high educated men maximize their utility with high educated women. This mechanism of partner selection is observed on the French Marriage Market all over the period of our study. The graphic representation of the joint surplus generated by marriage is a generally a concave surface with local maxima located on the subsection $x = y$. The nonparametric estimation shows that the joint surplus has decreased from 1962 to 1968, has increased on the period 1968-1982 and then has decreased on the period 1982-2011. This dynamic is also observed on the individual surpluses of the spouses. On the period 1962-1968, men utility and women utility have both decreased. On the period 1968-1982, men utility remains stable whereas women utility has increased. This is why the joint surplus has increased on this period. From 1982 to 2011 men utility and women utility have both decreased, with an intensity of decrease greater for men than for women. So men have lost more utility from marriage than women. This difference has had an important consequence mainly on the surplus gap between men and women. Estimations revealed that the sharing of marriage surplus is not egalitarian; men have the greater part of the joint surplus. But this difference has been reduced overtime. Today men remain happier in their union than women but this gap is very few now. We investigate on how the level of education of the spouses impact their surpluses. Nonparametric estimations shows that the averaged effect of the education level of the spouses on their surpluses (joint and individual) are negative on the population. The intensity of the impact of men education on their individual surplus has increased on the period 1968-1999 and decreased on the period 1999-2011. The intensity of the impact of men education on the women surplus has globally decreased over time, so today men education has less impact on women surplus than by the past. On the joint surplus the impact of men education has been relatively stable on the period 1968-1999 and then it decreased in intensity from 1999 to 2011. The intensity of the impact of men education on their own surplus is weaker than on women surplus. With women things work a bit differently.

We remarked that the impact of women education on their own surplus, is negative but has remained relatively constant on the period 1968-1999 and then has decreased in intensity. The impact of women education on men surplus and on joint surplus has significantly decreased in intensity over time. The intensity of the impact of women education on their own surplus is weaker than on men surplus. Finally the nonparametric estimation of the endogamy reveals that the endogamy has been positive all over the time in France. This is equivalent with a positive assortative mating. People do not choose partners who are different from them in term of level of education. The nearer the partner is the higher is the attractiveness between the two. The analysis of the evolution of the endogamy shows a significant increase on the period 1962-1999 and a decrease on the period 1999-2011. It seems to be stable now. This is the general evolution of endogamy in France. Depending on the geographic regions, the level of endogamy vary. During the period 1962-1999, regions where we notice the greatest increase are the Center and the North-West except Bretagne. The parisian area has had the weakest increase on that period.

We refine the analysis with a parametric estimation. We have specified the joint surplus as the sum of a Cobb-Douglas function in x and y and a polynomial function in $(x - y)$. The model fits perfectly the data. The parametric joint surplus has increased on the period 1962-1982 and has decreased significantly on the period 1982-2011. More precisely, the intensity of increase of the joint surplus on the period 1962-1982 has been higher for very educated couples than for the less educated; and the decrease of the joint surplus on the period 1982-2011 has been higher for less educated couples than for the very educated couples. There has been a negative shock on the marriage surplus in 1982 and this shock has more impacted the less educated couples than the very educated couples. When we focus on couples whose spouses have same education level, we derive the conditional effect of x and y on the joint surplus with respect to $x = y$. Men education level has had a negative conditional effect on the joint surplus with respect to $x = y$ on the period 1962-1990. This effect turns positive after 1990 untill 2011. Women education level has had a positive conditional effect on the surplus with respect to $x = y$ all over time. But the intensity of the effect of men education level on the joint surplus is higher than women effect today. The positivity of these two effects has a consequence: very educated couples where spouses have same education level are happier today than less educated couples where spouses have same education level. Finally the parametric of the endogamy shows a significant increase of the endogamy on the period 1962-1999 and a decrease on the period 1999-2011. On the period 1962-1982, the intensity of the increase of endogamy has been higher for very educated couples than for less educated couples. But on the period 1982-1990, the reinforcement of the endogamy within less educated couples became higher than within very educated couples. And finally, the decrease of endogamy on the period 1999-2011 has been higher for less educated couples than for the very educated.

References

[1] Becker, G. 1973, A Theory of Marriage: Part I, The Journal of Political Economy, Vol. 81, No. 4, pp. 813-846

[2] Becker, G. 1974, A Theory of Marriage: Part II, The Journal of Political Economy, vol. 82, issue 2, S11-S26

[3] Bernardi. 2003, Who Marries Whom in Italy ?, H.-P. BLOSSFELD, A. TIMM (eds.), Who Marries Whom ? Educational Systems as Marriage Markets in Modern Societies, Dordrecht, Kluwer Academic Publishers, p. 113-139.

[4] Birkelund G. E., Heldal J., 2003, Who Marries Whom ? Educational Homogamy in Norway, Demographic Research, 8, 1, p. 1-30.

[5] Bouchet-Valat. 2011, Mémoire de sociologie, Master 2 mention Sociologie et Statistique, École des Hautes Etudes en Sciences Sociales

[6] Bouchet-Valat. 2014, Les évolutions de l'homogamie de diplôme, de classe et d'origine sociales en France (1969-2011) : ouverture d'ensemble, repli des élites, École des Hautes Études en Sciences Sociales

[7] Bozon, Michel & Francois Heran. 2006, La formation du couple. Textes essentiels pour la sociologie de la famille, La Découverte, coll. Grands Repères

[8] Chiappori, Pierre André, Bernard Salanié & Yoram Weiss. 2014, Partner Choice and the Marital College Premium: Analyzing Marital Patterns Over Several Decades, Columbia University

[9] Choo, Eugene & Aloysius Siow. 2006, Who marries whom and why, Journal of Political Economy, Vol. 114, No. 1, pp. 175-201

[10] Esteve A. & Cortina C., 2006, Changes in Educational Aassortative Mating in Contemporary Spain, Demographic Research, 14, 17, p. 405-428.

[11] Forsé & Chauvel. 1995, L'évolution de l'homogamie en France. Une méthode pour comparer les diagonalités de plusieurs tables, Revue Francaise de Sociologie

[12] Galichon & Salanié. 2012, Cupid's invisible hand: social surplus and identification in matching models, Institut d'études politiques

[13] Girard. 1964, Le choix du conjoint. Une enquête psycho-sociologique en France, Présentation d'un cahier de l'I.N.E.D

[14] Goux & Morin. 2003, Who marries whom in France ? An analysis of the cohorts born between 1934 and 1978, Oxford University Press

[15] Halpin B. & Chan T. W., 2003, Educational Homogamy in Ireland and Britain : Trends and Patterns, The British Journal of Sociology, 54, 4, p. 473-495.

[16] Henz U. & Jonsson J.O., 2003, Who Marries Whom in Sweden ? dans H.-P. BLOSSFELD, A. TIMM (eds.), Who Marries Whom ? Educational Systems as Marriage Markets in Modern Societies, Dordrecht, Kluwer Academic Publishers, p. 235-266.

[17] Hou F. & Myles J. 2008, The Changing Role of Education in the Marriage Market : Assortative Marriage in Canada and the United States Since the 1970s, Canadian Journal of Sociology / Cahiers canadiens de sociologie, 33, 2, p. 337-366.

[18] MARE R. D. 1991, Five Decades of Educational Assortative Mating, American Sociological Review, 56, 1, p. 15-32.

[19] Schwartz & Mare. 2005, Trends in Educational Assortative Marriage from 1940 to 2003, Demography, Volume 42-Number 4, p. 621-646.

[20] Smits J., Ultee W. & Lammers J. 1998, Educational Homogamy in 65 Countries : An Explanation of Differences in Openness Using Country-Level Explanatory Variables, American Sociological Review, 63, 2, p. 264-285.

[21] Ultee W. & Luijkx R., 1990, Educational Heterogamy and Father-to-Son Occupational Mobility in 23 Industrial Nations : General Societal Openness or Compensatory Strategies of Reproduction ?, European Sociological Review, 6, 2, p. 125-149.

[22] Vanderschelden. 2006, Homogamie socioprofessionnelle et ressemblance en termes de niveau d'études : constat et évolution au fil des cohortes dunions, Economie Et Statistique, Numero 398-399

2.7 Appendix

2.7.1 Social Surplus and Entropy

In the light of Galichon and Salanié (2012), we proceed as follows. We define the global social surplus as the sum of the expected utilities of all groups of men and women. So we have:

$$\mathcal{W} = \sum_{x\in\mathcal{X}} M_x G_x(U_{x.}) + \sum_{y\in\mathcal{Y}} N_y H_y(V_{.y})$$

where:

$$G_x(U_{x.}) = E(\max(\max_{y\in\mathcal{Y}}(U_{xy} + \varepsilon_y),\, \varepsilon^0)) \text{ and } H_y(V_{.y}) = E(\max(\max_{x\in\mathcal{X}}(V_{xy} + \eta_x),\, \eta^0))$$

$$G_x(U_{x.}) = \sum_{y\in\mathcal{Y}} \frac{\mu_{xy}}{M_x} U_{xy} - G_x^*(\mu_{.|x}) \text{ and } H_y(V_{y.}) = \sum_{x\in\mathcal{X}} \frac{\mu_{xy}}{N_y} V_{xy} - H_y^*(\mu_{.|y})$$

So the social surplus becomes:

$$\mathcal{W} = \sum_{x\in\mathcal{X}} \mu_{xy} U_{xy} + \sum_{y\in\mathcal{Y}} \mu_{xy} V_{xy} - \sum_{x\in\mathcal{X}} M_x G_x^*(\mu_{.|x}) - \sum_{y\in\mathcal{Y}} N_y H_y^*(\mu_{.|y})$$

i.e

$$\mathcal{W} = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mu_{xy} U_{xy} + \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} \mu_{xy} V_{xy} - \mathcal{E}(\mu)$$

where:

$$\mathcal{E}(\mu) = \sum_{x\in\mathcal{X}} M_x G_x^*(\mu_{.|x}) + \sum_{y\in\mathcal{Y}} N_y H_y^*(\mu_{.|y})$$

$$S = U + V$$

Then:

$$\mathcal{W} = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mu_{xy} S_{xy} - \mathcal{E}(\mu)$$

$\mathcal{E}(\mu)$ is the social entropy. It measures in some way the disorder among the population due to the idiosyncratic terms. If the variation of the heterogeneity is very few, the entropy becomes negligible and then the social surplus is very close to $\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \mu_{xy} S_{xy}$.

Expression of the social surplus $\mathcal{W}$ and the entropy $\mathcal{E}$

We can derive the expression of $\mathcal{W}$ and $\mathcal{E}$ in this model.

$$\mathcal{W} = \sum_{x\in\mathcal{X}} M_x G_x(U_{x.}) + \sum_{y\in\mathcal{Y}} N_y H_y(V_{.y})$$

$$G_x(U_{x.}) = \ln(1 + \sum_{y\in\mathcal{Y}} e^{U_{xy}}) \text{ and } H_y(V_{.y}) = \ln(1 + \sum_{x\in\mathcal{X}} e^{V_{xy}})$$

We also have:

$$e^{U_{xy}} = \frac{\mu_{xy}}{\mu_{x0}} \text{ and } e^{V_{xy}} = \frac{\mu_{xy}}{\mu_{0y}}$$

Hence

$$\begin{aligned}
\mathcal{W} &= \sum_{x\in\mathcal{X}} M_x \ln(1 + \sum_{y\in\mathcal{Y}} \frac{\mu_{xy}}{\mu_{x0}}) + \sum_{y\in\mathcal{Y}} N_y \ln(1 + \sum_{x\in\mathcal{X}} \frac{\mu_{xy}}{\mu_{0y}}) \\
&= \sum_{x\in\mathcal{X}} M_x \ln(\sum_{y\in\mathcal{Y}_0} \frac{\mu_{xy}}{\mu_{x0}}) + \sum_{y\in\mathcal{Y}} N_y \ln(\sum_{x\in\mathcal{X}_0} \frac{\mu_{xy}}{\mu_{0y}}) \\
&= \sum_{x\in\mathcal{X}} M_x \ln(\frac{M_x}{\mu_{x0}}) + \sum_{y\in\mathcal{Y}} N_y \ln(\frac{N_y}{\mu_{0y}}) \\
\mathcal{W} &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_0} \mu_{xy} \ln(\frac{M_x}{\mu_{x0}}) + \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}_0} \mu_{xy} \ln(\frac{N_y}{\mu_{0y}})
\end{aligned}$$

We have:

$$\mathcal{E} = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \mu_{xy} S_{xy} - \mathcal{W}$$

so,

$$\mathcal{E} = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \mu_{xy} S_{xy} - \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_0} \mu_{xy} \ln(\frac{M_x}{\mu_{x0}}) - \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}_0} \mu_{0y} \ln(\frac{N_y}{\mu_{0y}})$$

$$S_{xy} = \ln(\frac{\mu_{xy}^2}{\mu_{x0}\mu_{0y}})$$

hence, $\mathcal{E} = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \mu_{xy} \left(\ln(\frac{\mu_{xy}^2}{\mu_{x0}\mu_{0y}})\right) - \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_0} \mu_{xy} \ln(\frac{M_x}{\mu_{x0}}) - \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}_0} \mu_{0y} \ln(\frac{N_y}{\mu_{0y}})$
$= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \mu_{xy} \left(\ln(\frac{\mu_{xy}}{\mu_{x0}}) + \ln(\frac{\mu_{xy}}{\mu_{0y}})\right) - \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_0} \mu_{xy} \ln(\frac{M_x}{\mu_{x0}}) - \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}_0} \mu_{0y} \ln(\frac{N_y}{\mu_{0y}})$
$= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \mu_{xy} \left(\ln(\frac{\mu_{xy}}{\mu_{x0}})\right) + \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \mu_{xy} \left(\ln(\frac{\mu_{xy}}{\mu_{0y}})\right) - \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_0} \mu_{xy} \ln(\frac{M_x}{\mu_{x0}}) - \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}_0} \mu_{0y} \ln(\frac{N_y}{\mu_{0y}})$
$= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_0} \mu_{xy} \left(\ln(\frac{\mu_{xy}}{\mu_{x0}})\right) + \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}_0} \mu_{xy} \left(\ln(\frac{\mu_{xy}}{\mu_{0y}})\right) - \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_0} \mu_{xy} \ln(\frac{M_x}{\mu_{x0}}) - \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}_0} \mu_{0y} \ln(\frac{N_y}{\mu_{0y}})$

hence, we obtain

$$\mathcal{E} = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}_0} \mu_{xy} \ln\left(\frac{\mu_{xy}}{M_x}\right) + \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}_0} \mu_{xy} \ln\left(\frac{\mu_{xy}}{N_y}\right)$$

2.7.2 Semiparametric estimation

We consider in this subsection that utility is transferable. We suppose that the systematic joint surplus is divided into two parts: one nonparametric part and a linear parametric another one.

$$\ln\left(\frac{\mu_{xy}^2}{\mu_{x0} * \mu_{0y}}\right) = H(x, y) + T'\theta$$

$H(x, y)$ is a nonparametric unknown function on the observable characteristics and T is a subvector of (x, y). θ is a parameter to be estimated. The model can be rewritten as follows for any group g of marriage:

$$\Pi_g = T_g'\theta + r(Z_g) + e_g$$

where:

$$Z = (X, Y),\ \Pi_g = \ln(\frac{\mu_{x_g y_g}^2}{\mu_{x_g 0} * \mu_{0 y_g}})$$

r is an unknown function of $Z = (X, Y)$. Let define:

$$S_g^\theta = \Pi_g - T_g'\theta$$

The model becomes:

$$S_g^\theta = r_\theta(Z_g) + e_g$$

Conditionally to θ this model is semiparametric. So, by the Nadaraya-Waston estimator we have:

$$\hat{r_\theta}(Z) = \frac{\sum_{g=1}^G K(\frac{Z-Z_g}{h}) * S^\theta}{\sum_{g=1}^G K(\frac{Z-Z_g}{h})}$$

i.e

$$\hat{r_\theta}(Z) = \frac{\sum_{g=1}^G K(\frac{Z-Z_g}{h}) * (\Pi_g - T_g'\theta)}{\sum_{g=1}^G K(\frac{Z-Z_g}{h})}$$

G is the number of types of marriage.

$$G = |\mathcal{X}| * |\mathcal{Y}| = p * q$$

Then, an estimator for S conditionally to θ is:

$$\hat{S}_g^\theta = \hat{r_\theta}(Z_g)$$

Let's define:

$$\mathbf{H} = \left(\frac{K(\frac{Z_i - Z_j}{h})}{\sum_{g=1}^G K(\frac{Z_i - Z_g}{h})} \right)_{1 \leq i,\, j, \leq G}$$

and

$$\mathbf{S} = (S_g)_{1 \leq g \leq G},\, \Pi = (\Pi_g)_{1 \leq g \leq G}$$

and matrix $\mathbf{T}$ is such that:

$$L_g(\mathbf{T}) = T_g',\, 1 \leq g \leq G$$

We have:

$$\hat{\mathbf{S}}^\theta = \mathbf{H}(\Pi - \mathbf{T}\theta)$$

and

$$\hat{\Pi}^\theta = \hat{\mathbf{S}}^\theta + \mathbf{T}\theta$$

Now to estimate θ we minimize under θ the square of the norm of $\Pi - \hat{\Pi}^\theta$ so, we define:

$$\Sigma(\theta) = ||\Pi - \hat{\Pi}^\theta||^2$$

We have:

$$\begin{aligned}
\Sigma(\theta) &= ||\Pi - \mathbf{H}(\Pi - \mathbf{T}\theta) - \mathbf{T}\theta||^2 \\
&= ||(\mathbf{I} - \mathbf{H})(\Pi - \mathbf{T}\theta)||^2 \\
&= ||\mathbf{A}(\Pi - \mathbf{T}\theta)||^2 \\
\Sigma(\theta) &= (\mathbf{A}(\Pi - \mathbf{T}\theta))'(\mathbf{A}(\Pi - \mathbf{T}\theta))
\end{aligned}$$

where:

$$\mathbf{A} = \mathbf{I} - \mathbf{H}$$

$$\hat{\theta} = Arg \min_{\theta} \Sigma(\theta)$$

First order condition gives:

$$\frac{\partial \Sigma}{\partial \theta}(\hat{\theta}) = 0$$

$$\begin{aligned}
\frac{\partial \Sigma(\theta)}{\partial \theta} &= \frac{\partial(\Pi - \mathbf{T}\theta)}{\partial \theta} \frac{\partial \Sigma(\theta)}{\partial(\Pi - \mathbf{T}\theta)} \\
&= -\mathbf{T}' \frac{(\mathbf{A}(\Pi - \mathbf{T}\theta))'(\mathbf{A}(\Pi - \mathbf{T}\theta))}{\partial(\Pi - \mathbf{T}\theta)} \\
&= -\mathbf{T}'(2\mathbf{A}'\mathbf{A}(\Pi - \mathbf{T}\theta)) \\
\frac{\partial \Sigma(\theta)}{\partial \theta} &= -2\mathbf{T}'\mathbf{A}'\mathbf{A}(\Pi - \mathbf{T}\theta)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Sigma}{\partial \theta}(\hat{\theta}) = 0 &\iff \mathbf{T}'\mathbf{A}'\mathbf{A}(\Pi - \mathbf{T}\hat{\theta}) = 0 \\
&\iff \mathbf{T}'\mathbf{A}'\mathbf{A}\Pi = \mathbf{T}'\mathbf{A}'\mathbf{A}\mathbf{T}\hat{\theta} \\
&\iff \hat{\theta} = (\mathbf{T}'\mathbf{A}'\mathbf{A}\mathbf{T})^{-1}\mathbf{T}'\mathbf{A}'\mathbf{A}\Pi
\end{aligned}$$

We then estimate Π by:

$$\begin{aligned}
\hat{\Pi} &= \hat{S}^{\hat{\theta}} + \mathbf{T}\hat{\theta} \\
&= \mathbf{H}(\Pi - \mathbf{T}\hat{\theta}) + \mathbf{T}\hat{\theta} \\
&= \mathbf{H}\Pi + (\mathbf{I} - \mathbf{H})\mathbf{T}\hat{\theta} \\
&= \mathbf{H}\Pi + \mathbf{A}\mathbf{T}\hat{\theta} \\
&= \mathbf{H}\Pi + \mathbf{A}\mathbf{T}(\mathbf{T}'\mathbf{A}'\mathbf{A}\mathbf{T})^{-1}\mathbf{T}'\mathbf{A}'\mathbf{A}\Pi \\
&= (\mathbf{H} + (\mathbf{T}'\mathbf{A}'\mathbf{A}\mathbf{T})^{-1}\mathbf{T}'\mathbf{A}'\mathbf{A})\Pi \\
\hat{\Pi} &= \mathbf{Q}\Pi
\end{aligned}$$

where:

$$\mathbf{Q} = \mathbf{H} + (\mathbf{T}'\mathbf{A}'\mathbf{A}\mathbf{T})^{-1}\mathbf{T}'\mathbf{A}'\mathbf{A}$$

Quality of the estimator

We can evaluate the quality of the estimator by the mean averaged squared error MASE defined by:

$$MASE = E(||\hat{\Pi} - E(\Pi)||^2)$$

We suppose homoskedascity across marriage groups.

$$Var(\Pi) = \sigma^2\mathbf{I}$$

This assumption is clearly not necessary. We have:

$$\begin{aligned}
MASE &= E(||\hat{\Pi} - E(\Pi)||^2) \\
&= E(||\mathbf{Q}\Pi - E(\Pi)||^2) \\
&= ||(\mathbf{Q} - \mathbf{I})E(\Pi)||^2 + Tr(Var(\mathbf{Q}\Pi)) \\
MASE &= ||(\mathbf{Q} - \mathbf{I})E(\Pi)||^2 + \sigma^2 Tr(\mathbf{Q}\mathbf{Q}')
\end{aligned}$$

Remark: For any random vector X we have:

$$E(||X||^2) = Tr(Var(X)) + ||E(X)||^2$$

Choice of the smoothing parameter h

We simply take the smoothing parameter that minimizes the cross validation CV or the general cross validation GCV.

$$h = Arg\min_h GCV(h)$$

$$GCV = \frac{1}{G}\left(\frac{G}{G - Tr\mathbf{Q}}\right)^2 ||\Pi - \hat{\Pi}||^2$$

2.7.3 Computation

Iterative Projection Fitting Procedure (Galichon and Salanié, 2012)

Lets consider a particular vector θ. To determinate $\log L(\lambda)$ we compute μ^λ as follows. We define a generalized entropy E that can be applied to even non feasible matching. We adopt the following notations:

$$\bar{\mu} = (\mu, \, (\mu_{x0})_x, \, (\mu_{0y})_y)$$

The optimal μ^λ maximizes:

$$W = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \mu_{xy} \Phi_{xy}^\lambda - E(\bar{\mu})$$

under the constraints:

$$\forall x, y, \sum_{y \in \mathcal{Y}} \mu_{xy} + \mu_{x0} = M_x, \text{ and } \sum_{x \in \mathcal{X}} \mu_{xy} + \mu_{0y} = N_y$$

The Lagrangian associated to this problem is:

$$\mathcal{L}(\bar{\mu}, \, a, \, b) = \sum_{x \in \mathcal{X}} a_x \left(M_x - \sum_{y \in \mathcal{Y}_0} \mu_{xy} \right) + \sum_{y \in \mathcal{Y}} b_y \left(N_y - \sum_{x \in \mathcal{X}_0} \mu_{xy} \right) + \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \mu_{xy} \Phi_{xy}^\lambda - E(\bar{\mu})$$

where:

$$a = (a_x)_x \text{ and } b = (b_y)_y$$

a and b are the Lagrangian multipliers. First conditions order give:

$$\frac{\partial E}{\partial \mu_{xy}} = \Phi_{xy} - a_x - b_y;\ \frac{\partial E}{\partial \mu_{x0}} = -a_x;\ \frac{\partial E}{\partial \mu_{0y}} = -b_y$$

We will iterate this procedure with the initial conditions at step $k = 0$, $(\bar{\mu}, a, b) = (\bar{\mu}^0, a^0, b^0)$. At any step $2k+1$, we suppose $b = b^{2k}$ and the Lagrangian becomes $\mathcal{L}(\bar{\mu}, a, b^{2k})$. First order conditions are:

$$\frac{\partial E}{\partial \mu_{xy}} = \Phi_{xy} - a_x - b_y^{2k};\ \frac{\partial E}{\partial \mu_{x0}} = -a_x;\ \frac{\partial E}{\partial \mu_{0y}} = -b_y^{2k}$$

Then we compute the solution $(\bar{\mu}^{2k+1}, a^{2k+1})$ with the equation:

$$M_x = \sum_{y \in \mathcal{Y}_0} \mu_{xy}$$

And at any step (2k+2), we assume $a = a^{2k+1}$ and the Lagrangian becomes $\mathcal{L}(\bar{\mu}, a^{2k+1}, b)$. First order conditions give:

$$\frac{\partial E}{\partial \mu_{xy}} = \Phi_{xy} - a_x^{2k+1} - b_y;\ \frac{\partial E}{\partial \mu_{x0}} = -a_x^{2k+1};\ \frac{\partial E}{\partial \mu_{0y}} = -b_y$$

Then we compute the solution $(\bar{\mu}^{2k+2}, b^{2k+2})$ with the equation:

$$N_y = \sum_{x \in \mathcal{X}_0} \mu_{xy}$$

We stop the procedure when the distance between $\bar{\mu}^{2k+2}$ and $\bar{\mu}^{2k}$is few enough.

Alternative Computation

The matching matrix is linked to the surplus matrix by the relation:

$$\ln(\frac{\mu_{xy}^2}{\mu_{x0} * \mu_{0y}}) = \Phi_{xy}^{\lambda}$$

At the equilibrium of the marriage market, μ_{xy} is a function of Φ_{xy}^{λ}, μ_{x0} and μ_{0y}. Let f be the link function. f can be known if the distribution of the utilities is known. In this framework the utilities follow a standard Gumbel law.

$$\mu_{xy} = f\left(\Phi^{\lambda}_{xy}, \mu_{x0}, \mu_{0y}\right)$$

We have the constraints:

$$\mu_{x0} = M_x - \sum_{y \in \mathcal{Y}} \mu_{xy} \quad \text{and} \quad mu_{0y} = N_y - \sum_{x \in \mathcal{X}} \mu_{xy}$$

By replacing μ_{xy} in the equations of the singles, we obtain:

$$\forall x \in \mathcal{X},\ \mu_{x0} = M_x - \sum_{y \in \mathcal{Y}} f\left(\Phi^{\lambda}_{xy}, \mu_{x0}, \mu_{0y}\right)$$

and

$$\forall y \in \mathcal{Y},\ \mu_{0y} = N_y - \sum_{x \in \mathcal{X}} f\left(\Phi^{\lambda}_{xy}, \mu_{x0}, \mu_{0y}\right)$$

We start the resolution by an arbitrary point:

$$\mu^0_{0y} = \hat{\mu}_{0y}$$

At step $(2k+1)$, we solve:

$$\forall x \in \mathcal{X},\ \mu_{x0} = M_x - \sum_{y \in \mathcal{Y}} f\left(\Phi^{\lambda}_{xy}, \mu_{x0}, \mu^{(2k)}_{0y}\right)$$

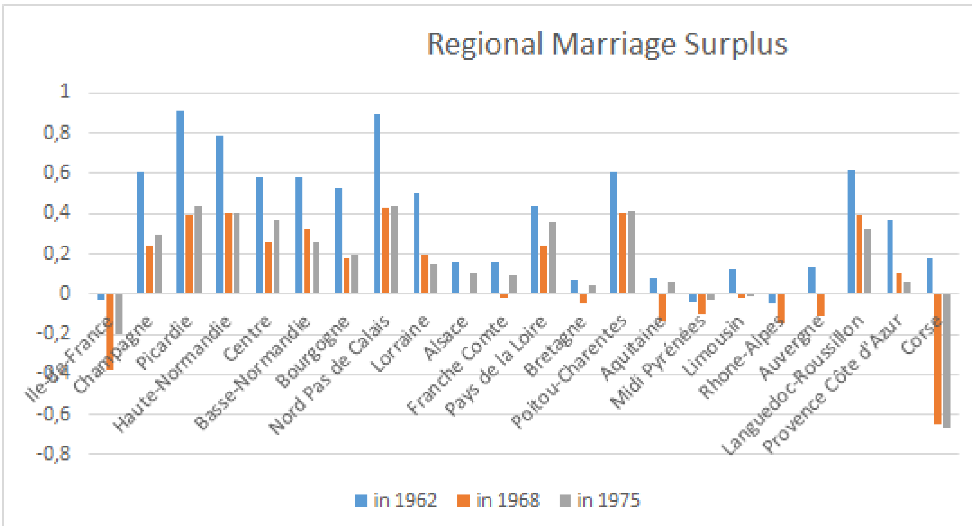
At step $(2k+2)$, we solve:

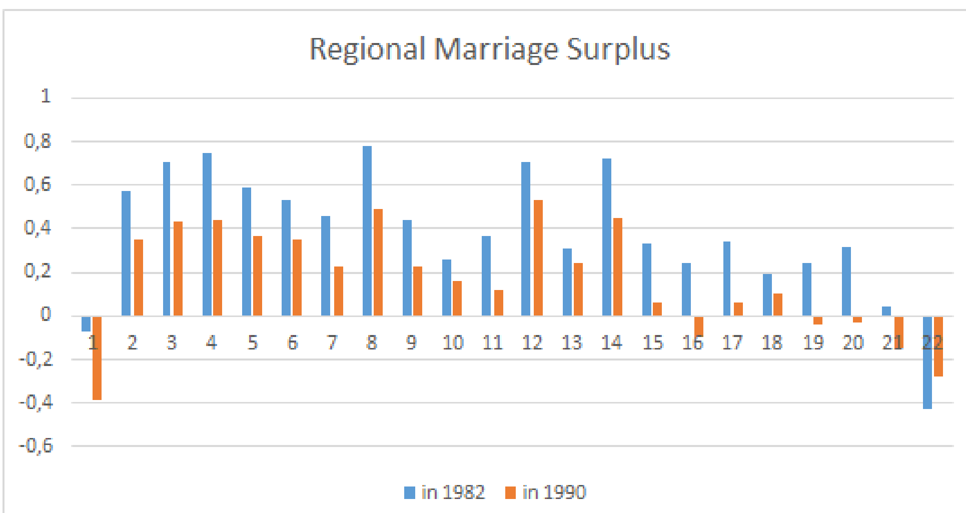
$$\forall y \in \mathcal{Y},\ \mu_{0y} = N_y - \sum_{x \in \mathcal{X}} f\left(\Phi^{\lambda}_{xy}, \mu^{(2k+1)}_{x0}, \mu_{0y}\right)$$

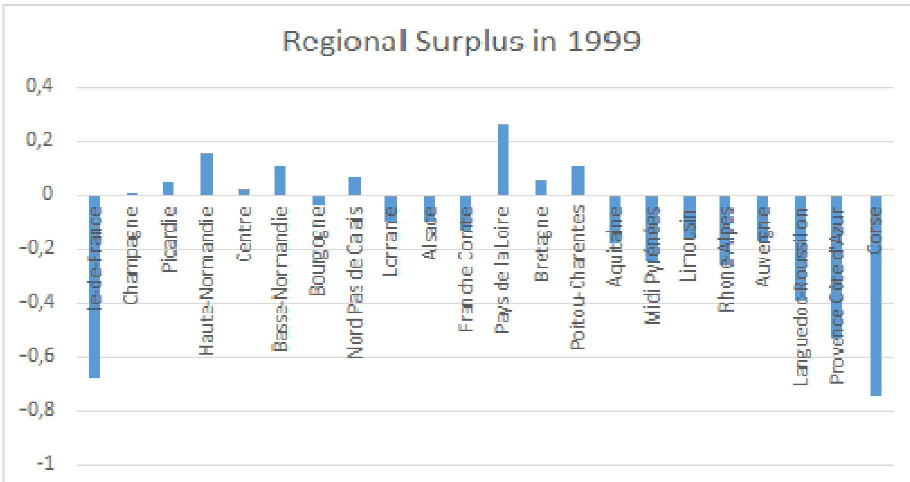
We stop the algorithm when $\left\|\mu^{(2k+2)} - \mu^{(2k)}\right\|^2$ is few enough. We then compute $LL(\lambda)$. We take the one that has the maximum likelihood.

2.7.4 Empirical results of the nonparametric estimation

Marriage Surpluses







Endogamy

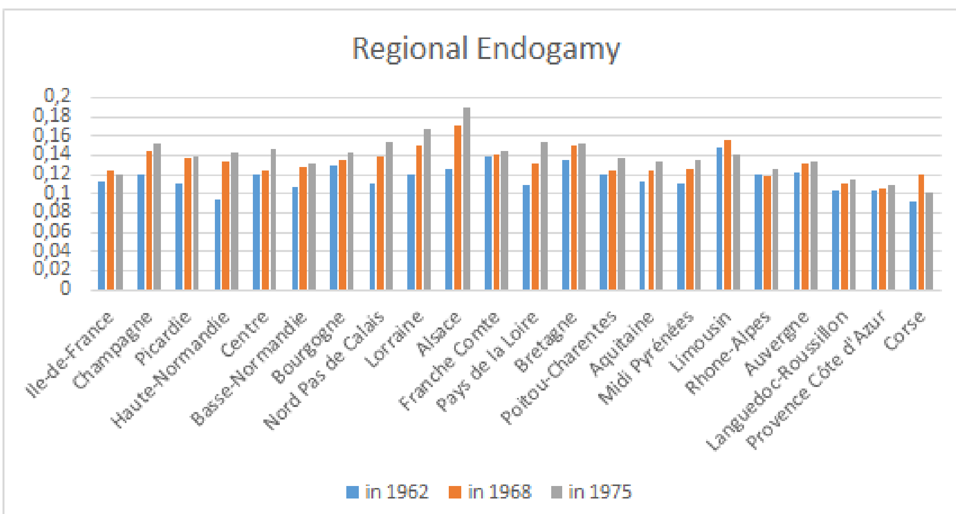


Figure 18: Regional Endogamy in 1962, 1968 and 1975

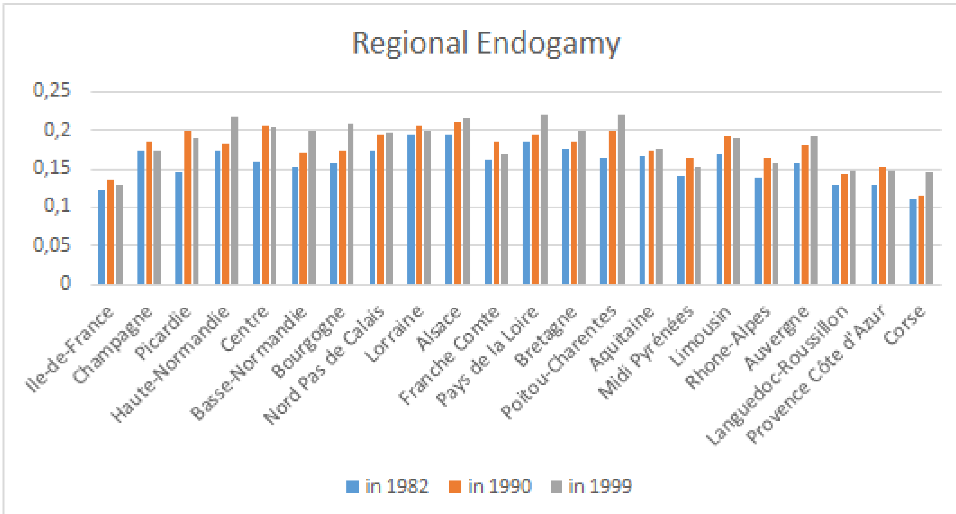


Figure 19: Regional Endogamy in 1982, 1990 and 1999

Figure 20: Evolution Rate of Regional Endogamy on the period 1962-1999

3 Marriage and Income Inequality: A New Model and some Evidence from U.S.

Abstract

In this paper we develop a continuous matching model with transferable utility. This model is a continuous extension of the Choo and Siow's (2006) model in the light of Dupuy and Galichon's (2014) that will differ from their approach in the way we will incorporate singles. One advantage of our approach is that it will allow an intuitive introduction of the attractiveness of the individuals in the matching market and it is empirically tractable. We then use this model to investigate the impact of marriage on income inequality in The U.S from 1968 to 2001 with PSID data.

3.1 Introduction

Trends in income inequality have been widely documented. Since the 1970s the growth rate of individual income in most OECD countries has become significantly lower in the bottom part of the income distribution than at the top. Over the same period, family structure and behavior experienced deep changes. Firstly, changes in demographic behavior have led to an increased likelihood of single families: decline in marriage rates and rise in divorce rates, etc. Secondly, gender inequality have narrowed significantly in both educational attainment and labor force participation, leading to a significant increase in the share of skilled employed women. Finally, in modern marriages husbands and wives tend to become more economically similar: couples are more often composed of two high- or two low-earning partners.

The socio-demographic changes observed during the last fifty years has, rather unsurprisingly, had consequences on the intra-household distribution of consumption, so that growth in income inequality actually hides more ambiguous trends in individual consumption. Lise and Seitz (2011) use structural methods to recover individual consumption from survey data and show that, in the U.K., inequalities in terms of individual consumption have increased more slowly than income inequalities. The shrinking wage gap among spouses explains a more equal distribution of total consumption within the household. Bargain et alii (2019) confirm this result with another identification strategy and stress that progress in education and in living standards are the key drivers of the increase in women' consumption.

Evaluating individual consumption is delicate, though, because a fraction of goods are jointly produced and jointly consumed by all the household members. The degree of joint consumption plays a particularly important role because it defines the size of scale economies in multi-person households (by comparison with single person households).[1] In particular, the technology of household consumption and production have changed over time: economies of scale associated with living in a couple – and more generally all the benefits of living in a couple – have not necessarily remained the same for 50 years.[2] Since the proportion of single persons and couples has changed significantly in recent years, and the costs and benefits of forming a couple have certainly not remained constant, it is difficult to correctly assess the observed increase in inequality.

In this paper, we thus adopt a more radical perspective. We assume that individual utility is perfectly transferable and, for single individuals, is equal to observable income, i.e., we use a particularly convenient cardinalization of utility. It means that utility, like money, can be redistributed between individuals at a constant rate. This formulation is convenient because it gives a way to compare utility of individuals living alone and living in a couple.[3] To identify individual utility, we then use an empirical marriage model à la Choo and Siow (2006) who start from the assumption that the surplus derived

[1] In applied work that attempt to evaluate inequality, scale economies are represented by ad-hoc equivalence scales that are largely arbitrary.

[2] Gottschalk and Mayer (2002) point out that the observed increase of the wages at the top of the distribution and the observed decline at the bottom of the distribution would, under most circumstances, increase home-production among low-income families and reduce home production among high-income households.

[3] It must be admitted, however, that transferability is a strong hypothesis that implies some restrictions on individual preferences. In particular, Bergstrom and Cornes (1983) derive the specific form that utility functions should have when consumption is made up of private and public goods.

from marriage breaks down into two parts: the first is deterministic and depends on a finite number of observable individual characteristics and the second consists of a continuous random term, with an extreme value distribution, representing what is not observable by the analyst. More precisely, in order to take into account the continuity of individual incomes, we consider the extension of this model proposed by Dupuy and Galichon (2014) who account for continuous observable individual characteristics (instead of finite number of characteristics). The second part of the surplus is then a max-stable stochastic process.

We also address another issue. It is generally admitted that much of the rise in overall inequality may be due to family composition shifts and other causes rather than the change in the sole pay patterns. Burtless (1999) shows that the growing positive correlation of husbands' and wives' earnings may account for 13% of the increase in overall inequality in the U.S. between 1979 and 1996 while the growing proportion of families headed by a single person may account for 21-25% of the increase. Similarly, Greenwood et alii (2014) find that, if mating patterns in the U.S. had remained the same as in 1960, the Gini coefficient would have been significantly lower in 2005.[4]

Our contribution is two-fold. In the initial formulation of the Dupuy and Galichon's (2014) marriage model, single individuals are ignored, i.e., all individuals are assumed to be married. Firstly, we propose an alternative version of this model to the case of presence of single individuals that is empirically tractable. We show that the individual utilities can be identified from observed marriages and propose a Maximum Likelihood methodology to estimate the parameters of the model. This extension is crucial because our main objective is to compare inequality among single individuals and married individuals. One must mention that Dupuy and Galichon have proposed an extension of their model to the integration of single individuals; the approach we propose here will precisely differ slightly from theirs in the way we will take into account the singles. The model is empirically tractable. One of the advantages of this approach that will be mainly shown in a later work in Chapter III, is that it will allow with particular specifications, analytical solutions. Secondly, we estimate this model using data from

[4]See also Martin (2006), Fernandez and alii (2004), Schwartz (2010) for similar results.

the PSID over the period 1968-2001. We show a reduction of income inequality by marriage and this inequality index has slightly increased on the period of the study. We have also measured the conversion rate between marriage surplus and income. We interpreted it as the 'monetary value of marriage'. It has increased on the period 1968-1971, decreased on the period 1971-1985 and has been relatively stable on the period 1985-2001.

3.2 The model

In this subsection, we consider the Dupuy-Galichon marriage market model and generalize it to the possibility of singles. We then show that individual surplus can be identified from observed matchings.

3.2.1 Marriage Market Description

Women $\mathcal{W}$ constitute one side of the marriage market and men $\mathcal{M}$ the other side. Individuals of each side decide either to match (i.e., marry) with an individual of the opposite side or to remain single. We assume the number of women and the number of men to be possibly unequal and infinite. The population of individuals is normalized to one, the number of men is denoted by ν and the number of women by $1-\nu$. Men and women are respectively characterized by vectors of attributes x and y (which may include continuous components, discrete components or the mix of the two). In our empirical application, we shall assume that attributes include individual incomes. We denote respectively by $\mathcal{X} \subseteq \mathbb{R}^p$ and $\mathcal{Y} \subseteq \mathbb{R}^q$ the support spaces of x and y. Let P and Q be respectively the probability distribution of x and y. In addition we assume these probability distributions to have given probability density functions. Let f and g be respectively the density function of x and the density function of y.

The marriage market aims to provide to each individual a partner or not by leaving him or her single. To characterize the matchings, we need to introduce some additional notions. We first define a variable M that gives at the equilibrium the indication of marriage for any individual. For any man m, if $M_m = 1$ then the man m is matched, and if $M_m = 0$ then he is single; and for any woman w, $M_w = 1$ if she is matched and $M_w = 0$ if she is single. We then define the probability α for a man to be married in the male population and, analogously, the probability β for a woman to be married. Since the number of married men and the number of married women have to be the same, we also have: $\alpha \cdot \nu = \beta \cdot (1-\nu)$.[5] In the case men and women are in equal proportions in the whole population, we get $\alpha = \beta$. We define f_1 and f_0 as the conditional probability density of x with respect to the event $M_m = 1$ and to the event $M_m = 0$. Analogously, we define respectively g_1 and g_0 the conditional probability density of y with respect to the event $M_w = 1$ and to the event $M_w = 0$. Since $f_1(x)$ and $f_0(x)$ are conditional probability density functions, we have:

$$f(x) = \alpha \cdot f_1(x) + (1-\alpha) \cdot f_0(x). \tag{1}$$

Similarly, we have:

$$g(y) = \beta \cdot g_1(y) + (1-\beta) \cdot g_0(y) \tag{2}$$

We can now define the notion of feasible matching as follows.

Definition 3.1 (Feasible Matching)**.** *A feasible matching is a joint density function $\pi(x,y)$ defined on $\mathcal{X} \times \mathcal{Y}$ and a positive $\kappa \in [0,1]$ representing the probability of marriage on the market.*

One must remark that:

$$\kappa = \nu\,\alpha = (1-\nu)\,\beta.$$

Then, as soon as one knows κ, one can derive the conditional probability of marriage on the set of men α and the conditional probability of marriage on the set of women β as the weights of the two sides of the market ν and $(1-\nu)$ are known.

A matching function $\pi(x,y)$ gives somehow the probability of occurrence of any particular type of marriage (x,y) in the population of married people. Since $f_1(x) =$

[5] In the setting of Dupuy and Galichon (2014), in which they consider a full marriage on the population, that assumption is equivalent to consider $\nu = \frac{1}{2}$, and $\alpha = \beta = 1$.

$\int_{\mathcal{Y}} \pi(x,y)dy$ and $g_1(y) = \int_{\mathcal{X}} \pi(x,y)dy$, the conditional probability density of X and Y with respect to the event $M = 0$ are given by:

$$\begin{aligned} f_0(x) &= \frac{1}{1-\alpha}\left(f(x) - \alpha \cdot \int_{\mathcal{Y}} \pi(x,y)dy\right) \qquad \text{and} \\ g_0(x) &= \frac{1}{1-\beta}\left(g(x) - \beta \cdot \int_{\mathcal{X}} \pi(x,y)dy\right), \end{aligned}$$

respectively. This model is a continuous extension of the discrete model proposed by Choo and Siow (2006) and an alternative approach to the Dupuy and Galichon's model by directly taking into account singles. Dupuy and Galichon (2014) have proposed a way to take into account singles and have shown that the integration of singles into their model does not fundamentally change the results they have obtained in the case of a full marriage on the population. Let denote by $\Pi(P,Q)$ the set of feasible matchings (π,κ). We denote by $\pi(x,y|M=1,\mathcal{H})$ the conditional probability density of the vector (X,Y) at the point (x,y) with respect to the event $\{M=1,\mathcal{H}\}$ i.e the probability density of choosing in the set of married men, a man of characteristics x married to a woman of characteristics y. We denote by $\pi(x,y|M=1,\mathcal{F})$ the conditional probability density of the vector (X,Y) at the point (x,y) with respect to the event $\{M=1,\mathcal{F}\}$ i.e the probability density of choosing in the set of married women, a woman of characteristics y and married to a man of characteristics x. These two distributions are actually equal since the number of married men and the number of married women are same. We have:

$$(X,Y)|\{M=1,\mathcal{H}\} \overset{\mathcal{L}}{=} (X,Y)|\{M=1,\mathcal{F}\},$$

and the density function of the distribution is $\pi(x,y)$. To explain more this relation, one must be convinced of the fact, the probability to choose in the set of married men, a man with characteristics x matched to a woman with characteristics y is exactly equal to the probability to choose in the set f married women, a woman with characteristics y matched to a man with characteristics x because the number of men with characteristics x matched to women with characteristics y is equal to the number of women with characteristics y matched to men with characteristics x.

3.2.2 Marriage surplus and Heterogeneity

The population is infinite. The key assumption here (previously suggested by Choo and Siow (2006) and formalized by Galichon and Salanié (2012) and Chiappori, Salanié and Weiss (2017)) is that the joint utility created when a man m with attributes x_m marries a woman w with attributes y_w rules out interactions between their unobserved characteristics, conditional on (x_m, y_w). More precisely, we assume that the joint utility from marriage can be written as

$$S(x_m, y_w) + \sigma\varepsilon_m(y_w) + \tau\eta_w(x_m)$$

where $S(x, y)$ is the joint systematic utility generated by the matching between a man of attributes x and a woman of attributes y, $\varepsilon_m(y)$ is a stochastic process that represent the specific gain in utility obtained by a man m with a woman with attributes y and $\eta_w(x)$ is a stochastic process that represent the specific gain in utility obtained by a woman w with a man with attributes x, and σ and τ are positive parameters. The joint utility from marriage represents the utility that partners will share between them, according to a rule that depends on the competition force in the marriage market.

We assume that any single man m of attributes $x = x_m$ has utility given by

$$u_m^0 = U^0(x) + \sigma\varepsilon_m^0,$$

where $U^0(x)$ is a known function and ε_m^0 as a stochastic term, specific to individual m, with a distribution that will be characterized below, and σ is a parameter. Similarly, we assume that any single woman w of attributes $y = y_w$ has utility given by

$$v_w^0 = V^0(y) + \tau\eta_w^0,$$

where $V^0(y)$ is a known function and η_w^0 as a stochastic term.

We suppose that the market equilibrium is stable in the sense of Gale and Shapley (1962), that is, there is no married person who would rather be single, and there is no pair of (married or unmarried) persons who prefer to form a new union. A man m marries a woman w if (a) the woman w belongs to the set of $\mathcal{W}$, (b) brings to him the highest level of utility and (c) brings a higher utility than his utility of remaining single. A woman w matches a man m if the converse of conditions (a)-(c) for individual w are satisfied. Let us denote respectively by u_m and v_w the utility at equilibrium of a man m of attributes x_m and the utility at equilibrium of a woman w of attributes y_w. One implication of this result is formulated in the following proposition.

Proposition 3.1. *The stable matching is such that, under the assumptions stated above, for any man m of attributes x_m,*

$$u_m(x_m) = \max\left(\max_{w\in\mathcal{W}}\left(U(x_m, y_w) + \sigma\varepsilon_m(y_w)\right), U^0(x_m) + \sigma\varepsilon_m^0\right) \tag{3}$$

and, for any woman w of attributes y_w,

$$v_w(y_w) = \max\left(\max_{m\in\mathcal{M}}\left(V(x_m, y_w) + \tau\eta_w(x_m)\right), V^0(y_w) + \tau\eta_w^0\right), \tag{4}$$

for some functions $U(x,y)$ and $V(x,y)$ such that $U(x,y)+V(x,y) = S(x,y)$.

Proof. The proof is then similar to that of Galichon and Salanié (2012). From well-known results (Shapley and Shubik, 1971), the equilibrium utilities solve the system of functional equations:

$$\begin{aligned} u_m(x_m) &= \max\left(\max_{w\in\mathcal{W}}(S(x_m, y_w) + \sigma\varepsilon_m(y_w) + \tau\eta_w(x_m) - v_w), U^0(x_m) + \sigma\varepsilon_m^0\right) \\ v_w(y_w) &= \max\left(\max_{m\in\mathcal{M}}(S(x_m, y_w) + \sigma\varepsilon_m(y_w) + \tau\eta_w(x_m) - u_m), V^0(y_w) + \tau\eta_w^0\right) \end{aligned}$$

The optimization problem in parentheses in the first equation can equivalently be written as:

$$\max_{y\in\mathcal{Y}}\left(S(x_m, y) + \sigma\varepsilon_m(y) + \max_{w:y_w=y}\left(\tau\eta_w(x_m) - v_w\right)\right)$$

or

$$\max_{y\in\mathcal{Y}}\left(S(x_m, y) + \sigma\varepsilon_m(y) - \min_{w:y_w=y}\left(v_w - \tau\eta_w(x_m)\right)\right).$$

If we define

$$V(x_m, y_w) = \min_{w:y_w=y}(v_w - \tau\eta_w(x_m)) \quad \text{and} \quad U(x_m, y_w) = S(x_m, y_w) - V(x_m, y_w),$$

we obtain:

$$u_m(x_m) = \max\left(\max_{y\in\mathcal{Y}}(U(x_m, y) + \sigma\varepsilon_m(y)), U^0(x_m) + \sigma\varepsilon_m^0\right).$$

Similarly, we obtain the value for v_w. ■

By now we will denote by $\bar{U}(x, y)$ and $\bar{V}(x, y)$ the individual net surpluses. In other words, we have:

$$\bar{U}(x, y) = U(x, y) - U^0(x) \ \text{ and } \ \bar{V}(x, y) = V(x, y) - V^0(y)$$

The net joint surplus is denoted $\bar{S}(x, y)$ i.e

$$\bar{S}(x, y) = \bar{U}(x, y) + \bar{V}(x, y) = S(x, y) - U^0(x) - V^0(y).$$

We also adopt the following notations:

$$\bar{u}_m = u_m - U^0 \ \text{ and } \ \bar{v}_w = v_w - V^0.$$

With these notations, in other words, the proposition stated above asserts that, the stable matching is such that, under the assumptions stated previously, for any man m of attributes x_m,

$$\bar{u}_m(x_m) = \max\left(\max_{w\in\mathcal{W}}\left(\bar{U}(x_m, y_w) + \sigma\varepsilon_m(y_w)\right), \sigma\varepsilon_m^0\right) \tag{5}$$

and, for any woman w of attributes y_w,

$$\bar{v}_w(y_w) = \max\left(\max_{m\in\mathcal{M}}\left(\bar{V}(x_m, y_w) + \tau\eta_w(x_m)\right), \tau\eta_w^0\right), \tag{6}$$

for some functions $\bar{U}(x, y)$ and $\bar{V}(x, y)$ such that $\bar{U}(x, y) + \bar{V}(x, y) = \bar{S}(x, y)$.

3.2.3 Identification of the net individual surpluses

In what follows, we shall focus on the net expected utility of each individual conditional on x and y, respectively. To simplify notation, we write

$$G_x(\bar{U}) = \mathbb{E}_P(\bar{u}\,|\,x)$$

and

$$H_y(\bar{V}) = \mathbb{E}_Q(\bar{v}\,|\,y)$$

the net expected value of the utility of man with attributes x and the net expected value of the utility of woman with attributes y, respectively. This notation emphasizes the fact that the net expected values depend on the net utility functions $\bar{U}(x, y)$ and $\bar{V}(x, y)$. Recall that $(X, Y)|\{M = 1, \mathcal{H}\}$ and $(X, Y)|\{M = 1, \mathcal{F}\}$ have the same distribution and we will denote it by Π. We have:

$$\Pi(x, y|M = 1, \mathcal{H}) = \Pi(x, y|M = 1, \mathcal{F})$$

The joint density function of $(X, Y)|\{M = 1, \mathcal{H}\}$ is $\pi(x, y|M = 1, \mathcal{H})$, i.e the joint density of (X, Y) on the population of married men where Y represents the characteristics of their partner. In that case, it as if the econometrician observes married men, observes their characteristics X and asks them the characteristics Y of their partners. The joint density function of $(X, Y)|\{M = 1, \mathcal{F}\}$ is $\pi(x, y|M = 1, \mathcal{F})$ that is the joint density of (X, y) on the population of married women; in that situation, imagine that the econometrician observes married women, observes their characteristics Y and asks them the characteristics X of their partners. We have:

$$\pi(x, y|M = 1, \mathcal{H}) = \pi(x, y|M = 1, \mathcal{F}) =: \pi(x, y)$$

The function $\pi(x, y)$ is then actually the joint density function of (X, Y) conditionally to marriage on each side of the market. The quantity $\pi(y, M = 1|x, \mathcal{H})$ is the probability density of choosing in the global population a woman such that her attributes are y and such that she belongs to the set of matched women with men of attributes x. We have:

$$\pi(y, M = 1|x, \mathcal{H}) = \frac{\pi(y, M = 1, x|\mathcal{H})}{f(x)} = \frac{\pi(x, y|M = 1, \mathcal{H}).P(M = 1|\mathcal{H})}{f(x)} = \frac{\alpha\, \pi(x, y)}{f(x)}$$

For women we have: the probability density of choosing in global population a man such that his attributes are x and such that he belongs to the set of matched men with women of attributes y is:

$$\pi(x, M = 1|y, \mathcal{F}) = \frac{\pi(x, M = 1, y|\mathcal{F})}{g(y)} = \frac{\pi(x, y|M = 1, \mathcal{F}).P(M = 1|\mathcal{F})}{g(y)} = \frac{\beta\,\pi(x, y)}{g(y)}$$

The problems (5) and (6) can be seen as the primal problems. The corresponding dual problems, which will be used hereafter, are defined as:

$$\max_{\bar{U}(x,\cdot)} \left(\int_{\mathcal{Y}} \frac{\alpha\pi(x, y)}{f(x)} \cdot \bar{U}(x, y) \cdot dy - G_x(\bar{U}) \right) \tag{7}$$

where $\alpha\pi(x, y)/f(x)$ is density of men of attributes x married to a woman of attributes y in the whole population of men, and

$$\max_{\bar{V}(\cdot,y)} \left(\int_{\mathcal{X}} \frac{\beta\pi(x, y)}{g(y)} \cdot \bar{V}(x, y) \cdot dx - H_y(\bar{V}) \right) \tag{8}$$

where $\beta\pi(x, y)/g(y)$ is the density of women of attributes y married to a man of attributes x in the whole population of women.

As we have:

$$\kappa = \nu.\alpha = (1 - \nu).\beta,$$

the conjugate (7) and (8) can respectively be rewritten as:

$$G_x^* \left(\frac{\kappa\pi}{f} \right) = \max_{\bar{U}(x,\cdot)} \left(\int_{\mathcal{Y}} \frac{\kappa\pi(x, y)}{f(x)} \cdot \bar{U}(x, y) \cdot dy - \nu G_x(\bar{U}) \right) \tag{9}$$

and

$$H_y^* \left(\frac{\kappa\pi}{g} \right) = \max_{\bar{V}(\cdot,y)} \left(\int_{\mathcal{X}} \frac{\kappa\pi(x, y)}{g(y)} \cdot \bar{V}(x, y) \cdot dx - (1 - \nu) H_y(\bar{V}) \right) \tag{10}$$

As pointed out by Galichon and Salanié (2012), we have:

$$\nu \frac{\partial G_x(\bar{U})}{\partial \bar{U}}(x, y) = \frac{\kappa\pi(x, y)}{f(x)} \quad \text{i.e} \quad \frac{\partial G_x(\bar{U})}{\partial \bar{U}}(x, y) = \frac{\alpha\pi(x, y)}{f(x)} \tag{11}$$

and

$$(1-\nu)\frac{\partial H_y(\bar{V})}{\partial \bar{V}}(x,y) = \frac{\kappa\pi(x,y)}{g(y)} \quad \text{i.e} \quad \frac{\partial H_y(\bar{V})}{\partial \bar{V}}(x,y) = \frac{\beta\pi(x,y)}{g(y)}, \tag{12}$$

from the envelop theorem, where $\alpha\pi(x,y)/f(x)$ is the probability for man of attributes x of being married to a woman of attributes y and $\beta\pi(x,y)/g(y)$ is the probability for woman of attributes y of being married to a man of attributes x.

The first order conditions of the optimization problems (9) and (10) are satisfied if conditions (11) and (12) hold, so that these optimization problems are indeed the dual of (3) and (4).[6] From the envelop theorem, we then have:

$$\frac{\partial G_x^*(\kappa\pi/f)}{\partial(\kappa\pi/f)}(x,y) = \bar{U}(x,y) \qquad \text{and}$$
$$\frac{\partial H_y^*(\kappa\pi/g)}{\partial(\kappa\pi/g)}(x,y) = \bar{V}(x,y).$$

The above identities show that utility functions can be recovered from observed probabilities.

3.2.4 Social Net Surplus and Optimal Matching

As shown by Galichon and Salanié (2012), the equilibrium state of the market aims to maximize the total expected utility and this is equivalent to maximize the total net surplus as the functions U^0 and V^0 are assumed to be exogenous to the model. The social net surplus is:

$$\bar{\mathcal{W}} = \nu\mathbb{E}\left[\bar{u}_m\right] + (1-\nu)\mathbb{E}\left[\bar{v}_w\right]$$

We can then rewrite the social net surplus for a particular feasible matching as:

$$\bar{\mathcal{W}} = \nu \int_{\mathcal{X}} G_x(\bar{U})(x)\cdot f(x)\cdot dx + (1-\nu)\int_{\mathcal{Y}} H_y(\bar{V})(y)\cdot g(y)\cdot dy \tag{13}$$

using the previously defined notation.

We then get the following proposition:

[6]Technically, the functions $G_x^*(\kappa\pi/f)$ and $H_y^*(\kappa\pi/g)$ are respectively the conjugates of $G_x(\bar{U})$ and $H_y(\bar{V})$ by the Legendre-Fenchel transformation.

Proposition 3.2. *The stable matching* (π, κ) *maximizes the social net surplus, i.e.,*

$$\bar{\mathcal{W}} = \max_{\{\pi,\kappa\}} \left(\kappa \int_{\mathcal{X}} \int_{\mathcal{Y}} \pi(x,y) \cdot \bar{S}(x,y) \cdot dx \cdot dy - \mathcal{E}(\pi,\kappa) \right)$$

where

$$\mathcal{E}(\pi,\kappa) = \int_{\mathcal{X}} G_x^* \left(\frac{\kappa\pi}{f} \right) \cdot f(x) \cdot dx + \int_{\mathcal{Y}} H_y^* \left(\frac{\kappa\pi}{g} \right) \cdot g(y) \cdot dy.$$

Proof. Recalling that from (11) and (12)

$$\nu \frac{\partial G_x}{\partial \bar{U}}(\bar{U})(x,y) = \frac{\kappa\pi(x,y)}{f(x)} \qquad \text{and} \qquad (1-\nu)\frac{\partial H_y}{\partial \bar{V}}(\bar{V})(x,y) = \frac{\kappa\pi(x,y)}{g(y)},$$

that is, the first order conditions of (9) and (10) are satisfied, with $\kappa = \nu\alpha = (1-\nu)\beta$. Therefore, at the equilibrium, we have:

$$\nu G_x(\bar{U}) = \int_{\mathcal{Y}} \frac{\kappa\pi(x,y)}{f(x)} \bar{U}(x,y) dy - G_x^* \left(\frac{\kappa\pi}{f} \right)$$

and

$$(1-\nu) H_y(\bar{V}) = \int_{\mathcal{X}} \frac{\kappa\pi(x,y)}{g(y)} \bar{V}(x,y) dx - H_y^* \left(\frac{\kappa\pi}{g} \right)$$

The expected value of u and v, respectively, can be broken down into two terms: the first one is the expected value of the systematic part and the second one the expected value of the stochastic part. From (13) we then write the social surplus as:

$$\begin{aligned} \bar{\mathcal{W}} &= \kappa \int_{\mathcal{X}} \int_{\mathcal{Y}} \pi(x,y) \cdot \bar{U}(x,y) \cdot dx \cdot dy + \kappa \int_{\mathcal{X}} \int_{\mathcal{Y}} \pi(x,y) \cdot \bar{V}(x,y) \cdot dx \cdot dy \\ &\quad - \left(\int_{\mathcal{X}} G_x^* \left(\frac{\kappa\pi}{f} \right) \cdot f(x) \cdot dx + \int_{\mathcal{Y}} H_y^* \left(\frac{\kappa\pi}{g} \right) \cdot g(y) \cdot dy \right) \end{aligned}$$

Finally, using $\bar{U}(x,y) + \bar{V}(x,y) = \bar{S}(x,y)$ and the definition of $\mathcal{E}(\pi,\kappa)$, we obtain:

$$\bar{\mathcal{W}}(\pi,\kappa) = \kappa \int_{\mathcal{X}} \int_{\mathcal{Y}} \pi(x,y) \cdot \bar{S}(x,y) \cdot dx \cdot dy - \mathcal{E}(\pi,\kappa),$$

which is maximized with respect to (π, κ). ∎

The first part of the expression of the social net surplus is the averaged net joint systematic surplus of couples. The second part $\mathcal{E}(\pi, \kappa)$ is a generalized entropy term (in Galichon and Salanié's (2012) terminology), linked to the heterogeneity of sympathy shocks among the population. If σ and τ converge to zero, then the second term vanishes.

3.2.5 A Convenient Specification and its Interpretation

The previous proposition holds under general conditions. To obtain a convenient specification, we follow Dupuy and Galichon (2014) and characterize the stochastic structure of the model as follows.

Assumption S.

(i) The stochastic processes $\varepsilon^m(y)$ and $\eta^w(x)$ are max-stable process of the form:

$$\begin{aligned} \varepsilon^m(y) &= \max_k(\varepsilon_k^m : y_k = y) \text{ if the set } \{k : y_k = y\} \text{ is non-empty} \\ &= -\infty \text{ otherwise,} \end{aligned}$$

where $\{(y_k^m, \varepsilon_k^m),\ k \in \mathbb{K}\}$ follows a Poisson process on the space $\mathcal{Y} \times \mathbb{R}$ of intensity $e^{-\varepsilon} dy\, d\varepsilon$, and

$$\begin{aligned} \eta^w(x) &= (\eta_l^w : x_l = x) \text{ if the set } \{l : x_l = x\} \text{ is non-empty} \\ &= -\infty \text{ otherwise.} \end{aligned}$$

where $\{(x_k^w, \eta_k^w), l \in \mathbb{L}\}$ follows a Poisson process on $\mathcal{X} \times \mathbb{R}$ with intensity $e^{-\eta} dx\, d\eta$.

(ii) The stochastic terms ε_m^0 and η_w^0 follow a standard Gumbel distribution, independent of ε_k^m for $k \in \mathbb{K}$ and independent of η_l^w for $l \in \mathbb{L}$, respectively.

Intuitively, each man m of attributes $x = x_m$ meets (or considers seriously) only a random subset of the population of women to make his choice. These women he knows will be called "acquaintances". This subset can be infinite and we index the acquaintances by $k \in \mathbb{K}$. The attributes of these women, potential partners of man m will denoted by y_k^m for any particular k. Thus, from each acquaintance k, man m gets a utility equal to: $U(x, y_k^m) + \sigma\varepsilon_k^m$. The intensity $e^{-\varepsilon} dy\, d\varepsilon$ represents the probability a man has an acquaintance with a woman k such that y_k^m is in a set of size dy and the sympathy shock ε_k^m lies in a set of size $d\varepsilon$. The decrease of the intensity in ε translates the idea according to which there is a low probability to have an acquaintance with women with whom the sympathy shock is high and there is a high probability to have an acquaintance with women with whom the sympathy is low. In addition, for any disjoints subsets $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{Y} \times \mathbb{R}$, the fact that m has an acquaintance in $\mathcal{A}$ is independent from the fact that m has an acquaintance in $\mathcal{B}$. Similarly, we assume for any woman w of attributes $y = y_w$, that she has acquaintances in a random subset of the global population of men, and each acquaintance with a man $l \in \mathbb{L}$ of attributes x_l^w generates for the woman w a utility equal to $V(x_l^w, y) + \tau\eta_l^w$.

Thanks to the above assumption, the choice of a partner can be seen as a discrete choice problem. That is, each individual does his/her choice in a set of acquaintances. In addition, the distribution of the utilities u_m and v_w individuals get at the equilibrium can be derived and the conditional expectation of these utilities can then be computed.

Proposition 3.3. *Assume a stable matching and S. Then, for any man m of attributes x_m, whose utility at the equilibrium is $u_m(x_m)$, and for any woman w of attributes y_w whose utility at the equilibrium is $v_w(y_w)$ we have:*

$$u_m(x_m) \sim Gumbel\left(U^0(x_m) + \sigma \ln\left(1 + \int_{\mathcal{Y}} e^{\frac{\bar{U}(x_m,y)}{\sigma}} dy\right), \sigma\right)$$

and

$$v_w(y_w) \sim Gumbel\left(V^0(y_w) + \tau \ln\left(1 + \int_{\mathcal{X}} e^{\frac{\bar{V}(x,y_w)}{\tau}} dy\right), \tau\right)$$

Proof. We consider the case of the distribution of men' utility and start from the preceding result:

$$u_m(x_m) = \max\left(\max_{w \in \mathcal{W}} \left(U(x_m, y_w) + \sigma\varepsilon_m(y_w)\right), U^0(x_m) + \sigma\varepsilon_m^0\right).$$

Using assumption S, the problem can be written as:

$$u_m(x_m) = \max(\max_{k\in\mathbb{K}} \left(U(x_m, y_k^m) + \sigma\varepsilon_k^m\right), U^0(x_m) + \sigma\varepsilon_m^0)$$

where (y_k^m, ε_k^m) is following a Poisson process of intensity $e^{-\varepsilon}d\varepsilon dy$. If the cumulative distribution of $u_m(x)$ is denoted by F_{u_m}, then:

$$F_u(t) = \Pr\left(\varepsilon_m^0 \leq \frac{t - U^0(x)}{\sigma}\right) \cdot \Pr\left(\varepsilon_k^m \leq \frac{t - U(x_m, y_k^m)}{\sigma}, \forall k\right)$$

because ε_m^0 and ε_k^m are independent. Since ε_m^0 is following a standard Gumbel distribution and (ε_k^m, y_k) is following a Poisson distribution of intensity $e^{-\varepsilon}d\varepsilon dy$, we have:

$$F_u(t) = \exp\left(-\exp\left(-\frac{t - U^0(x_m)}{\sigma}\right)\right) \times \exp\left(-\int_{\mathcal{Y}} \int_{\frac{t-U(x_m,y)}{\sigma}}^{+\infty} \exp\left(-\varepsilon\right) d\varepsilon dy\right)$$

In this expression, the second term of the right-hand-side represents the probability of not observing (y_k^m, ε_k^m) in the set $\mathcal{Y} \times [(t - U(x_m, y))/\sigma, +\infty]$. Integrating the second term of this expression with respect ot ε gives:

$$F_u(t) = \exp\left(-\exp\left(-\frac{t - U^0(x_m)}{\sigma}\right)\right) \times \exp\left(-\int_{\mathcal{Y}} \exp\left(\frac{t - U(x_m, y)}{\sigma}\right) dy\right)$$

Simplifying gives:

$$F_u(t) =$$
$$\exp\left(-\exp\left(\frac{1}{\sigma}\left(U^0(x_m) + \sigma\log\left(1 + \int_{\mathcal{Y}} \exp\left(\frac{U(x_m, y) - U^0(x_m)}{\sigma}\right) dy\right) - t\right)\right)\right)$$

That is, u_m has a Gumbel distribution of parameters

$$\left\{U^0(x_m) + \sigma\log\left(1 + \int_{\mathcal{Y}} \exp\left(\frac{U(x_m, y) - U^0(x_m)}{\sigma}\right) dy\right), \sigma\right\}.$$

We do the same demonstration for $v_w(y)$ to prove its distribution. ■

From well-known results regarding the Gumbel distribution, we directly get:

$$G_x(\bar{U}) = \sigma\ \ln\left(1 + \int_{\mathcal{Y}} \exp\left(\frac{\bar{U}(x, y)}{\sigma}\right) dy\right) + \sigma\gamma \tag{14}$$

and

$$H_y(\bar{V}) = \tau \ln\left(1 + \int_{\mathcal{X}} \exp\left(\frac{\bar{V}(x,y)}{\tau}\right) dy\right) + \tau\gamma, \tag{15}$$

where $\gamma = \ln(\ln(2))$ is the Euler-Mascheroni constant. This result is very similar to the expressions obtained in the discrete case, which are $\sigma \ln(1 + \sum_{y_i \in \mathcal{Y}} \exp(\bar{U}(x_j, y_i)/\sigma)$ and $\tau \ln(1 + \sum_{x_j \in \mathcal{X}} \exp(\bar{V}(x_j, y_i)/\tau)$, where x_j and y_i are discrete variables. We remark that conditionally on x, the expected value of the utility of a man is higher at the equilibrium than the expected value of his utility of remaining single and the difference between the two is $\sigma \ln\left(1 + \int_{\mathcal{Y}} \exp\left(\frac{\bar{U}(x,y)}{\sigma}\right) dy\right)$.

From Proposition (3.3), we can also derive an expression for the systematic utility of each individual. If we apply the identity (11) with the specification for men's conditional utility (14), we obtain:

$$\frac{\exp\left(\frac{\bar{U}(x,y)}{\sigma}\right)}{1 + \int_{\mathcal{Y}} \exp\left(\frac{\bar{U}(x,y)}{\sigma}\right) dy} = \frac{\alpha\pi(x,y)}{f(x)} \tag{16}$$

Similarly, if we apply the identity (12) with the specification for women' conditional utility (15), we obtain:

$$\frac{\exp\left(\frac{\bar{V}(x,y)}{\tau}\right)}{1 + \int_{\mathcal{X}} \exp\left(\frac{\bar{V}(x,y)}{\tau}\right) dx} = \frac{\beta\pi(x,y)}{g(y)}. \tag{17}$$

Together with identity (1), the right-hand-side of expression (16) can be written as:

$$\frac{\alpha}{1-\alpha} \frac{\pi(x,y)}{f_0(x)} \left(1 + \frac{\alpha}{1-\alpha} \frac{f_1(x)}{f_0(x)}\right)^{-1}.$$

In addition, the feasibility of matchings implies that

$$f_1(x) = \int_{\mathcal{Y}} \pi(x,y) dy.$$

Thus, we have

$$\frac{\exp\left(\frac{\bar{U}(x,y)}{\sigma}\right)}{1 + \int_{\mathcal{Y}} \exp\left(\frac{\bar{U}(x,y)}{\sigma}\right) dy} = \frac{\alpha}{1-\alpha} \frac{\pi(x,y)}{f_0(x)} \left(1 + \int_{\mathcal{Y}} \frac{\alpha}{1-\alpha} \frac{\pi(x,y)}{f_0(x)} dy\right)^{-1},$$

which implies that:

$$\exp\left(\frac{\bar{U}(x,y)}{\sigma}\right) = \frac{\alpha}{(1-\alpha)}\frac{\pi(x,y)}{f_0(x)}. \tag{18}$$

In other words, the systematic utility obtained by a man of attributes x married to a woman of attributes y is equal to the ratio of the joint dentity of observing matches of man of attributes x and woman of attributes y and the density of observing a single of attributes x. Similarly, we have:

$$\exp\left(\frac{\bar{V}(x,y)}{\tau}\right) = \frac{\beta}{1-\beta}\frac{\pi(x,y)}{g_0(y)}. \tag{19}$$

with a similar interpretation. This straightforwardly generalizes the expressions obtained by Choo and Siow (2006). By comparison, our results are quite similar the results found by Dupuy and Galichon (2014). In fact, for $U^0(x) = 0$ and $V^0(y) = 0$, our result can be written at the form:

$$U(x,y) = \sigma \ln\left(\alpha\pi(x,y)\right) - \sigma \ln\left((1-\alpha)f_0(x)\right)$$

and

$$V(x,y) = \tau \ln\left(\beta\pi(x,y)\right) - \tau \ln\left((1-\beta)g_0(y)\right)$$

And Dupuy and Galichon (2014) found that:

$$U(x,y) = \sigma \ln\left(\pi(x,y)\right) - a(x) \quad \text{and} \quad U(x,y) = \tau \ln\left(\pi(x,y)\right) - b(y)$$

where:

$$a(x) = \frac{\sigma}{2}\ln\left(\int_{\mathcal{Y}} \frac{e^{[U(x,y')/(\sigma/2)]}}{f(x)}dy'\right) \quad \text{and} \quad b(y) = \frac{\sigma}{2}\ln\left(\int_{\mathcal{X}} \frac{e^{[V(x',y)/(\sigma/2)]}}{g(y)}dx'\right)$$

In the following theorem we formalize the theoretical results obtained in this framework. We recall some notations: $\bar{U}(x,y) = U(x,y) - U^0(x)$ is the net surplus of men, $\bar{V}(x,y) = V(x,y) - V^0(y)$ is the net surplus of women and by $\bar{S}(x,y) = S(x,y) - U^0(x) - V^0(y) = \bar{U}(x,y) + \bar{V}(x,y)$ is the net joint surplus.

Theorem 3.1. *Assume a stable matching and Assumption S. Then we have:*

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

$$\bar{S}(x,y) = \ln\left(\left(\frac{\alpha}{1-\alpha}\right)^{\sigma}\left(\frac{\beta}{1-\beta}\right)^{\tau}\frac{(\pi(x,y))^{\sigma+\tau}}{(f_0(x))^{\sigma}\,(g_0(y))^{\tau}}\right)$$

2. *The systematic surplus of a man of attributes x from a matching with a woman of attributes y is such as:*

$$\bar{U}(x,y) = \frac{\sigma}{\sigma+\tau}\left(\bar{S}(x,y) - \tau\ln\left(\frac{1-\alpha}{\alpha}f_0(x)\right) + \tau\ln\left(\frac{1-\beta}{\beta}g_0(y)\right)\right)$$

3. *The systematic surplus of a woman of attributes y from a matching with a man of attributes x is such as:*

$$\bar{V}(x,y) = \frac{\tau}{\sigma+\tau}\left(\bar{S}(x,y) + \sigma\,\ln\left(\frac{1-\alpha}{\alpha}f_0(x)\right) - \sigma\,\ln\left(\frac{1-\beta}{\beta}g_0(y)\right)\right)$$

4. *for any $x \in \mathcal{X}$,*

$$f_0(x) = \frac{1}{(1-\alpha)\left(1+\int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}}dy\right)}f(x) \;\; \textit{and} \;\; f_1(x) = \frac{\int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}}dy}{\alpha\left(1+\int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}}dy\right)}f(x)$$

5. *for any $y \in \mathcal{Y}$,*

$$g_0(y) = \frac{1}{(1-\beta)\left(1+\int_{\mathcal{X}} e^{\frac{\bar{V}(x,y)}{\tau}}dx\right)}g(y) \;\; \textit{and} \;\; g_1(y) = \frac{\int_{\mathcal{X}} e^{\frac{\bar{V}(x,y)}{\tau}}dx}{\beta\left(1+\int_{\mathcal{X}} e^{\frac{\bar{V}(x,y)}{\tau}}dx\right)}g(y)$$

6. *The probabilities $\alpha = \Pr(M=1|\mathcal{H})$ and $\beta = \Pr(M=1|\mathcal{F})$ are determined by:*

$$\alpha = \int_{\mathcal{X}}\int_{\mathcal{Y}} \frac{e^{\frac{\bar{U}(x,y)}{\sigma}}}{1+\int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}}dy}f(x)dx\,dy \;\; \textit{and} \;\; \beta = \int_{\mathcal{X}}\int_{\mathcal{Y}} \frac{e^{\frac{\bar{V}(x,y)}{\tau}}}{1+\int_{\mathcal{X}} e^{\frac{\bar{V}(x,y)}{\tau}}dx}g(y)dx\,dy$$

The proof is given in the appendix.

3.3 Estimation Method and Additional Results

The estimation method is based on the maximization of an approximated log-likelihood function with respect to a set of parameters. We proceed as follows. Firstly, we derive the maximum likelihood function assuming that systematic utility functions are known. Secondly, we propose an approximation for systematic utility functions that depend on the primitives of the model, i.e., the total surplus from marriage, the distributions of characteristics and the proportion of men or women (the latter is directly observed). To simplify we suppose hereafter that $\sigma = \tau = 1$ that is, the distribution parameters are equal to one. We also suppose that x and y are scalars (coinciding with log incomes in the empirical application). We then specify the net joint surplus as:

$$\bar{S}(x,y) = -a_2^2 x^2 + c\,x\,y - b_2^2 y^2 + a_1 x + b_1 y = (1 \;\; x \;\; x^2) \begin{pmatrix} 0 & b_1 & -b_2^2 \\ a_1 & c & 0 \\ -a_2^2 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ y \\ y^2 \end{pmatrix}$$

where a_1, b_1, a_2^2, b_2^2 and c are parameters to be estimated. With this particular form, and if the coefficients of x^2 and y^2 are negative, the integrability of $\exp S(x,y)$ on the cartesian space $\mathcal{X} \times \mathcal{Y}$ is guaranteed. This is necessary because the optimal matching joint density $\pi(x,y)$ is a bounded function on $\mathcal{X} \times \mathcal{Y}$. The matrix:

$$K = \begin{pmatrix} 0 & b_1 & -b_2^2 \\ a_1 & c & 0 \\ -a_2^2 & 0 & 0 \end{pmatrix}$$

can be interpreted as the affinity matrix (in Dupuy and Galichon (2014)'s terminology). Finally, we assume that x and y follows normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \text{and} \quad g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

The negativity of $-a_2^2$ and $-b_2^2$ will imply a concavity of the surpluses with a maximum. This is realistic because when for example sympathy shocks are negligible, the optimal choice is given by the maximization of the surpluses, because the utility is quasi equal to its deterministic part then, and the concavity of a quadratic joint surplus function guarantees the existence of a maximum. If the affinity matrix is definite negative, this guarantees that the extremum of $\bar{S}(x, y)$ is a maximum. The density functions $f(x)$ and $g(y)$ depend on exogenous parameters that will be estimated independently of the model.

3.3.1 The Maximum Likelihood

To derive the likelihood function, we denote the set of couples, the set of single men and the set of single women by $\mathcal{C}$, $\mathcal{S}_m$ and $\mathcal{S}_w$ respectively. Considering a man m with characteristics $x_m = x$ and a woman w with characteristics $y_w = y$, the likelihood of the union (m, w) is the product of the conditional probability density the man m is matched to a woman with characteristics y with respect to $x_m = x$ with the conditional probability density the woman w is matched to a man with characteristics x with respect to $y_w = y$. The probability density to observe in the male population a man with characteristics x matched to a woman with characteristics y (here only the man is observed since the observation is done on the male population) is the product of the probability of marriage on the male population with the probability density to observe in the set of married men, a man with characteristics x matched with a woman with characteristics y, i.e $\alpha\pi(x, y)$. The conditional probability density to observe in the male population such a man m with respect to $x_m = x$ is then:

$$\frac{\alpha\pi(x, y)}{f(x)}$$

Analogously the probability density to observe in the female population a woman with characteristics y matched to a man with characteristics x (here only the woman we observe as the observation is done on the female population) is the product of the probability of marriage on the female population with the probability density to observe in the set of married women, a woman with characteristics y matched with a man with characteristics x, that is $\beta\pi(x, y)$. Then the conditional probability density to observe in the female population such a woman w with respect to $y_w = y$ is:

$$\frac{\beta\pi(x, y)}{g(y)}$$

Things work similarly for single individuals. The probability density to observe in the male population, a single man with characteristics x is the probability of being single on the male population with the probability density in the set of single men, a man with characteristics x, that is: $(1-\alpha)f_0(x)$. And then the conditional probability density to observe in the male population such a man m with respect to $x_m = x$ is:

$$\frac{(1-\alpha)f_0(x)}{f(x)}$$

Same reasoning for single women leads to derive that the probability density to observe in the female population a single woman w with respect to $y_w = y$ is:

$$\frac{(1-\beta)g_0(y)}{g(y)}$$

The likelihood is then:

$$\mathcal{L} = \prod_{(m,w)\in\mathcal{C}} \frac{\alpha\pi(x_m, y_w)}{f(x_m)} \frac{\beta\pi(x_m, y_w)}{g(y_w)} \times \prod_{m\in\mathcal{S}_m} \frac{(1-\alpha)f_0(x_m)}{f(x)} \times \prod_{w\in\mathcal{S}_w} \frac{(1-\beta)g_0(y_w)}{g(y)}$$

The first term of the right-hand side gives the density for a man m and a woman w to be married; the second term the probability for a man m to be single, and the third term the probability for a woman w to be single.

To have a well-behaved likelihood function, we have to express it as a function of the primitives of the model. The density functions are related to individual utilities as summarized by the following lemma.

Lemma 3.1. *Assume a stable matching and S. Then, we have:*

$$\begin{aligned}
(a) \quad &: \quad \frac{\alpha\pi(x_m, y_w)}{f(x_m)} = \frac{\exp(\bar{U}(x_m, y_w))}{1 + \int_{\mathcal{Y}} \exp(\bar{U}(x_m, y))dy}, \\
(b) \quad &: \quad \frac{\beta\pi(x_m, y_w)}{g(y_w)} = \frac{\exp(\bar{V}(x_m, y_w))}{1 + \int_{\mathcal{X}} \exp(\bar{V}(x, y_w))dx}, \\
(c) \quad &: \quad \frac{(1-\alpha)f_0(x_m)}{f(x)} = \frac{1}{1 + \int_{\mathcal{Y}} \exp(\bar{U}(x_m, y))dy}, \\
(d) \quad &: \quad \frac{(1-\beta)g_0(y_w)}{g(y)} = \frac{1}{1 + \int_{\mathcal{X}} \exp(\bar{V}(x, y_w))dx}.
\end{aligned}$$

Proof. The conditions (a) and (b) are simply expressions (16) and (17). To obtain condition (c), we integrate expression (18) with respect to y and obtain:

$$\int_{\mathcal{Y}} \exp\left(\bar{U}(x,y)\right) dy = \frac{\alpha}{(1-\alpha)} \frac{\int_{\mathcal{Y}} \pi(x,y)dy}{f_0(x)} = \frac{\alpha}{(1-\alpha)} \frac{f_1(x)}{f_0(x)}$$

Together with (1), we obtain condition (c). Similarly, to obtain condition (d), we integrate expression (19) with respect to x. ■

Using this result, the likelihood function can be written as a function of systematic utilities. The next step is to express systematic utilities as a function of the primitives $\bar{S}(x,y)$, $f(x)$, $g(y)$ and ν. The following lemma gives equations that define systematic utilities for given primitives.

Lemma 3.2. *Assume a stable matching and S. Then, the men's systematic utility is implicitly defined as:*

$$\bar{U}(x,y) = \frac{1}{2}\bar{S}(x,y) + \frac{1}{2}\ln\left(\frac{(1-\nu)g(y)}{\nu f(x)}\right) + \frac{1}{2}A(\bar{U})(x,y), \tag{20}$$

where

$$\begin{aligned} A(\bar{U})(x,y) &= \ln\left(1 - \frac{\nu}{1-\nu}\int_{\mathcal{X}} \frac{f(x)}{g(y)} \frac{\exp\left(\bar{U}(x,y)\right)}{1+\int_{\mathcal{Y}} \exp\left(U(x,y)\right) dy} dx\right) \\ &\quad + \ln\left(1 + \int_{\mathcal{Y}} \exp\left(\bar{U}(x,y)\right) dy\right) \end{aligned}$$

and the women's systematic utility is implicitly defined as:

$$\bar{V}(x,y) = \frac{1}{2}\bar{S}(x,y) + \frac{1}{2}\ln\left(\frac{\nu f(x)}{(1-\nu)g(y)}\right) + \frac{1}{2}B(\bar{V})(x,y), \tag{21}$$

where

$$\begin{aligned} B(\bar{V})(x,y) &= \ln\left(1 - \frac{1-\nu}{\nu}\int_{\mathcal{Y}} \frac{g(y)}{f(x)} \frac{\exp\left(\bar{V}(x,y)\right)}{1+\int_{\mathcal{X}} \exp\left(\bar{V}(x,y)\right) dx} dy\right) \\ &\quad + \ln\left(1 + \int_{\mathcal{X}} \exp\left(\bar{V}(x,y)\right) dx\right) \end{aligned}$$

with $A(\bar{U})(x,y) + B(\bar{V})(x,y) = 0$.

Proof. Because of (16) and (17), we have the following identity:

$$\frac{\exp(\bar{V}(x,y))g(y)}{1+\int_{\mathcal{X}}\exp(\bar{V}(x,y))dy}\frac{1}{\beta} = \frac{\exp(\bar{U}(x,y))f(x)}{1+\int_{\mathcal{Y}}\exp(\bar{U}(x,y))dy}\frac{1}{\alpha},$$

i.e., the systematic utility functions are related. Then, using $\alpha \cdot \nu = \beta \cdot (1-\nu)$, we eliminate α and β and obtain:

$$\frac{\exp(\bar{V}(x,y))}{1+\int_{\mathcal{X}}\exp(\bar{V}(x,y))dx}\frac{1}{f(x)} = \frac{\nu}{1-\nu}\frac{\exp(\bar{U}(x,y))}{1+\int_{\mathcal{Y}}\exp(\bar{U}(x,y))dy}\frac{1}{g(y)} \tag{22}$$

If the right-hand side is denoted as $a(\bar{U})(x,y)$, we have:

$$\exp(\bar{V}(x,y)) = f(x)a(\bar{U})(x,y) + f(x)a(\bar{U})(x,y)\int_{\mathcal{X}}\exp(\bar{V}(x,y))dx.$$

Then, integrating left-hand side and right-hand side with respect to x gives:

$$\int_{\mathcal{X}}\exp(\bar{V}(x,y))dx = \int_{\mathcal{X}} f(x)a(\bar{U})(x,y)dx + \int_{\mathcal{X}} f(x)a(\bar{U})(x,y)dx\int_{\mathcal{X}}\exp(\bar{V}(x,y))dx$$

and rearranging gives::

$$1+\int_{\mathcal{X}}\exp(\bar{V}(x,y))dy = \frac{1}{1-\int_{\mathcal{X}} f(x)(\bar{U})a(x,y)dx}.$$

From (22) and the above expression, we deduce an expression for women' systematic utility:

$$\bar{V}(x,y) = \log\left(\frac{f(x)a(\bar{U})(x,y)}{1-\int_{\mathcal{X}} f(x)a(\bar{U})(x,y)dx}\right)$$

i.e

$$\bar{V}(x,y) = \log\left(f(x)\right) + \log\left(a(\bar{U})(x,y)\right) - \log\left(1-\int_{\mathcal{X}} f(x)a(\bar{U})(x,y)dx\right)$$

Using the identity $\bar{S}(x,y) = \bar{U}(x,y) + \bar{V}(x,y)$ gives:

$$\bar{S}(x,y) - \bar{U}(x,y) = \log\left(f(x)\right) + \log\left(a(\bar{U})(x,y)\right) - \log\left(1-\int_{\mathcal{X}} f(x)a(\bar{U})(x,y)dx\right)$$

Using the definition of $a(\bar{U})(x,y)$ and simplifying give expression (20). Similar reasoning gives expression (21). ■

The results of the lemma show the equations that relate utilities $U(x,y)$ and $V(x,y)$ to the primitives $\bar{S}(x,y)$, $f(x)$, $g(y)$, and the proportion ν. The analytical solution may be complicated. Yet, it is possible to obtain approximated solutions for $\bar{U}(x,y)$ and $\bar{V}(x,y)$. If we omit the last term on the right-hand side of (20) and (21), we have a first order approximation defined as:

$$\bar{U}_0(x,y) = \frac{1}{2}\bar{S}(x,y) + \frac{1}{2}\ln\left(\frac{(1-\nu)g(y)}{\nu f(x)}\right),$$

and

$$\bar{V}_0(x,y) = \frac{1}{2}\bar{S}(x,y) + \frac{1}{2}\ln\left(\frac{\nu f(x)}{(1-\nu)g(y)}\right).$$

Intuitively, each man with characteristics x married to a woman with characteristics y receives a little more or a little less than equal sharing of the total net surplus $\bar{S}(x,y)$, depending on whether the number of women with characteristics y is larger or smaller than the number of men with characteristics x. This first approximation is a good approximation only when the rate of marriage is very low and in the following chapter we prove actually that $\bar{U}_0(x,y)$ and $\bar{V}_0(x,y)$ are the analytical expressions of $\bar{U}(x,y)$ and $\bar{V}(x,y)$ in case of scarce matching on the market. For the estimation, we will then use a second order of approximation. Using the previous lemma, a second order approximation is defined as:

$$\bar{U}_1(x,y) = \frac{1}{2}\bar{S}(x,y) + \frac{1}{2}\ln\left(\frac{(1-\nu)g(y)}{\nu f(x)}\right) + \frac{1}{2}A(\bar{U}_0)(x,y)$$

and

$$\bar{V}_1(x,y) = \frac{1}{2}\bar{S}(x,y) + \frac{1}{2}\ln\left(\frac{\nu f(x)}{(1-\nu)g(y)}\right) + \frac{1}{2}B(\bar{V}_0)(x,y)$$

More generally, we define the sequences:

$$\forall\, k \in \mathbb{N},\ \bar{U}_{k+1}(x,y) = \frac{1}{2}\bar{S}(x,y) + \frac{1}{2}\ln\left(\frac{(1-\nu)g(y)}{\nu f(x)}\right) + \frac{1}{2}A(\bar{U}_k)(x,y)$$

and

$$\forall\, k \in \mathbb{N},\ \bar{V}_{k+1}(x,y) = \frac{1}{2}\bar{S}(x,y) + \frac{1}{2}\ln\left(\frac{\nu f(x)}{(1-\nu)g(y)}\right) + \frac{1}{2}B(\bar{V}_k)(x,y)$$

The convergence of these sequences as ensured by the fact that, from the lemma, the exact functions $\bar{U}$ and $\bar{V}$ verify the equations:

$$\forall\, x,\, \forall\, y,\ \bar{U}(x,y) = \frac{1}{2}\bar{S}(x,y) + \frac{1}{2}\ln\left(\frac{(1-\nu)g(y)}{\nu f(x)}\right) + \frac{1}{2}\mathbf{A}(\bar{U})(x,y)$$

and

$$\forall\, x,\, \forall\, y,\ \bar{V}(x,y) = \frac{1}{2}\bar{S}(x,y) + \frac{1}{2}\ln\left(\frac{\nu f(x)}{(1-\nu)g(y)}\right) + \frac{1}{2}\mathbf{B}(\bar{V})(x,y)$$

The uniqueness of V and of V guaranteed by their identification ensures the convergence of the sequences respectively to U and to V.

The procedure can then be iterated until the approximated systematic utility is sufficiently close to the true value. In the empirical application, we stop at the second order approximation.[7]

3.3.2 Discussion: Gender surplus gap and attractiveness

In this subsection, we try to analyze the gap in the systematic surpluses between partners within the households and to formulate a theoretical view on the question: What are the determinants of the surplus gap between partner within the households ? To do this, we base our development on showing the relation between the individual surpluses of the partners within the household shown in the proof of the previous lemma. In what follows we will first focus on defining the notion of attractiveness of individuals on the market. A natural measure of the attractiveness is the probability to be attractive for the other hand of the market. The attraction of women of characteristics y on men with characteristics x is the conditional probability density for a woman w to be matched with a man of characteristics x with respect to $y_w = y$. This conditional probability is:

$$\frac{\beta\pi(x,y)}{g(y)}$$

[7] As $\left|A(\bar{U})(x,y)\right| < 1$ and $\left|B(\bar{V})(x,y)\right| < 1$, we assume the approximation is reasonable.

This is the attraction of the woman w on the set of men with characteristics x. We can then derive the attractiveness of the woman with characteristics y on a particular man with characteristics x as the intensity of her attraction on this sub male population by:

$$\frac{\beta\pi(x,y)}{f(x)g(y)}$$

And analogously, the attractiveness of a man with characteristics x on a woman with characteristics y is then defined by:

$$\frac{\alpha\pi(x,y)}{f(x)g(y)}$$

We can remark that, when $\alpha = \beta$, i.e when the proportion of men is equal to the proportion of women in the whole population, the reciprocal attractions between a man with characteristics x and a woman with characteristics y are equal. This makes sense even if the number of men with characteristics x married to women with characteristics y at equilibrium is always equal to the number of women with characteristics y married to men with characteristics x, a difference in the proportions of the sides of the market implies a difference in the reciprocal strengths of attraction between the two sides and the less weighted side will be the more attractive because the more the side of the market is wide the lower is the conditional probability of marriage on it. This traduces somehow the scarcity effect on the attractiveness i.e attractive individuals are not the most frequent on the market. We have defined previously from (22) the following quantity:

$$a(\bar{U})(x,y) = \frac{\nu}{1-\nu}\frac{1}{g(y)}\frac{e^{\bar{U}(x,y)}}{1+\int_{\mathcal{Y}} e^{\bar{U}(x,y)}dy}$$

We define now as well:

$$b(\bar{V})(x,y) = \frac{1-\nu}{\nu}\frac{1}{f(x)}\frac{e^{\bar{V}(x,y)}}{1+\int_{\mathcal{X}} e^{\bar{V}(x,y)}dx}$$

From lemma, by using the relation:

$$\frac{\pi(x,y)}{f(x)} = \frac{1}{\alpha}\frac{e^{\bar{U}(x,y)}}{1+\int_{\mathcal{Y}} e^{\bar{U}(x,y)}dy} \quad \text{and} \quad \frac{\pi(x,y)}{g(y)} = \frac{1}{\beta}\frac{e^{\bar{V}(x,y)}}{1+\int_{\mathcal{X}} e^{\bar{V}(x,y)}dx}$$

we can remark that:

$$a(\bar{U})(x,y) = \frac{\beta\pi(x,y)}{f(x)g(y)} \quad \text{and} \quad b(\bar{V})(x,y) = \frac{\alpha\pi(x,y)}{f(x)g(y)}$$

Then the attractiveness of a woman with characteristics y on a man with characteristics x is $a(\bar{U})(x,y)$ and the reciprocal attractiveness of the man on her is $b(\bar{V})(x,y)$. The attractiveness of the woman on the man depends on $\bar{U}$ because it is subject to the individual net surplus of the men from a match with her and that determines their will to choose her. And reversely, the attractiveness of the man depends on how intensively women want to match with him and this is subject to their surplus from such union with him. We will in the following of this development show the determinant of the surplus gap between partners within the matching.

The surplus gap between a woman with characteristics and her partner with characteristics x i.e the difference $\bar{V}(x,y) - \bar{U}(x,y)$ From the proof of the lemma, we have the relations:

$$\bar{V}(x,y) = \ln\left(\frac{f(x)a(\bar{U})(x,y)}{1 - \int_{\mathcal{X}} f(x)a(\bar{U})(x,y)dx}\right)$$

and

$$\bar{U}(x,y) = \ln\left(\frac{g(y)b(\bar{V})(x,y)}{1 - \int_{\mathcal{Y}} g(y)b(\bar{V})(x,y)dy}\right)$$

These expressions can be rewritten. From (18), we have:

$$a(\bar{U})(x,y) = \frac{e^{\bar{V}(x,y)}}{1 + \int_{\mathcal{X}} e^{\bar{V}(x,y)}dx}\frac{1}{f(x)}$$

This implies that:

$$1 - \int_{\mathcal{X}} f(x)a(\bar{U})(x,y)dx = \frac{1}{1 + \int_{\mathcal{X}} e^{\bar{V}(x,y)}dx}$$

Note that the term $\int_{\mathcal{X}} f(x)a(\bar{U})(x,y)dx$ is the attractiveness of women with characteristics y on the whole male population, and the term $\frac{1}{1+\int_{\mathcal{X}} e^{\bar{V}(x,y)}dx}$ is the probability a woman with characteristics y to remain single. This relation shows that the attractiveness of a woman with characteristics y on the whole male population is simply the probability to be matched to a man. We can denote by $\beta(y)$ this probability, i.e:

$$\beta(y) = \int_{\mathcal{X}} f(x)a(\bar{U})(x,y)dx = \frac{\int_{\mathcal{X}} e^{\bar{V}(x,y)}dx}{1 + \int_{\mathcal{X}} e^{\bar{V}(x,y)}dx}$$

Analogously we have:

$$1 - \int_{\mathcal{Y}} g(y)b(\bar{V})(x,y)dy = \frac{1}{1 + \int_{\mathcal{Y}} e^{\bar{U}(x,y)}dy}$$

We also denote:

$$\alpha(x) = \int_{\mathcal{Y}} g(y)b(\bar{V})(x,y)dy = \frac{\int_{\mathcal{Y}} e^{\bar{U}(x,y)}dy}{1 + \int_{\mathcal{Y}} e^{\bar{U}(x,y)}dy}$$

This is the probability for a man with characteristics x to be matched and it is equal to the global attractiveness of a man with characteristics x on the whole female population. By using $\alpha\nu = \beta(1-\nu)$ combined to relations above, we can derive that:

$$\begin{aligned} \bar{V}(x,y) - \bar{U}(x,y) &= -\left[\ln\left((1-\nu)(1-\beta(y))g(y)\right) - \ln\left(\nu(1-\alpha(x))f(x)\right)\right] \\ &= -\ln\left[\frac{(1-\nu)(1-\beta(y))g(y)}{\nu(1-\alpha(x))f(x)}\right] \end{aligned}$$

The term $(1-\nu)(1-\beta(y))g(y)$ is the absolute probability density to choose in the whole population a single woman with characteristics y, in other words, it is the probability density absolute of celibacy of women with characteristics y in the whole market, not the conditional to their characteristics. And the term $\nu(1-\alpha(x))f(x)$ is the absolute probability density to choose in the whole population a single man with characteristics x, i.e the absolute probability density of single-hood of men with characteristics x in the whole population. The surplus gap within a union between the woman with characteristics y and her partner with characteristics x is a separable function in y and x it is equal to the opposite of the logarithm of the ratio between the absolute probability density of the woman to be in the set of single women with characteristics y and the absolute probability density of the man to be in the set of of single men with characteristics x. A consequence of this is the fact that the one who has the greater individual surplus within the couple is the partner who has the lower absolute probability density to be single in the population between the two, in other terms, the happier within the couple is the one whose characteristics' group has the fewer number of singles between the two, the more attractive then between the two partners. This is explained by the fact that a high attractiveness implies a high demand and then the partner who is the less attractive within the couple transfers utility to the more attractive (and diminish then his/her own utility) to be competitive regard to the concurrency. To illustrate this, we can consider some rural African villages in which only men own agricultural lands and herds of cattle. The resources are then quasi exclusively owned by the male population. To live, a woman needs to be in a safe household and that it is a security for her to not be in the need. Men are then the more attractive on this market. Wealthy men who have for example several lands or numerous cattle, have almost naturally more than one spouse. The demand is high for them that some families who want to built an alliance with rich men can even offer them one of their young girls. This typical example is written in the novel of the African writer Seydou Badian (*Sous l'orage*, 1957) in which a family was obliging their young and intelligent daughter even educated and motivated for long studies to stop her studies and marry the richest man of the village, a polygamous old man; but it won't be correct to spoil you on this interesting novel. So, women in such societies are somehow in a situation of resignation at the equilibrium of the market and men are clearly the winners of the game.

3.3.3 Income inequality and impact of marriage

There exists several studies on the measurement of Income Inequality : Pareto (1896), Hoover (1936), Atkinson (1970), Theil (1979), Palma (2011) who worked on the distribution of income inequality across the population. The inequality of composition of income is studied by Kaldor (1955), Foster and Wolfson (1992), Esteban and Ray (1991), Duclos and Taptue (2015), Araar (2008), Deutsch and Silber (2010) and Deutsch, Fusco, and Silber (2013), Milanovic (2017), Ranaldi (2017, 2018) with the introduction of the notion of polarization between the different income sources. Ranaldi (2018) proposes a measure of income composition inequality called Income Factor Polarization Index, a metric that summarizes the polarization of the two income sources across the whole population. Linking inequality and marriage is a subject discussed by many researchers in demography and economics. The unitary marriage model of Becker (1975) does not allow to get an idea about the impact of marriage on the welfare inequality. Chiappori (1988) has proposed a non unitary model of marriage and showed a welfare inequality with the household. Young (1952) had also noticed an inequality in consumption linked to the income inequality between spouses. Browning and alii (1994), using Canadian data showed the existence of inequality in clothing explained by the income inequality between spouses. Bonke (2015) empirically proved that the unequal repartition on consumption is correlated to the unequal distribution of the income with the household. Another interesting study is done by Pahl (1989) in which he considers different type of whole wage system that can be collective or individualized. Analogous studies are proposed by Laporte and Schellenberg (2011), Belleau and Lobet (2017), Ashby and Burgoyne (2008) and Belleau and Proulx (2011). The estimations of welfare in these systems shows an inequality of welfare (Phipps and Burton, 1995) in the advantage of men. Borooah et McKee (1994) proved that the income inequality is higher in an unequal sharing of income than in an equal between spouses. Fritzell (1999) finds a linear positive relation between the income distribution and the welfare distribution. Davies and Joshi (1994) Woolley and Marshall (1994) suggest the existence of relation between income inequality and welfare inequality. Stephane Crespo (2017) shows the impact of the wage sharing within the household on the inequality measurement. All these researches prove the existence of inequality of welfare within households, and this inequality is higher when the system of wage sharing is unequal. This suggests that marriage impacts the global welfare inequality because the inequality of income between two spouses (collective sharing system) is fewer than if they were not matched (individualized sharing system). To measure the impact of marriage on the income inequality, we consider that the marriage surplus must be taken in account in the individual income.

Before focusing on the income inequality and how it is impacted by marriage, we need to introduce first an important notion. Let m be a man with income R^M_m and w a woman with income R^W_w. We will intuitively consider as joint welfare of a couple, the sum of the income of both spouses with the joint net utility from marriage. The net utility from marriage is the difference between marriage utility and utility when remaining single. This intuitive definition has the advantage to maintain the welfare equal to the income for single individuals. The limit of this definition is simply the fact that we are summing a quantity (income) with a well defined unit to another quantity (utility) with a less defined unit. We must take into account this difference of unit. This leads us to adopt formally the following definition. We denote respectively by Z_m^M, Z_w^W the welfare of the man m and the welfare of the woman w. We assume the existence of a constant λ such that:

$$Z_m^M = R_m^M + \lambda(u_m - \varepsilon_m^0) \text{ and } Z_w^W = R_w^W + \lambda(v_w - \eta_w^0)$$

The terms $u - \varepsilon^0$ and $v - \eta^0$ represent respectively the net utility of men and the net utility of women. We can remark that one implication of this definition is that, when the man m and the woman w are single, then their welfare is equal to their income since the net utility of singles is zero. We then have for singles:

$$Z_m^M = R_m^M \text{ and } Z_w^W = R_w^W$$

Value of Marriage The constant λ can be interpreted as the exchange rate between the income and utility. In other words, $1/\lambda$ is somehow the value in monetary unit of the net utility from marriage. Its evolution will indicate how the value of marriage has changed over the period of our study. We denote:

$$E^M = u - \varepsilon^0, \;\; E^W = v - \eta^0$$

The variable E^M is the net marriage utility of men and E^W is the net marriage utility of women. These two variables are non negative random variables. We then define two variables R and E as follows:

$$R = R^M \mathbb{1}_{\mathcal{M}} + R^W \mathbb{1}_{\mathcal{W}} \text{ and } E = E^M \mathbb{1}_{\mathcal{M}} + E^W \mathbb{1}_{\mathcal{W}}$$

where $\mathbb{1}_{\mathcal{M}}$ is the indication to belong to the male population, and $\mathbb{1}_{\mathcal{W}}$ is the indication to belong to the female population.

We also define on the population, the variable F that indicates the side of the market individuals belong to. Formally, F is such that for any individual k:

$$F_k = 1 \text{ if and only if } k \in \mathcal{W} \text{ and } F_k = 0 \text{ if and only if } k \in \mathcal{M}$$

Determination of λ : The constant λ can be determined with the following approach. Let $\mathcal{G}$ be a function evaluating the social inequality of any non negative random vector observable on the population. We can take for $\mathcal{G}$ the Gini index or its extension to the multidimensional case. Note that a such function $\mathcal{G}$ induces an ordering relation among random vectors or variables observed on a particular population, and an ordering relation among populations on which a particular random vector or variable is observed. We can immediately remark that we can define a relative equivalence relation on the set of non negative random vectors or on variables observed on a particular population. Given a population $\mathcal{P}$, two non negative random vectors will be considered as equivalent if they have the same inequality on $\mathcal{P}$. On the whole market $\mathcal{P}$, we assume the vector (R, E) be equivalent to the vector $Z = R + \lambda E$ by the ordering $\mathcal{G}$. In other terms, they are assumed to satisfy the equation:

$$\mathcal{G}(R, E) = \mathcal{G}(R + \lambda E)$$

This equation guarantees somehow the substitution between income and net utility from marriage such that the total social welfare inequality is maintained unchanged. The weakness of this approach is its arbitrariness. It will just allow us to measure changes over the time in terms of inequality. We will then compare the inequality within married people to the inequality within single people. The Gini index has the advantage to be the common measure of inequality. For its two-dimensional version, we will consider its extension proposed by G. A. Koshevoy and Mosler (1997). for non negative multivariate distributions. We give their definition below. We suggest in Lawogni (2020, *Measures of Inequality in Vectors Distributions*, International Journal of Statistical Analysis) a generalization of the Gini index and a slight extension of the Pietra index for multivariate distributions in which the coordinates can be negative or positive where the chosen distance is not necessarily the norm N_1. Although, we will simply consider the definition of G. A. Koshevoy and Mosler for the coordinates of the vector (R, E) are non negative.

Definition 3.2. *(G. A. Koshevoy and Mosler, 1997). Let* $X = (X^1, ..., X^d)$ *be a random vector in* $\mathbb{R}^d_+$ *and independently observed on* n *individuals constituting a population* $\mathcal{P}$ *with* $X_i = (X^1_1, ..., X^d_1)$ *is the observation on the individual* i*, and* $X^s = (X^s_1, ..., X^s_n)$

is the list of the s-th coordinate of X observed on the n individuals. The Gini index of X is given by:

$$\mathcal{G}(X) = \frac{1}{2d \times n^2} \sum_{j=1}^{n} \sum_{i=1}^{n} \left(\sum_{s=1}^{d} \frac{(x_i^s - x_j^s)^2}{(\bar{x^s})^2} \right)^{\frac{1}{2}}$$

where $\bar{x^s}$ is the average of x^s.

In this definition, d is the dimension of the vector X, its i-th observation is denoted X_i and its s-th coordinate is X^s. Note that the net marriage utility E is non negative. The income R is observable contrarily to the marriage net utilities that is not observable. Let μ be the number of couples in the population. We denote by $\mathcal{M}a$ the set of matched people. In our case, we have: $d = 2$ because the vector (R, E) has two dimensions, $n = 2 \times \mu$ is the number of matched people. Then, following the given definition, the index $\mathcal{G}$ takes here the form:

$$\mathcal{G}_{\mathcal{M}a}(Z) = \frac{1}{16\mu^2} \sum_{j \in \mathcal{M}a} \sum_{i \in \mathcal{M}a} \left(\left(\frac{R_i - R_j}{\bar{R}} \right)^2 + \left(\frac{E_i - E_j}{\bar{E}} \right)^2 \right)^{\frac{1}{2}}$$

As we don't observe the term $\left(\frac{E_i - E_j}{\bar{E}}\right)^2$, we can estimate it by:

$$\frac{\hat{E^2}_i + \hat{E^2}_j - 2\hat{E}_i\hat{E}_j}{(\mathbb{E}[E])^2}$$

We estimate E by its conditional expectation with respect to the income R, to the gender and to the marriage. We do same for E^2.

$$\hat{E} = \mathbb{E}[E \,|\, R, M = 1, F], \text{ and } \hat{E^2} = \mathbb{E}[E^2 \,|\, R, M = 1, F]$$

Note that:

$$\mathbb{E}[E \,|\, R, M = 1, F = 0] = \mathbb{E}[u - \varepsilon^0 \,|\, R = R^M, M = 1]$$

and

$$\mathbb{E}[E \,|\, R, M = 1, F = 1] = \mathbb{E}[v - \eta^0 \,|\, R = R^W, M = 1]$$

Analogously, we have:

$$\mathbb{E}[E^2 \,|\, R, M = 1, F = 0] = \mathbb{E}[(u - \varepsilon^0)^2 \,|\, R = R^M, M = 1]$$

and

$$\mathbb{E}[E^2 \,|\, R, M = 1, F = 1] = \mathbb{E}[(v - \eta^0)^2 \,|\, R = R^W, M = 1]$$

To get theoretically the different expectations above, we need to derive the distribution of the net utilities $u - \varepsilon^0$ and $v - \eta^0$. For more details, you can refer to the appendix. The index is then estimated by:

$$\hat{\mathcal{G}}_{\mathcal{M}a}(Z) = \frac{1}{16\mu^2} \sum_{j \in \mathcal{M}a} \sum_{i \in \mathcal{M}a} \left(\left(\frac{R_i - R_j}{\bar{R}} \right)^2 + \left(\frac{\hat{E}_i^2 + \hat{E}_j^2 - 2\hat{E}_i \hat{E}_j}{(\mathbb{E}[E])^2} \right) \right)^{\frac{1}{2}}$$

The characteristics on which the marriage surpluses are defined respectively for men and for women as follows:

$$X = \frac{\ln(R^M) - \mathbb{E}(\ln(R^M))}{\sqrt{\mathbb{V}(\ln(R^M))}} \quad \text{and} \quad Y = \frac{\ln(R^W) - \mathbb{E}(\ln(R^W))}{\sqrt{\mathbb{V}(\ln(R^W))}}$$

These variables R^M and R^W are the hourly incomes. They are proportional to the incomes of full-time employment.

3.4 Data and Empirical Results

3.4.1 Data

We use the PSID data on the period 1968-2001. We exclude from the data under 17 years old individuals. Our main variable is the income per hour. As we dealt with missing data on the income, we used the Heckman method to estimate the unobserved income on some individuals as the decision of unemployment can be partially explained by some observable determinants such as the age, the education or the sex. We estimate by the Heckman method of selection the income per hour for all the individuals by using as variables explaining the selection, the age, the square of age, the sex, and the education and these variables are present in the data basis. The hourly income will be the variable on which we will compute the gain from marriage. This variable is somehow the income of full worked time (by normalizing the worked time to 1 for all the individuals).

3.4.2 Descriptive Statistics

We present below some descriptive statistics. The following figures show the size of the data we used.The figure 21 indicates the numbers of men and of women in the data for each year of the period of the study. We observe that the number of men is lower than the number of women in all the data for every year of the period of the study. and the numbers of married people and of single people. The figure 22 precises the number of singles and married individuals in the different data for each year of the period of the study.

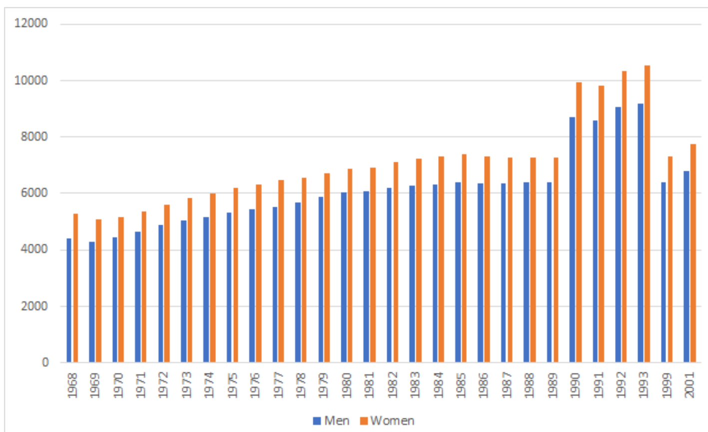


Figure 21: Size of the data by gender

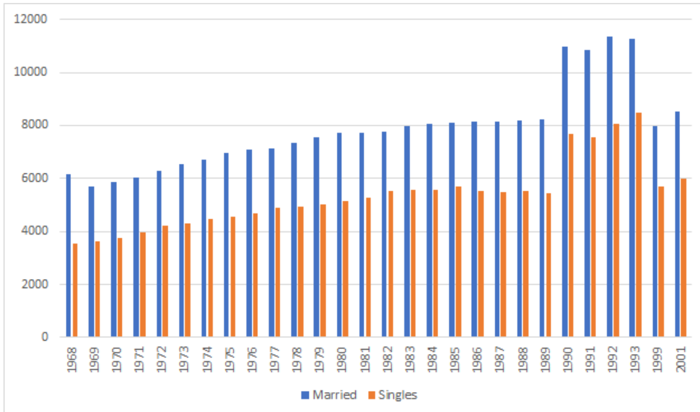


Figure 22: Size of the data by marital status

The number of singles in the data is lower than the number of married people. The figure 23 shows the evolution of the rate of marriage on this market over the period 1968-2001.

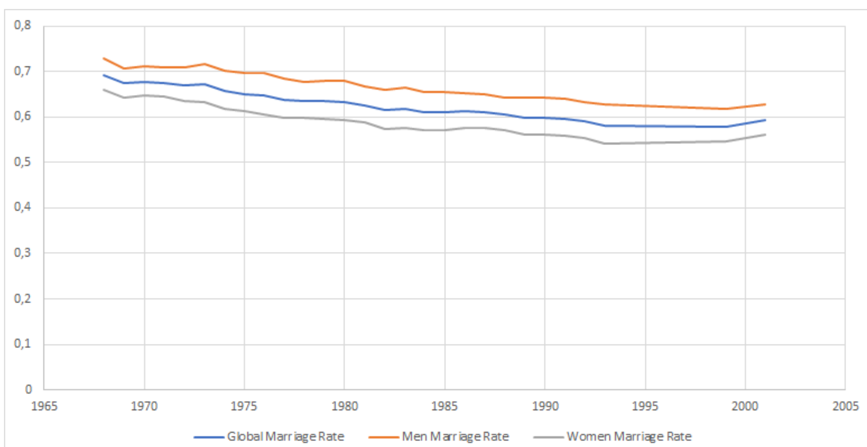


Figure 23: Evolution of the rates of marriage

As we can clearly remark it, the rate of marriage has drastically decreased over the period of the study.

The following figure shows the evolution of the estimated averaged income per hour conditionally to the gender and conditionally to the sex.

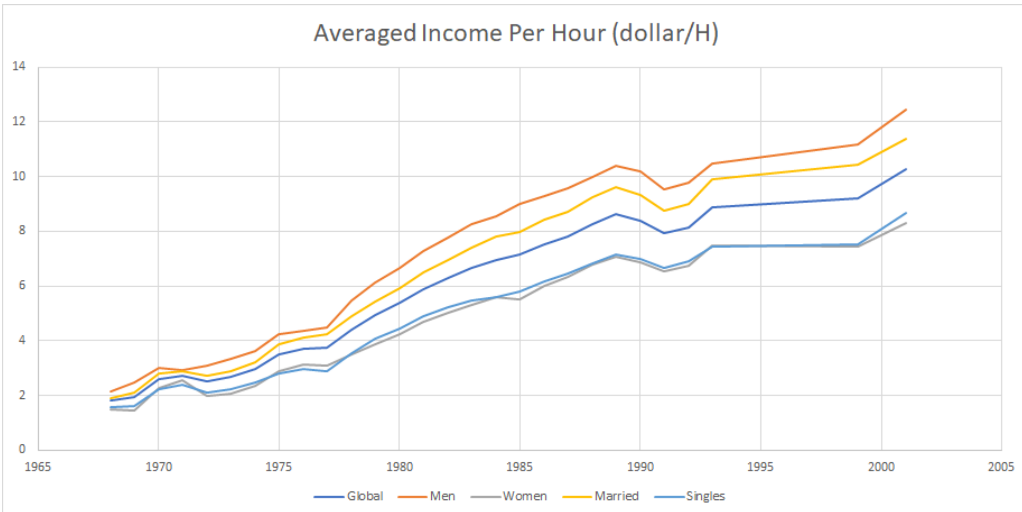


Figure 24: Evolution of the averaged hourly income (in dollar/hour)

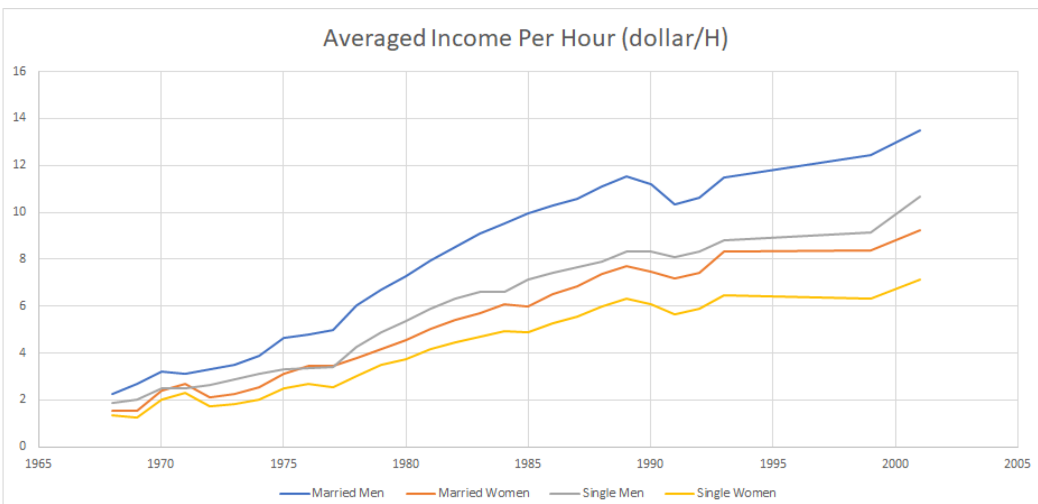


Figure 25: Evolution of the averaged hourly income (in dollar/hour)

We remark a global increase of the hourly income over the period of the study. The averaged hourly income of men is greater than the averaged hourly income of women, and the averaged hourly income of married individuals is greater than the averaged income of singles. The greatest hourly incomes are earned by married men and the lowest by single women. We can also point out that the hourly income of single men is higher than the hourly income of married women. A global observation that married individuals earn higher hourly income than single individuals.

Income Inequality In what follows, we describe the hourly income inequality by gender and by marital status. We observe a general increase of the inequalities on the period 1970-1999 and a decline from 1999 to 2001. We observe the highest inequality on the singles over the whole period. The men hourly income inequality is greater than the women' on the period 1968-1990 and then the women hourly income inequality becomes more important than the men' from 1990 to 2001. The hourly income inequality of singles is greater than the hourly income inequality of married. The absolute gap in inequality between men and women is extremely lower than the gap in equality between singles and married. This suggests us a reduction of the inequality by marriage. This is confirmed the figure 27.

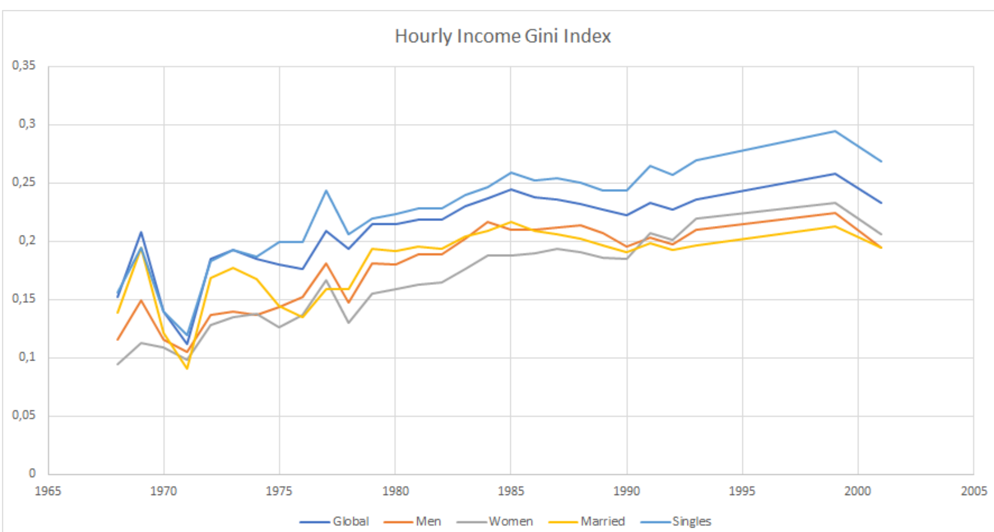


Figure 26: Evolution of the averaged hourly income Gini Index (in dollar/hour)

The hourly income inequality of married men is lower than the hourly income inequality of single men and same for women, the inequality of the hourly income of married women is lower than for single women. We also remark that there more inequality in the married men hourly income than in the married women hourly income and this over all the period of the study. For singles, we remark that on the periods 1970-1974, and 1990 2001 the hourly income inequality of single women is higher than for single men whereas inversely on the periods 1968-1970 and 1974-1990 the hourly income inequality of single women is lower than for single men.

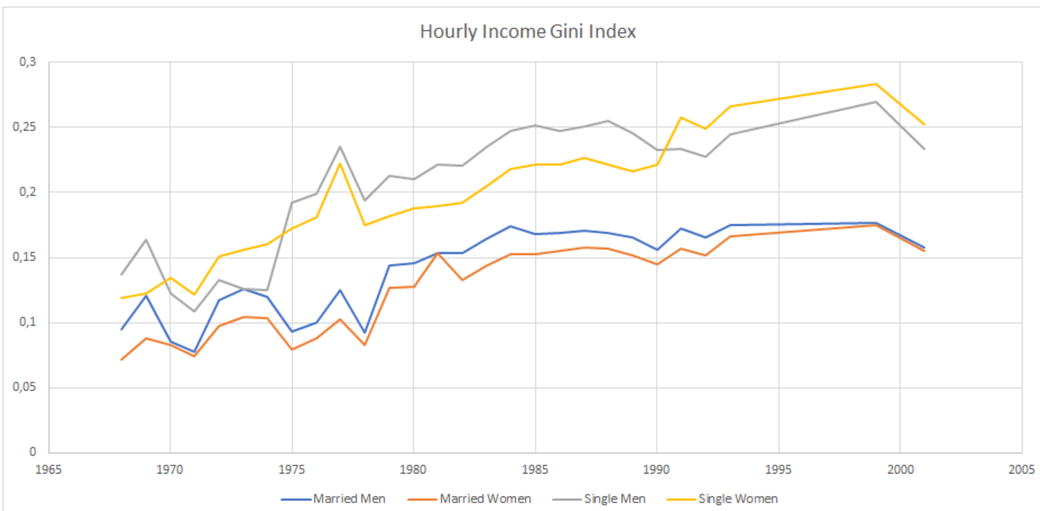


Figure 27: Evolution of the averaged hourly income Gini Index (in dollar/hour)

3.4.3 Empirical Results

The parameters and their evolution over the time We estimate the model. We recall the specification of the joint surplus:

$$S(x, y) = -a_2^2 x^2 + c\, x\, y - b_2^2 y^2 + a_1\, x + b_1\, y$$

where x and y are the normalization of the logarithm of the incomes of men and of women. The figure 28 shows the evolution of the parameters over the time.

We can mainly remark the positivity of the coefficient c that is:

$$c = \frac{\partial^2 S}{\partial x \partial y}(x, y) > 0$$

That shows a positive assortative mating between men and women with respect to the income. We can notice a relative stability of this parameter c on the period 1968-1987 and a general increase on the period 1987-2001. Other evidence is the fact the stability of the parameters a_2^2 and b_2^2 over the period of the study and we have $a_2^2 < b_2^2$. We observe a global stability for the parameters a_1 and b_1 too. To compare the effects of x and y on the joint surplus, we will focus on the sign of the first derivatives of $S(x, y)$:

$$\frac{\partial S}{\partial x}(x, y) = -2a_2^2 x + cy + a_1 \text{ and } \frac{\partial S}{\partial y}(x, y) = -2b_2^2 y + cx + b_1$$

Let's assume a couple with characteristics (x, y). We have:

$$\frac{\partial S}{\partial x}(x, y) > 0 \iff x < \frac{c}{2\,a_2^2} y + \frac{a_1}{2\,a_2^2} := x^*(y)$$

and

$$\frac{\partial S}{\partial y}(x, y) > 0 \iff y < \frac{c}{2\,b_2^2} x + \frac{b_1}{2\,b_2^2} := y^*(x)$$

Note that $x^*(y)$ and $y^*(x)$ are actually not the optimal corresponding characteristics of partner respectively for the woman and for the man. They are optimal only if the sympathy shocks are negligible. From the estimation, we observe that $c > 3$, $a_2^2 < 1.5$ and $b_2^2 > 1.5$, so $\frac{c}{2\,a_2^2} > 1$ and $\frac{c}{2\,b_2^2} < 1$. The term $\frac{a_1}{2\,a_2^2}$ is positive so $x^*(y) > y$.

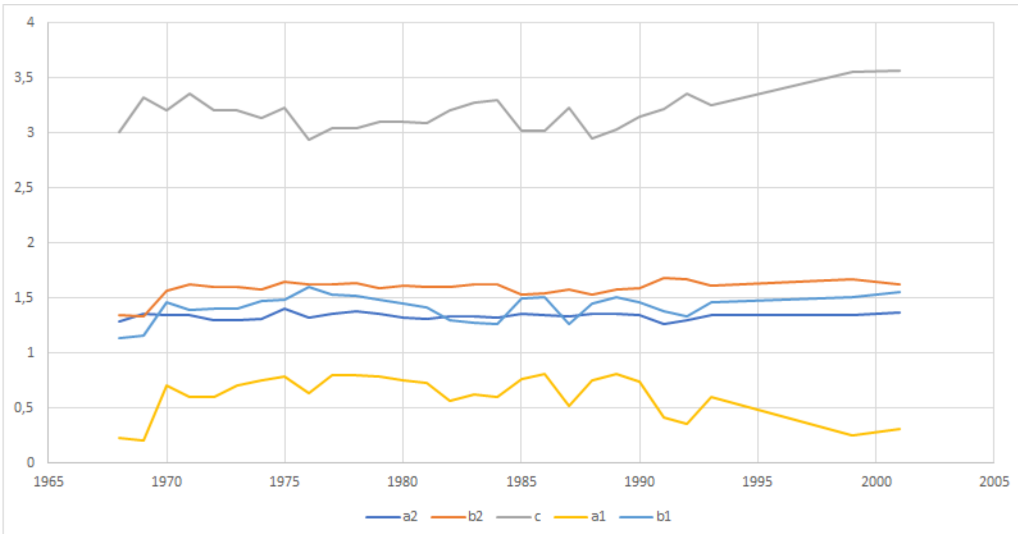


Figure 28: Evolution of the parameters of the model

Evolution of the inequality of net utility from marriage We have observed that there is less income inequality for married individuals than for singles. As the formation of unions generates a net utility for matched individuals with respect to singles, we must consider too the inequality of this net utility. Note that the net utility is positive quantity since it is the difference between the obtained utility at the equilibrium and the utility from remaining single. The estimation shows the presence of inequality of the net utility from marriage. The following graph shows its evolution over the period of the study.

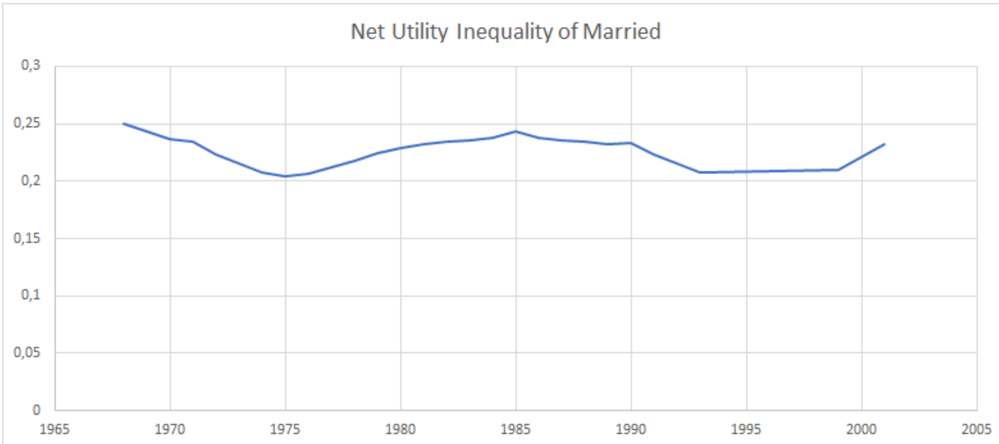


Figure 29: Evolution of the inequality of net utility

Sharing of the joint net utility The estimation revealed that the sharing of the net joint utility from marriage is very similar to the weights of the incomes. The following graph shows the evolution of the weight of the income of married men in the total income of married individuals and also the evolution of the weight of their net utility in the total net utility of married individuals. And we can remark a very similar evolution of the two proportions. We notice on the period 1968-1975 an increase of the part of married men in the total net utility and a very little decline on the period 1975-2001. The part of men in the total net utility is slightly higher than their part in the total income and represents about 60% of the total net utility. The fact that the part of net utility is close to the part of income confirms the idea that the reduction of the gender gap in income leads to a more egalitarian distribution of the consumption within household since the consumption can quantify utility from marriage. This will be done one of the next chapters.

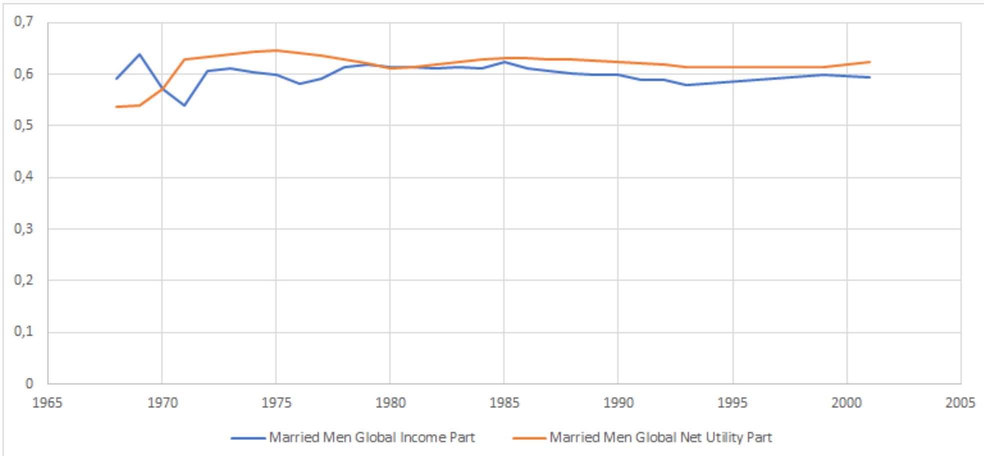


Figure 30: Evolution of the men' parts of income and net utility

Evolution of the value of marriage: $1/\lambda$ The conversion rate between the net utility and income is $1/\lambda$. It represents somehow the value of marriage. The following graph shows its evolution over the period of the study. We can remark an increase of the value of marriage on the period 1968-1971 and a general drastic decrease on the period 1971-1985 and has been reltively stable then on the period 1985-2001.

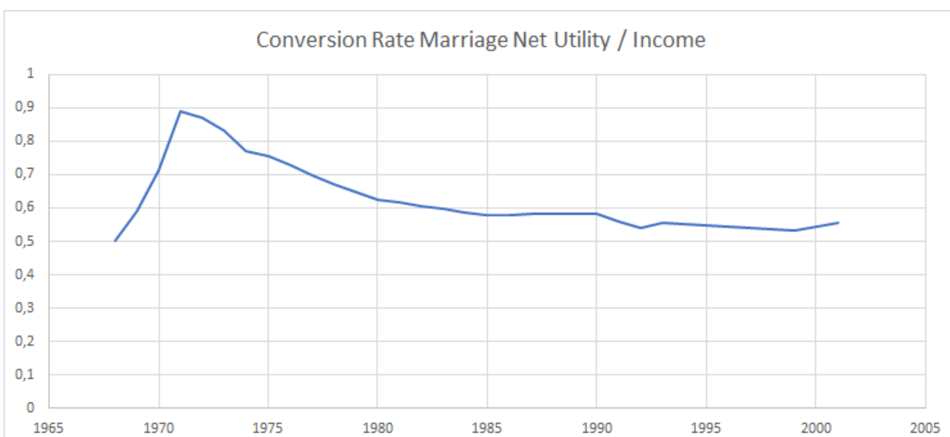


Figure 31: Evolution of the value of Marriage:$1/\lambda$

Impact of marriage on the income inequality We have observed that the income inequality of married is lower than the income inequality of singles. But before really concluding into a reduction of inequality by marriage we must take into account the existence of net utility inequality generated by marriage. That is why, to measure the impact of marriage on the income inequality, we have computed the inequality index of the income on the population of married people and on the population of the singles. And we have also computed the inequality index of the refined income $Z = Income + \lambda\, Net\, Utility$ on the population of married people taking into account the marriage surplus. We graph the evolution of these inequality indexes over the time.

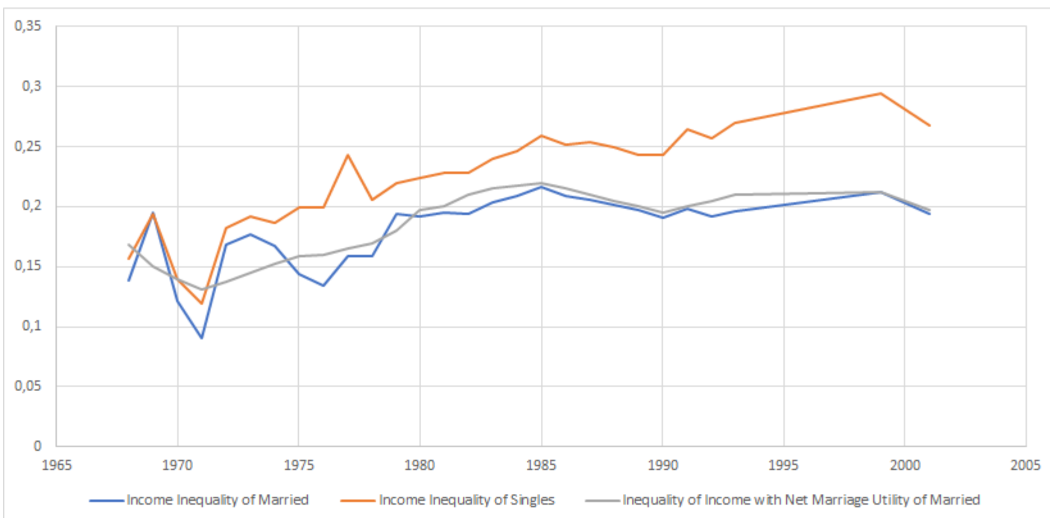


Figure 32: Impact of Marriage on the Inequality

We can remark that the inequality of the income on the population of the married people is lower than the inequality of the income on the population of the singles. The integration of the net utility into the income does slightly increases the inequality on the population of married individuals because of the the inequality of the net utility but lets the inequality be still lower to the inequality income of single individuals. We can then affirm that marriage reduces inequality relatively to singles.

3.5 Conclusion

Dupuy and Galichon (2014) have proposed a continuous model extending the Choo and Siow's model. In the initial formulation of the Dupuy and Galichon's model, single individuals are ignored, i.e., all individuals are assumed to be married. In this paper, in the light of the Dupuy and Galichon's model, we propose a slightly different extension of the Choo and Siow's model to the continuous case with perfectly transferable utility and taking into account singles. This model is empirically tractable. We have applied it to measure changes in utility inequality. The study shows that there is more income inequality among single individuals than among married people. The income inequality has globally increased. To measure the impact of marriage on the income inequality, in the computation of the inequality, we take into account the marriage surplus and its inequality. We still obtain a lower inequality for married than the income inequality of singles. This shows somehow a reduction of inequality by marriage. This confirms a positive impact of marriage on welfare inequality. This inequality index has slightly increased on the period of the study. We estimated too in this study the conversion rate between marriage surplus and income to measure a monetary value of marriage. We observed an increase of this value on the period 1968-1971, a significant decrease on the period 1971-1985, and has been relatively stable on the period 1985-2001.

References

[1] Araar. 2008, On the decomposition of polarization indices: Illustrations with Chinese and Nigerian household surveys, CIRPEE Working Paper No. 08-06

[2] Ashby & Burgoyne. 2008, Separate financial entities? : Beyond categories of money management. The Journal of Socio-Economics. Vol. 37. No 2. p. 458-480.

[3] Atkinson. 1970, On the Measurement of Inequality, Journal of Economic Theory, Vol. 2, Issue 3, p 244-263

[4] Badian, Seydou. 1957, Sous l'Orage, Edition Présence Africaine

[5] Bargain, Olivier & Damien Echevin & Nicolas Moreau & Adrien Pacifico. 2019, "Inefficient Couples: Non-minimization of the Tax Burden among French Cohabiting Couples," Working Papers hal-02365239, HAL.

[6] Becker, G. 1973, A Theory of Marriage: Part I, The Journal of Political Economy, Vol. 81, No. 4, pp. 813-846

[7] Becker, G. 1974, A Theory of Marriage: Part II, The Journal of Political Economy, vol. 82, issue 2, S11-S26

[8] Belleau & Lobet. 2017, L'amour et l'argent, Guide de survie en 60 questions, Montréal, Les éditions du remue-ménage, 2017, p. 57-59

[9] Belleau & Proulx. 2011, Le revenu familial, un concept vague et insidieux : analyse critique et historique des relations économiques familiales, Revue Enfances, Familles, Générations Numéro 15, p. 78109

[10] Bergstrom, Theodore & Richard Cornes. 1983, Independence of Allocative Efficiency from Distribution in the Theory of Public Goods, Econometrica, 51, 1753-1765

[11] Bonke. 2015, Lifetime Earnings Inequality in Germany, Journal of Labor Economics, vol. 33, issue 1, p 171 - 208

[12] Borooah. V. K. & P. M. Mckee. 1994, Intra-household income transfers and implications for poverty and inequality in the UK, in J. Creedy (dir.), Taxation, Poverty and Income Distribution, Aldershot, Edward Elgar: p 69-86

[13] Browning & Bourguignon & Chiappori. 1994, Income and Outcomes: A Structural Model of Intrahousehold Allocation, Journal of Political Economy, 1994, vol. 102, issue 6, 1067-96

[14] Burtless. 1999, Effects of growing wage disparities and changing family composition on the U.S. income distribution, European Economic Review, vol. 43, issue 4-6, 853-865

[15] Chiappori, Pierre André. 1988, Rational household labor supply, Econometrica, Vol. 56, No. 1, pp. 63-90

[16] Chiappori, Salanié & Weiss. 2017, Partner Choice and the Marital College Premium, American Economic Review, vol. 107, no. 8, pp. 2109-67

[17] Choo, Eugene & Aloysius Siow. 2006, Who marries whom and why, Journal of Political Economy, Vol. 114, No. 1, pp. 175-201

[18] Crespo. 2017, Et si les conjoints ne partageaient pas tous leurs revenus ? Conséquences sur la mesure des inégalités du bien-être, Cahiers québécois de démographie

[19] Davies & Joshi. 1994, The foregone earnings of Europe's mothers?, Standards of Living and Families : Observation and Analysis, Congress Colloquia, pp.101-135

[20] Deutsch, J., Fusco, A. & Silber, J. 2013, The BIP Trilogy (Bipolarization, Inequality and Polarization): One Saga but Three Different Stories. Economics: The Open-Access, Open-Assessment, E-Journal, 7 (2013-22): 133.

[21] Deutsch, J. & Silber, J. 2010, Analyzing the Impact of Income Sources on Changes in BiPolarization in The Measurement of Individual Well-Being and Group Inequalities: Essays in Memory of Z. M. Berrebi, edited by Deutsch, J., and Silber, J., Routledge Economics, Taylor and Francis Group, pp. 127-152

[22] Duclos & Taptue. 2015, Polarization, Handbook of Income Distribution, Vol. 2A.

[23] Dupuy, Arnaud & Alfred Galichon, 2014. "Personality Traits and the Marriage Market," Journal of Political Economy, University of Chicago Press, vol. 122(6), p 1271-1319.

[24] Esteban, J. & Ray, D. 1991, On the Measurement of Polarization, Boston University, Institute for Economic Development, Working Paper

[25] Fernandez, Raquel, Alessandra Fogli & Claudia Olivetti, Mothers and Sons. 2004, Preference Formation and Female Labor Force Dynamics, Quarterly Journal of Economics, vol. 119, no 4, pp. 124999

[26] Foster & Wolfson. 1992, Polarization and the Decline of the Middle Class: Canada and the US, OPHI WORKING PAPER NO. 31

[27] Gale & Shapley. 1962, College Admissions and the Stability of Marriage, Mathematical Association of America, Vol. 69, No. 1, pp. 9-15

[28] Galichon, Alfred & Bernard Salanié. 2012, Cupid's invisible hand: social surplus and identification in matching models, Sciences Po publications

[29] Gottschalk, Peter & Susan Mayer. 2002, Changes in Home Production and Trends in Economic Inequality. In Daniel Cohen, Thomas Piketty, and Gilles Saint-Paul (eds), The New Economics of Rising Inequality. Oxford: Oxford University Press, 265-284.

[30] Greenwood, Guner, Kocharkov & Santos. 2014, Marry Your Like: Assortative Mating and Income Inequality, AMERICAN ECONOMIC REVIEW VOL. 104, NO. 5, pp. 348-53

[31] Hoover Jr. 1936, The Measurement of Industrial Localization, Review of Economics and Statistics, Vol. 18, No. 4, pp. 162-171

[32] Kaldor. 1955, Alternative Theories of Distribution, The Review of Economic Studies, Vol. 23, No. 2 (1955 - 1956), pp. 83-100

[33] Koshevoy & Mosler. 1997, Multivariate Gini Indices, Journal of Multivariate Analysis, vol. 60, issue 2, p 252-276

[34] Laporte & Schellenberg. 2011, The Income Management Strategies of Older Couples in Canada, Statistics Canada Analytical Branch Studies Working Paper No. 335

[35] Lawogni, Ulysse. 2020, Measures of Inequality in Vectors Distributions, International Journal of Statistical Analysis. V1(3): 1-16.

[36] Lise, Jeremy & Shannon Seitz. 2011, Consumption Inequality and Intra-household Allocations, Review of Economic Studies, vol. 78, pp. 328-355.

[37] Martin. 2006, Practising Gender at Work: Further Thoughts on Reflexivity, Wiley Online Library, Gender, Work and Organization, Volume 13, Issue 3, Pages 254-276

[38] Milanovic, B. 2017, Increasing Capital Income Share and its Effect on Personal Income Inequality. In After Piketty, The Agenda for Economics and Inequality, Edited by Heather Boushey, J.Bradford DeLong, Marshall Steinbaum

[39] Pahl. 1989, Housing, Work and Life Style, Wiley Online Library, Volume 80, Issue 2 Pages 75-87

[40] Palma. 2011, Homogeneous Middles vs. Heterogeneous Tails, and the End of the "Inverted U": It's All About the Share of the Rich, Development and Change, 42(1) p 87 - 153

[41] Pareto. 1896, Cours d'Economie Politique, Collection : Travaux de Sciences Sociales, Editeur : Librairie Droz

[42] Phipps, S. A. & P. S. Burton. 1995, Sharing within families : implications for the measurement of poverty among individuals in Canada, Canadian Journal of Economics, vol. 28, issue 1, p 177-204

[43] Ranaldi, M. 2017, Income-Factor Polarization: A Methodological Approach, Documents de travail du Centre d'Economie de la Sorbonne

[44] Ranaldi, M. 2018, On the Measurement of Functional Income Distribution, Documents de Travail du Centre d'Economie de la Sorbonne

[45] Shapley & Shubik. 1971, The assignment game I: The core, International Journal of Game Theory volume 1, pp 111-130

[46] Schwartz. 2010, Earnings Inequality and the Changing Association Between Spouses' Earnings, American Journal of Sociology, vol. 115, pp. 1524-1557

[47] Theil. 1979, The Measurement of Inequity by Components of Income, Economics Letters

[48] Young. 1952, Distribution of income within the family, British journal of sociology, Vol. 3, No. 4, pp. 305-321

3.6 Appendix

3.6.1 Proof of Theorem 3.1

Expression of the joint surplus

From the first order conditions, we derive the relations:

$$e^{\frac{\bar{U}(x,y)}{\sigma}} = \frac{\alpha}{1-\alpha}\frac{\pi(x,y)}{f_0(x)} \text{ and } e^{\frac{\bar{V}(x,y)}{\tau}} = \frac{\beta}{1-\beta}\frac{\pi(x,y)}{g_0(y)}$$

This leads to:

$$\bar{U}(x,y) = \sigma \ln\left(\frac{\alpha}{1-\alpha}\frac{\pi(x,y)}{f_0(x)}\right), \text{ and } \bar{V}(x,y) = \tau \ln\left(\frac{\beta}{1-\beta}\frac{\pi(x,y)}{g_0(y)}\right)$$

The sum of the two equations gives:

$$\bar{S}(x,y) = \ln\left(\left(\frac{\alpha}{1-\alpha}\right)^{\sigma}\left(\frac{\beta}{1-\beta}\right)^{\tau}\frac{(\pi(x,y))^{\sigma+\tau}}{(f_0(x))^{\sigma}(g_0(y))^{\tau}}\right)$$

Expressions of U and V

To derive the expressions of $U(x,y)$ and of $V(x,y)$, we recall:

$$e^{\frac{\bar{U}(x,y)}{\sigma}} = \frac{\alpha}{1-\alpha}\frac{\pi(x,y)}{f_0(x)} \text{ and } e^{\frac{\bar{V}(x,y)}{\tau}} = \frac{\beta}{1-\beta}\frac{\pi(x,y)}{g_0(y)}$$

The ratio of these two expressions gives:

$$e^{\frac{\bar{U}(x,y)}{\sigma} - \frac{\bar{V}(x,y)}{\tau}} = \frac{\alpha}{(1-\alpha)f_0(x)}\frac{(1-\beta)g_0(y)}{\beta}$$

Then we get:

$$\tau\bar{U}(x,y) - \sigma\bar{V}(x,y) = -\sigma\tau\ln\left(\frac{1-\alpha}{\alpha}f_0(x)\right) + \sigma\tau\ln\left(\frac{1-\beta}{\beta}g_0(y)\right)$$

By using in addition $\bar{S}(x,y) = \bar{U}(x,y) + \bar{V}(x,y)$, we obtain:

$$\bar{U}(x,y) = \frac{\sigma}{\sigma+\tau}\left(\bar{S}(x,y) - \tau\ln\left(\frac{1-\alpha}{\alpha}f_0(x)\right) + \tau\ln\left(\frac{1-\beta}{\beta}g_0(y)\right)\right)$$

and

$$\bar{V}(x,y) = \frac{\tau}{\sigma+\tau}\left(\bar{S}(x,y) + \sigma\ln\left(\frac{1-\alpha}{\alpha}f_0(x)\right) - \sigma\ln\left(\frac{1-\beta}{\beta}g_0(y)\right)\right)$$

Expression of f_0 and f_1

$$e^{\frac{\bar{U}(x,y)}{\sigma}} = \frac{\alpha}{1-\alpha}\frac{\pi(x,y)}{f_0(x)}$$

This leads to:

$$\int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy = \frac{\alpha}{1-\alpha}\int_{\mathcal{Y}}\frac{\pi(x,y)}{f_0(x)}dy$$

So, we get:

$$f_0(x)\int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy = \frac{\alpha}{1-\alpha}\sigma f_1(x)$$

Since we know that $f(x) = \alpha f_1(x) + (1-\alpha)f_0(x)$, we directly derive:

$$f_0(x) = \frac{1}{(1-\alpha)\left(1+\int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy\right)} f(x)$$

and

$$f_1(x) = \frac{\int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy}{\alpha\left(1+\int_{\mathcal{Y}} e^{\frac{U(x,y)}{\sigma}} dy\right)} f(x)$$

An analogous reasoning gives the expressions of $g_0(y)$ and $g_1(y)$.

Expression of α and β

We use:

$$f_1(x) = \frac{\int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy}{\alpha \left(1 + \int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy\right)} f(x)$$

This implies:

$$\alpha \int_{\mathcal{X}} f_1(x)dx = \int_{\mathcal{X}} \frac{\int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy}{\left(1 + \int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy\right)} f(x)dx$$

Since $\int_{\mathcal{X}} f_1(x)dx = 1$, we immediately get:

$$\alpha = \int_{\mathcal{X}} \int_{\mathcal{Y}} \frac{e^{\frac{\bar{U}(x,y)}{\sigma}}}{1 + \int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy} f(x)dxdy$$

We find the expression of β by deduction.

$$\pi(x,y) = \frac{f(x)}{\alpha} \frac{e^{\frac{\bar{U}(x,y)}{\sigma}}}{1 + \int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy} = \frac{g(y)}{\beta} \frac{e^{\frac{\bar{V}(x,y)}{\tau}}}{1 + \int_{\mathcal{X}} e^{\frac{\bar{V}(x,y)}{\tau}} dx}$$

Therefore, we have:

$$\frac{f(x).e^{\frac{\bar{U}(x,y)}{\sigma}}}{1 + \int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy} = \alpha.\frac{g(y)}{\beta} \frac{e^{\frac{\bar{V}(x,y)}{\tau}}}{1 + \int_{\mathcal{X}} e^{\frac{\bar{V}(x,y)}{\tau}} dx}$$

Using the proved expression of α, we get:

$$\alpha = \int_{\mathcal{X}} \int_{\mathcal{Y}} \alpha.\frac{g(y)}{\beta} \frac{e^{\frac{\bar{V}(x,y)}{\tau}}}{1 + \int_{\mathcal{X}} e^{\frac{\bar{V}(x,y)}{\tau}} dx} dxdy$$

We immediately deduce the expression of β:

$$\beta = \int_{\mathcal{X}} \int_{\mathcal{Y}} g(y).\frac{e^{\frac{\bar{V}(x,y)}{\tau}}}{1 + \int_{\mathcal{X}} e^{\frac{\bar{V}(x,y)}{\tau}} dx} dxdy$$

3.6.2 Other approach of estimation: Likelihood Maximization with penalization

The approximation functions for $U(x,y)$ and $V(x,y)$ are quite huge. In this paragraph, we will discuss about an alternative approach to estimate the model. The main idea will be to specify as most as possible "correctly" the densities $f_0(x)$ and $g_0(y)$. We define two functions $h(x)$ and $\tilde{h}(y)$ such that:

$$f_0(x) = f(x).h(x) \quad \text{and} \quad g_0(y) = g(y).\tilde{h}(y)$$

We will investigate on a convenient specification for $h(x)$ and $\tilde{h}(y)$. The justification of this approach is based on the fact that we have shown that $U(x,y)$ and $V(x,y)$ can be decomposed at the form:

$$U(x,y) = \frac{1}{2}\left(S(x,y) - \ln(f(x)) + \ln(g(y)) + \ln\left(\frac{1-\nu}{\nu}\right) + A(U)(x,y)\right)$$

and

$$V(x,y) = \frac{1}{2}\left(S(x,y) + \ln(f(x)) - \ln(g(y)) - \ln\left(\frac{1-\nu}{\nu}\right) + B(V)(x,y)\right)$$

with $A(U) = -B(V)$ and $\|A(U)\|_\infty < 1$. So the function $A(U)(x,y)$ is bounded. Using the expression of $U(x,y)$ from the theorem and replacing $f_0(x)$ and $g_0(y)$ respectively by $f(x)h(x)$ and by $g(y)\tilde{h}(y)$, we obtain:

$$U(x,y) = \frac{1}{2}\left(S(x,y) - \ln(f(x)) + \ln(g(y)) - \ln((1-\alpha)h(x)) + \ln((1-\beta)\tilde{h}(y)) + \ln\left(\frac{\alpha}{\beta}\right)\right)$$

Let's remark that:

$$\ln\left(\frac{\alpha}{\beta}\right) = \ln\left(\frac{1-\nu}{\nu}\right)$$

The difference of these two expressions of $U(x,y)$ gives:

$$0 = A(U)(x,y) + \ln((1-\alpha)h(x)) - \ln((1-\beta)\tilde{h}(y))$$

i.e

$$A(U) = -\ln((1-\alpha)h(x)) + \ln((1-\beta)\tilde{h}(y))$$

We want to specify $A(U)(x,y)$ as the addition of a function of x with a function of y such that $A(U)$ is bounded. And as $|A(U)|$ is lower than 1, a flexible specification for $A(U)$ can sufficiently fit it. A necessary condition is the fact that the functions $h(x)$ and $\tilde{h}(y)$ are bounded too and are non negative because of the logarithm. So we specify $h(x)$ and $\tilde{h}(y)$ such that they are both bounded and non negative and such that $h(x)f(x)$ and $\tilde{h}(y)g(y)$ are integrable respectively on $\mathcal{X}$ and $\mathcal{Y}$. In other words, we must satisfy:

$$\forall\, x \in \mathcal{X},\; h(x) > 0,\; \max_{x\in\mathcal{X}} |h(x)| < \infty,\; \text{and } \int_{\mathcal{X}} h(x)f(x)dx = 1$$

and

$$\forall\, y \in \mathcal{X},\; \tilde{h}(y) > 0,\; \max_{y\in\mathcal{Y}} |\tilde{h}(y)| < \infty,\; \text{and } \int_{\mathcal{Y}} \tilde{h}(y)g(y)dy = 1$$

To provide convenient specifications for $h(x)$ and $\tilde{h}(y)$, we make the following assumption.

Assumption 3.1.

$$\forall\, a > 0,\, b > 0,\; \int_{\mathcal{X}} (f(x))^a dx < \infty \textit{ and } \int_{\mathcal{Y}} (g(y))^b dy < \infty$$

Under this assumption, the following forms satisfy the constraints stated above for $h(x)$ and $\tilde{h}(y)$.

$$h(x) = a_0 \left(a_1 + a_2.(f(x))^{a_3^2}\right)^2 \text{ and } \tilde{h}(y) = b_0 \left(b_1 + b_2.(g(y))^{b_3^2}\right)^2$$

We then deduce the specification of $f_0(x)$ and $g_0(y)$.

$$f_0(x) = a_0 \left(a_1 + a_2.(f(x))^{a_3^2}\right)^2 f(x) \text{ and } g_0(y) = b_0 \left(b_1 + b_2.(g(y))^{b_3^2}\right)^2 g(y)$$

To satisfy the constraint:

$$\int_{\mathcal{X}} f_0(x)dx = 1 \quad\text{and}\quad \int_{\mathcal{Y}} g_0(y)dy = 1,$$

the parameters a_0 and b_0 are positive and given by:

$$a_0 = \frac{1}{\int_{\mathcal{X}} \left(a_1 + a_2.(f(x))^{a_3^2}\right)^2 f(x)dx} \quad\text{and}\quad b_0 = \frac{1}{\int_{\mathcal{Y}} \left(b_1 + b_2.(g(y))^{b_3^2}\right)^2 g(y)dy}$$

In the particular case we have:

$$f(x) = \frac{1}{2\pi}e^{-\frac{1}{2}x^2} \quad \text{and} \quad g(y) = \frac{1}{2\pi}e^{-\frac{1}{2}y^2}$$

The specification of $f_0(x)$ and $g_0(y)$ becomes then:

$$f_0(x) = a_0\left(a_1 + a_2.e^{-a_3^2x^2}\right)^2.e^{-\frac{1}{2}x^2} \text{ and } g_0(y) = b_0\left(b_1 + b_2.e^{-b_3^2x^2}\right)^2.e^{-\frac{1}{2}y^2}$$

and that is equivalent to the form:

$$f_0(x) = a_0\left(a_1.e^{-\frac{1}{4}x^2} + a_2.e^{-a_3^2x^2}\right)^2 \text{ and } g_0(y) = b_0\left(b_1.e^{-\frac{1}{4}y^2} + b_2.e^{-b_3^2x^2}\right)^2$$

The parameters a_0 and b_0 are given by:

$$a_0 = \frac{1}{\int_{\mathcal{X}}\left(a_1.e^{-\frac{1}{4}x^2} + a_2.e^{-a_3^2x^2}\right)^2 dx} \text{ and } b_0 = \frac{1}{\int_{\mathcal{Y}}\left(b_1.e^{-\frac{1}{4}y^2} + b_2.e^{-b_3^2x^2}\right)^2 dy}$$

Specifying $f_0(x)$ and $g_0(y)$ has a cost. In fact, normally, for a given specification of $S(x,y)$ and for a given $f(x)$ and $g(y)$ and ν, it is theoretically possible to derive the optimal matching $\pi(x,y)$ in the set of the feasible matching. The specification of $f_0(x)$ and $g_0(y)$ is equivalent to specify $A(U)(x,y)$ and with the decomposition of $U(x,y)$, this specification of $A(U)(x,y)$ gives the expression of $U(x,y)$, and by deduction, the expression of $V(x,y)$ as we know that $U(x,y) + V(x,y) = S(x,y)$. Once $U(x,y)$ and $V(x,y)$ are known, all the model is determined. So we will get an expression for $\pi(x,y)$ derived from the specification of $f_0(x)$ and $g_0(y)$. But this expression of $\pi(x,y)$ is not the exact analytical expression of $\pi(x,y)$ but actually a fitting to its real expression. Then the matching derived from the specification of $f_0(x)$ and $g_0(y)$ is not necessarily feasible. The cost of the specification will be this loss of feasibility. We will penalize the distance of the derived matching from the specification of $f_0(x)$ and $g_0(y)$ to the set of the feasible matching by:

$$\rho\int_{\mathcal{X}}\left(\frac{1}{\alpha}\left(f(x) - (1-\alpha)f_0(x)\right) - \int_{\mathcal{Y}}\pi(x,y)dy\right)^2 dx + \tilde{\rho}\int_{\mathcal{Y}}\left(\frac{1}{\beta}\left(g(y) - (1-\beta)g_0(y)\right) - \int_{\mathcal{X}}\pi(x,y)dx\right)^2 dy$$

with:

$$\pi(x,y) = \left[\frac{1-\alpha}{\alpha}f_0(x)\right]^{\frac{1}{2}}\left[\frac{1-\beta}{\beta}g_0(y)\right]^{\frac{1}{2}} e^{\frac{1}{2}S(x,y)}$$

where the penalization parameters ρ and $\tilde{\rho}$ will be arbitrarily chosen. To understand this penalization, it suffices to remark that:

$$\frac{1}{\alpha}\left(f(x)-(1-\alpha)f_0(x)\right)=f_1(x) \quad \text{and} \quad \frac{1}{\beta}\left(g(y)-(1-\beta)g_0(y)\right)=g_1(y)$$

The feasibility of π requires:

$$f_1(x)=\int_{\mathcal{Y}}\pi(x,y)dy \quad \text{and} \quad g_1(y)=\int_{\mathcal{X}}\pi(x,y)dx$$

We have just taken as penalization:

$$\rho\|f_1-\int_{\mathcal{Y}}\pi(.,y)dy\|^2+\tilde{\rho}\|g_1-\int_{\mathcal{X}}\pi(x,.)dx\|^2$$

In this penalization, remark that we have:

$$\beta=\frac{\alpha\nu}{1-\nu}$$

The parameters α and β must satisfy another constraint which is:

$$\int_{\mathcal{X}}\int_{\mathcal{Y}}\pi(x,y)dx\,dy=1 \;\text{ with }\; \pi(x,y)=\left[\frac{1-\alpha}{\alpha}f_0(x)\right]^{\frac{1}{2}}\left[\frac{1-\beta}{\beta}g_0(y)\right]^{\frac{1}{2}}e^{\frac{1}{2}S(x,y)}$$

This implies:

$$\sqrt{\frac{(1-\alpha)(1-\beta)}{\alpha\beta}}=\frac{1}{\int_{\mathcal{X}}\int_{\mathcal{Y}}[f_0(x)]^{\frac{1}{2}}\,[g_0(y)]^{\frac{1}{2}}\,e^{\frac{1}{2}S(x,y)}dxdy}$$

We will denote this quantity by K. So:

$$K=\frac{1}{\int_{\mathcal{X}}\int_{\mathcal{Y}}[f_0(x)]^{\frac{1}{2}}\,[g_0(y)]^{\frac{1}{2}}\,e^{\frac{1}{2}S(x,y)}dxdy}$$

Combining the relations:

$$K=\sqrt{\frac{(1-\alpha)(1-\beta)}{\alpha\beta}} \quad \text{and} \quad \beta=\frac{\alpha\nu}{1-\nu}$$

We obtain:

$$\alpha = \frac{-1+\sqrt{1+4\,\nu(1-\nu)\left(\frac{K^2}{1-\nu}-1\right)}}{2\,\nu\left(\frac{K^2}{1-\nu}-1\right)} \quad \text{and} \quad \beta = \frac{-1+\sqrt{1+4\,\nu(1-\nu)\left(\frac{K^2}{1-\nu}-1\right)}}{2(1-\nu)\left(\frac{K^2}{1-\nu}-1\right)}$$

The likelihood for a given λ can be written as follows:

$$L(\lambda) = \prod_{(m,w)\in\mathcal{C}} \left(\frac{\alpha\beta(\pi(x_m,y_w))^2}{f(x)g(y_w)}\right) \prod_{m\in\mathcal{S}_h} \left(\frac{(1-\alpha)f_0(x_m)}{f(x_m)}\right) \prod_{w\in\mathcal{S}_f} \left(\frac{(1-\beta)g_0(y_w)}{g(y_w)}\right)$$

We specify $f_0(x)$ and $g_0(y)$ as described above and we maximize the log-likelihood with the penalization we have defined. In other terms, we maximize:

$$\ln(L(\lambda)) + \rho \int_{\mathcal{X}} \left(\frac{1}{\alpha}\left(f(x)-(1-\alpha)f_0(x)\right) - \int_{\mathcal{Y}} \pi(x,y)dy\right)^2 dx$$

$$+\tilde{\rho} \int_{\mathcal{Y}} \left(\frac{1}{\beta}\left(g(y)-(1-\beta)g_0(y)\right) - \int_{\mathcal{X}} \pi(x,y)dx\right)^2 dy$$

with the constraints:

$$\alpha = \frac{-1+\sqrt{1+4\,\nu(1-\nu)\left(\frac{K^2}{1-\nu}-1\right)}}{2\,\nu\left(\frac{K^2}{1-\nu}-1\right)} \quad \text{and} \quad \beta = \frac{-1+\sqrt{1+4\,\nu(1-\nu)\left(\frac{K^2}{1-\nu}-1\right)}}{2(1-\nu)\left(\frac{K^2}{1-\nu}-1\right)}$$

where

$$K = \frac{1}{\int_{\mathcal{X}}\int_{\mathcal{Y}} \left[f_0(x)\right]^{\frac{1}{2}} \left[g_0(y)\right]^{\frac{1}{2}} e^{\frac{1}{2}S(x,y)} dxdy}$$

3.6.3 Derivation of the distribution of the net utilities

To derive the distribution of the net marriage surpluses $u-\varepsilon^0$ and $v-\eta^0$, we adopt the following little notations. Let u_m^{-0} and v_w^{-0} be respectively the maximum utility of the man m from his different acquaintances and the maximum utility of the woman w from her different acquaintances. We have:

$$u_m = \max\left\{\varepsilon_m^0, u_m^{-0}\right\} \quad \text{and} \quad v_w = \max\left\{\eta_w^0, v_w^{-0}\right\}$$

The variables ε_m^0 and u_m^{-0} are independent and so are the variables η_w^0 and v_w^{-0}. The distributions of u_m^{-0} and v_w^{-0} are given below. Refer to the Proposition 4 for its immediate proof.

$$u_m^{-0}|X_m = x \ \sim \ \text{Gumbel}\left(\ln\left(\int_{\mathcal{Y}} e^{U(x,y)}dy\right), 1\right)$$

and

$$v_w^{-0}|Y_w = y \ \sim \ \text{Gumbel}\left(\ln\left(\int_{\mathcal{X}} e^{V(x,y)}dx\right), 1\right)$$

Here, we have:

$$\mathcal{X} = \mathcal{Y} =]-\infty, +\infty[$$

We can rewrite the net utilities from marriage as follows:

$$E_m^M = u_m - \varepsilon_m^0 = \max\{0, u_m^{-0} - \varepsilon_m^0\} = (u_m^{-0} - \varepsilon_m^0)_+$$

and

$$E_w^W = v_w - \eta_w^0 = \max\{0, u_w^{-0} - \eta_w^0\} = (v_w^{-0} - \eta_w^0)_+$$

As we know that ε_m^0 and η_w^0 are standard Gumbel, and independent respectively with u_m^{-0} and v_w^{-0}, we can then derive the joint distribution of $(\varepsilon_m^0, u_m^{-0})$ and of (η_0, v_w^{-0}). We then derive by convolution, the distribution of $u_m^{-0} - \varepsilon_m^0$ and the distribution of $v_w^{-0} - \eta_w^0$. We obtain the probability density function of $E_m^M = (u_m^{-0} - \varepsilon_m^0)_+$ and the probability density function of $E_w^W = (v_w^{-0} - \eta_w^0)_+$

$$f_{E^M}(u) = \mathbb{1}(u \geq 0).\left[\int_{-\infty}^{+\infty} e^{u-2s}\left(\int_{-\infty}^{+\infty} e^{U(x,y)}dy\right)\exp\left\{-e^{-s}\left(e^u + \int_{-\infty}^{+\infty} e^{U(x,y)}dy\right)\right\}ds\right]$$

Analogously, we have:

$$g_{E^W}(v) = \mathbb{1}(v \geq 0).\left[\int_{-\infty}^{+\infty} e^{v-2s}\left(\int_{-\infty}^{+\infty} e^{V(x,y)}dx\right)\exp\left\{-e^{-s}\left(e^v + \int_{-\infty}^{+\infty} e^{V(x,y)}dx\right)\right\}ds\right]$$

These distributions have as support space $\mathbb{R}_+$ with a mass at the point 0 that is the probability of being single conditionally to the characteristics. In fact we have:

$$\mathbb{P}(E^M \leq 0) = \mathbb{P}(u_m = \varepsilon^0|x) = 1 - \alpha(x) = 1 - \int_0^{+\infty} f_{E^M}(u)du$$

and

$$\mathbb{P}(E^W \leq 0) = \mathbb{P}(v_w = \eta^0 | y) = 1 - \beta(y) = 1 - \int_0^{+\infty} g_{E^W}(v)dv$$

where $\alpha(x)$ and $\beta(y)$ are respectively the conditional probability of marriage of men and of women with respect to the characteristics.

We can now derive the conditional distribution of E^M and of E^W with respect to marriage. Note that, there is marriage if and only if the net utility from marriage is strictly greater than 0. We the distributions we are searching are respectively the conditional distributions of E^M and of E^W with respect to the event $\{E^M > 0\}$ and to the event $\{E^W > 0\}$.

$$f_{E^W|M=1}(u) = \frac{f_{E^M}(u)}{\int_0^{+\infty} f_{E^M}(u)du} \text{ and } g_{E^W|M=1}(v) = \frac{g_{E^W}(v)}{\int_0^{+\infty} g_{E^W}(v)dv}$$

4 Matching In Closed Forms

Abstract

In this paper, we consider an equally weighted bipartite matching market with transferability of the utility. We try to show the ways in which the model suggested in the previous chapter can be extended to the case $\alpha \in \{0, 1\}$ where α is the probability of matching on each of the two sides of the market. We first prove that the deterministic part of the utility of the singles can be expressed in function of the probability of matching and that allows us to rewrite the theoretical results of the model of Chapter II at a new form that extends directly the model of Chapter II to the case $\alpha \in \{0, 1\}$. We then consider a particular case of full matching i.e $\alpha = 1$, with a quadratic specification and Gaussian distributions of the characteristics. We find entirely the analytical expression the optimal matching and of the individual surpluses gained by the partners from the sharing of the joint surplus at the equilibrium. We provide two ways to estimate the model.

4.1 Introduction

There exists a rich literature on matching models. Gale and Shapley (1962) defined theoretically an equilibrium for two-sided matching markets. The mathematical exploitation of this definition of equilibrium will be one the main later researches in this field. Shapely and Shubik (1971) proved an important result that is the maximization of the total social welfare by the optimal matching. But the resolution of their equation requires too strong assumptions. The first econometric framework to estimate matching models has been proposed by Choo and Siow (2006) with transferability of the utility. They considered a discrete logit framework to model the gain from marriage. They have been able to identify individual gains from marriage and the matching matrix in function of the primitives of the model. They obtain closed forms that allow easily a nonparametric estimation of the model. Chiappori, Salanié and Weiss (2008) used an heteroskedastic version of the Choo and Siow model to study the assortative matching on the marriage market. Lindenlaub (2017), considering multidimensional setting, studied the equilibrium properties and particularly on a matching problem with quadratic surplus and Gaussian distributions of the populations, and such that there is no unobserved heterogeneity and bivariate observed characteristics. Galichon and Salanié (2012) give a general discrete framework with transferable utility and with equilibrium in the sense of Gale and Shapley. From their framework we can essentially derive as particular cases, the results of Choo and Siow and the heteroskedastic model of Chiappori, Salanié and Weiss. One of the main results of their study is to prove that the optimal matching maximizes the total social welfare and this social welfare can be decomposed as the sum of the averaged social surplus and an entropy which quantifies somehow the statistical disorder in the population. Dupuy and Galichon (2014) have proposed a continuous extension of the econometric matching model of Choo and Siow (2006). That model considered mainly a full matching on an equally weighted bipartite matching market i.e all the individuals get matched almost surely at the equilibrium of the market. They treat the integration of singles and show that the equilibrium does not change when we take into account singles. They found that the individual surpluses can be identified up to two existing, unique but undetermined functions; the individual surplus of a man in a union with a woman is identified up to a function $a(.)$ that depends on the characteristics of the man, and analogously, the individual surplus of woman in a union with a man is identified up to a function $b(.)$ that depends on the characteristics of the woman. Bojilov and Galichon (2016) with the setting of Dupuy and Galichon, show that the assignment problem can be solved analytically in quadratic specification with Gaussian distributions of the observable characteristics, i.e, they derive analytically the optimal matching function by identifying the affinity matrix, and they also derive the conditional distributions $X|Y$ and $Y|X$ where X and Y are respectively the characteristics of men and of women. But what they do not do it to provide analytical expression of the individual surpluses and the functions $a(.)$ and $b(.)$ are then not precised analytically. The interest to have analytical expression of the individual is the possibility to compare the surplus gap for instance within the unions. Actually the surplus gap between partners is a separable function in the characteristics of the partners and it involves directly the functions $a(x)$ and $b(y)$. Finding analytically these functions implies automatically the analytical expressions of the individual surplus and reversely. Our main goal in this work is to solve entirely analytically the matching problem.

In Chapter II, we have proposed in the light of Dupuy and Galichon another approach of a continuous extension of the Choo and Siow model that takes systematically into account the possibility of existence of singles at the equilibrium, i.e the probability of marriage is completely endogenous. The model has provided some theoretical results quite similar to the ones found of Dupuy-Galichon. We consider the same framework as in the chapter 2, a bipartite matching market with transferable utility, possibly infinite and where the two sides have the same weight and in addition, we assume that the deterministic of the utility of single individuals is a constant parameter. We rewrite the equilibrium obtained in the chapter 2 and this rewriting has the advantage to allow directly the generalization of the model to the case $\alpha \in \{0, 1\}$ where α is the probability of matching on each of the two sides of the market. We show the equivalence of our approach with the model of Dupuy and Galichon in the case of full matching. We also show briefly that the framework we suggest here finds analytically the equilibrium of the market when the matching tends to be scarce. We then go on the fundamental purpose of this work. We consider in addition to the framework, as Bojilov and Galichon (2016), a quadratic specification of the joint surplus, with Gaussian distributions of the observable characteristics. We provide analytical expressions for the matching function and we retrieve the theorem of Bojilov and Galichon (2016) identifying the affinity matrix. Furthermore we find entirely the analytical expressions of individual surpluses. We finally present two ways to estimate parametrically the model, by maximum likelihood eased by the fact that it is completely expressed analytically, and the estimation by moments based on the identification of the affinity matrix.

4.2 The model

4.2.1 Matching Market Description

We assume the market to be bipartite, one-to-one with transferable utility. One part of the market is $\mathcal{H}$, and the second one is $\mathcal{F}$. The whole population will be denoted by $\mathcal{P}$. We will use the generic terms m and w respectively for individuals in $\mathcal{H}$ and for individuals in $\mathcal{F}$. Without losing in generality, the set $\mathcal{H}$ will be considered as the set of men and the set $\mathcal{F}$ the set of women. This framework can perfectly also be applied to any matching market such as employment market, Real Estate market, etc... The market we consider here is possibly infinite and its two sides are assumed to be in equal proportions.

$$\mathbb{P}(\mathcal{H}) = \mathbb{P}(\mathcal{F}) = \frac{1}{2}$$

Individuals of each side decide either to match with another individual of the opposite side or to remain single. The men are characterized by a continuous vector X and observable on the market. Analogously, women are characterized by a continuous vector Y observable on the market too. The support spaces of X and Y are respectively $\mathbb{R}^p$ and $\mathbb{R}^q$. We denote respectively by P and Q the probability distributions of X and Y. These probability distributions are assumed to have respectively a probability density f and a probability density g. These functions are assumed to be exogenous to the model. On this market, each individual aims to match with another individual of the opposite side. He or she remains unmatched in the case he or her does not find the "ideal" partner. We will bring more precision about the mechanism of matching on the market. To go further, we need to define the term of matching.

Definition 4.1 (Matching)**.** *A matching is any density function π associated to P and Q and defined on the Cartesian space $\mathcal{X} \times \mathcal{Y}$.*

A matching describes precisely the manner any particular union of attributes (x, y) is formed relatively to the set of couples. For more precision, it is important to know that all the density functions on $\mathcal{X} \times \mathcal{Y}$ are not realistic. Before refining the definition of matching, we need to adopt some useful notations. We define the variable M that is the indication of the matching status. As we can guess it, this variable is endogenous because it is completely determined by the market.

$$\forall\, k \in \mathcal{P}, \;\; M_k = \mathbb{1}\{ \text{ k is matched } \}$$

We also define the variables $\tilde{X}$ and $\tilde{Y}$ respectively observable on the set of women $\mathcal{F}$ and on the set of men $\mathcal{H}$. For any particular man m, the variable $\tilde{Y}_m$ represents the attributes of his partner. And for any particular woman w, $\tilde{X}_w$ represents the attributes of her partner. In the case the partner does not exist, the attributes are $\emptyset$, i.e the individual is unmatched. So $\tilde{X}$ and $\tilde{Y}$ have respectively as support spaces $\mathcal{X} \cup \emptyset$ and $\mathcal{Y} \cup \emptyset$. We can then define for any man m, the variable $(X_m, \tilde{Y}_m)$ and for any woman w the variable $(\tilde{X}_w, Y_w)$. We will denote by α the probability to choose at the equilibrium a matched man in the set of the men $\mathcal{H}$. This probability is also equal to the probability to choose at the equilibrium a matched woman in the set of women $\mathcal{F}$.

$$\alpha = \mathbb{P}(M = 1|\mathcal{H}) = \mathbb{P}(M = 1|\mathcal{F})$$

Let f_1 and f_0 be respectively the conditional probability density of X with respect to the event $\{M = 1, \mathcal{H}\}$ and to the event $\{M = 0, \mathcal{H}\}$. Analogously, we define respectively g_1 and g_0 the conditional probability density of Y with respect to the event $\{M = 1, \mathcal{F}\}$ and to the event $\{M = 0, \mathcal{F}\}$.

$$f_1(x) = f(x|M = 1, \mathcal{H}) \quad \text{and} \quad f_0(x) = f(x|M = 0, \mathcal{H})$$

and

$$g_1(y) = g(y|M = 1, \mathcal{F}) \quad \text{and} \quad g_0(y) = g(y|M = 0, \mathcal{F})$$

The density functions f and g are exogenous. The model will aim to identify α, f_0, f_1, g_0 and g_1. In some particular cases, the functions f_1 and g_1 can be exogenous; for instance when the matching is assume to be full i.e all the individuals are matched, then the conditional densities to marriage f_1 and g_1 are exactly equal respectively to f and g. We will treat later that particular case.

All the matchings are not feasible. We refine the definition as follows.

Definition 4.2 (Feasible Matching)**.** *A feasible matching is a joint density function* $\pi(x, y)$ *defined on* $\mathcal{X} \times \mathcal{Y}$ *and a probability* α*.*

Note that $\pi(x, y)$ is the conditional joint density of $(X, \tilde{Y})$ with respect to $(M = 1, \mathcal{H})$ and equivalently, it is also the conditional joint density of $(\tilde{X}, Y)$ with respect to $(M = 1, \mathcal{F})$. We have:

$$\pi(x, y) = \pi(x, y|M = 1, \mathcal{H}) = \pi(x, y|M = 1, \mathcal{F})$$

Let's denote by $\Pi(P,Q)$ the set of feasible matchings. In the following subsection, we will introduce the notion of matching surplus and we will add heterogeneity to the model.

4.2.2 Matching Surplus and Heterogeneity

The population is infinite. We make the separability assumption (previously suggested by Choo and Siow (2006) and formalized by Galichon and Salanié (2012) and Chiappori, Salanié and Weiss (2017)): that is the joint utility generated when a man m with characteristics x_m matches with a woman w with characteristics y_w does not depend on interactions between their unobserved characteristics, conditional on (x_m, y_w).

Assumption 4.1 (Separability)**.** *The joint utility of a couple (m,w) with characteristics (x,y) is written:*

$$S(x,y) + \sigma\varepsilon_m(y) + \sigma\eta_w(x)$$

where $S(x,y)$ is the deterministic part of the joint utility, σ is a positive constant, $\varepsilon_m(y)$ and $\eta_w(x)$ are two stochastic terms.

The function $S(x,y)$ can be interpreted as the joint systematic surplus generated by the matching between a man of characteristics x and a woman of characteristics y. The term $\varepsilon_m(y)$ is a stochastic process that represents for the man m his utility from sympathy shock with a woman with attributes y. The term $\eta_w(x)$ is a stochastic process that represents for the woman w her utility from sympathy shock with a man with attributes x, and σ is a positive parameter. The joint utility from marriage represents the utility that partners will share between them, according to a rule that depends on the competition force in the marriage market.

The individuals compare their utility from marriage with what they would have if they remain single when they are single. We model in the following assumption, the utility of single individuals.

Assumption 4.2. *The utility of a single man m is:*

$$u_m^0 = A^0 + \sigma\varepsilon_m^0$$

and the utility of a single woman w is:

$$v_w^0 = B^0 + \sigma\eta_w^0$$

where A^0 is a general parameter common to all single men and B^0 a general parameter common to all single women; σ is a positive constant, ε_m^0 is a stochastic term specific to man m and η_w^0 also a stochastic term specific to woman w.

Remark that this assumption is equivalent to the general case in which the deterministic parts of the utility of the single are functions $A^0(x)$ and $B^0(y)$. In that case, it suffices to redefine a new joint surplus as $S(x,y) - A^0(x) - B^0(y) + \mathbb{E}(A^0(X)) + \mathbb{E}(B^0(Y))$ and new utilities of singles $u^0 = \mathbb{E}(A^0(X)) + \sigma\varepsilon^0$ for single men and $v^0 = \mathbb{E}(B^0(X)) + \sigma\eta^0$ for single women and this redefinition makes us come back to verify Assumption 4.2.

Dupuy and Galichon Specification The following specification is in the light of Dupuy and Galichon (2014). This has also been considered by in the previous chapter. We describe it in the following lines. Each man m of attributes x meets a subset, possibly infinite, of the set of women to make his choice. We assume ε_m^0 to have a standard Gumbel distribution. We index his acquaintances by $\mathbb{N}^*$. From a particular acquaintance k with a woman of attributes y_k^m, the man gets the utility:

$$u_m^k = U(x, y_k^m) + \sigma\varepsilon_k^m$$

We assume the process $\{(y_k^m, \varepsilon_k^m)_k\,,\, k \in \mathbb{N}^*\}$ to be a Poisson process on the space $\mathcal{Y} \times \mathbb{R}$ with intensity $e^{-\varepsilon-\gamma}$, where γ is the Euler-Mascheroni constant. Analogously, for each woman w of attributes y, we assume η_w^0 to have have standard Gumbel distribution; the woman w meets a possibly infinite subset of the men to make her choice. Her acquaintances are indexed by $\mathbb{N}^*$. From a particular acquaintance k with a man of attributes x_k^j, she gets the utility:

$$v_w^k = V(x_k^w, y) + \sigma\eta_k^w$$

The process $\{(x_k^w, \eta_k^w)_k\,,\, k \in \mathbb{N}^*\}$ is assumed to be a Poisson process on the space $\mathcal{X} \times \mathbb{R}$ with intensity $e^{-\eta-\gamma}$.

We assume that for any man m of attributes $x_m = x$, the term ε_m^0 is independent from any sympathy shock ε_k^m from an acquaintance k. And analogously for women, the term η_w^0 is independent from any sympathy shock η_k^w from an acquaintance k.

We summarize that in the following assumption.

Assumption 4.3. *(Dupuy and Galichon, 2014)*

(i) The stochastic processes $\varepsilon^m(y)$ and $\eta^w(x)$ are max-stable process of the form:

$$\begin{aligned} \varepsilon^m(y) &= \max_k(\varepsilon_k^m : y_k = y) \text{ if the set } \{k : y_k = y\} \text{ is non-empty} \\ &= -\infty \text{ otherwise,} \end{aligned}$$

where $\{(y_k^m, \varepsilon_k^m),\ k \in \mathbb{K}\}$ follows a Poisson process on the space $\mathcal{Y} \times \mathbb{R}$ of intensity $e^{-\varepsilon-\gamma} dy\, d\varepsilon$, and

$$\begin{aligned} \eta^w(x) &= (\eta_l^w : x_l = x) \text{ if the set } \{l : x_l = x\} \text{ is non-empty} \\ &= -\infty \text{ otherwise.} \end{aligned}$$

where $\{(x_k^w, \eta_k^w), l \in \mathbb{L}\}$ follows a Poisson process on $\mathcal{X} \times \mathbb{R}$ with intensity $e^{-\eta-\gamma} dx\, d\eta$.

(ii) The stochastic terms ε_m^0 and η_w^0 follow a standard Gumbel distribution, independent of ε_k^m for $k \in \mathbb{K}$ and indepedent of η_l^w for $l \in \mathbb{L}$, respectively.

We will denote respectively by $\bar{U}(x,y)$ and $\bar{V}(x,y)$ the net individual surpluses of men of attributes x and women of attributes y from their matching and we denote by $\bar{S}(x,y)$ the net joint surplus.

$$\bar{U}(x,y) = U(x,y) - A^0,\ \bar{V}(x,y) = V(x,y) - B^0 \quad \text{and} \quad \bar{S}(x,y) = S(x,y) - A^0 - B^0.$$

4.2.3 Equilibrium and Identification

We suppose that the market equilibrium is stable in the sense of Gale and Shapley (1962), that is, there is no married person who would rather be single, and there is no pair of (married or unmarried) persons who prefer to form a new union. The equilibrium is the solution of the following maximization problems. Each man must choose in the set of his acquaintances the partner that brings him the optimal utility; in the case this utility is higher than his proper utility he chooses her. And reversely, each woman must choose in the set of her acquaintances the partner that brings her the optimal utility; she chooses this man in the case this utility she gets from the union is higher than her proper utility. A marriage between a man and a woman occurs when they choose each other. Let's denote respectively by $u_m(x)$ and $v_w(y)$ the utility at equilibrium of a man m of attributes x and the utility at equilibrium of a woman w of attributes y. We also denote:

$$\bar{u}_m = u_m - A^0 \text{ and } \bar{v}_w = v_w - B^0.$$

A man m of attributes x solves:

$$\bar{u}_m(x) = \max\{\sigma\varepsilon_m^0, \max_{k\in\mathbb{N}^*}\{\bar{U}(x, y_k) + \sigma\varepsilon_m^k\}\}$$

And a woman w of attributes y solves:

$$\bar{v}_w(y) = \max\{\sigma\eta_w^0, \max_{k\in\mathbb{N}^*}\{\bar{V}(x_k, y) + \sigma\eta_w^k\}\}$$

At the equilibrium, each individual does his choice in the set of his or her acquaintances. A match occurs between a man m and a woman w if the man is an acquaintance of the man m and w brings to him the maximum utility in his set of acquaintances and this utility is higher to his utility of remaining single. And reversely, man m is an acquaintance of the woman w and brings to her the maximum of utility in the set of her acquaintances and this utility is higher to her utility of being single. Dupuy and Galichon (2014) have shown that the optimal matching maximizes the social surplus.

Social Surplus

The social surplus in this framework is the averaged utility individuals obtain at the equilibrium of the market. We widely adopt the notation of Dupuy and Galichon and so denote it by $\mathcal{W}$. We denote by $\bar{\mathcal{W}}$ the net social surplus. As the two sides of the market are equally weighted, we have:

$$\bar{\mathcal{W}} = \frac{1}{2}\mathbb{E}(\bar{u}) + \frac{1}{2}\mathbb{E}(\bar{v}) = \frac{1}{2}\mathbb{E}_P\mathbb{E}(\bar{u}|X) + \frac{1}{2}\mathbb{E}_Q\mathbb{E}(\bar{v}|Y)$$

By remarking that:

$$\mathbb{E}_{\alpha\pi}(\bar{S}(X,Y)) = \mathbb{E}_{\alpha\pi}(\bar{U}(X,Y)) + \mathbb{E}_{\alpha\pi}(\bar{V}(X,Y))$$

The social surplus can then be written as follows:

$$\bar{\mathcal{W}}(\pi,\alpha) = \frac{1}{2}\mathbb{E}_{\alpha\pi}(\bar{S}(X,Y)) - \frac{1}{2}\left(\mathbb{E}_{\alpha\pi}(\bar{U}(X,Y)) - \mathbb{E}(\bar{u})\right) - \frac{1}{2}\left(\mathbb{E}_{\alpha\pi}(\bar{V}(X,Y)) - \mathbb{E}(\bar{v})\right)$$

As we have:

$$\mathbb{E}_{\alpha\pi}(\bar{U}(X,Y)) = \mathbb{E}_P\mathbb{E}_{\alpha\pi|X}(\bar{U}(X,Y|X)) \text{ and } \mathbb{E}(\bar{u}) = \mathbb{E}_P\mathbb{E}(\bar{u}|X)$$

and

$$\mathbb{E}_{\alpha\pi}(\bar{V}(X,Y)) = \mathbb{E}_Q\mathbb{E}_{\alpha\pi|Y}(\bar{V}(X,Y|Y)) \text{ and } \mathbb{E}(\bar{v}) = \mathbb{E}_Q\mathbb{E}(\bar{v}|Y)$$

we deduce that:

$$\begin{aligned}\bar{\mathcal{W}}(\pi,\alpha) = \frac{1}{2}\mathbb{E}_{\alpha\pi}(\bar{S}(X,Y)) - \frac{1}{2}\mathbb{E}_P\left(\mathbb{E}_{\alpha\pi|X}(\bar{U}(X,Y|X)) - \mathbb{E}(\bar{u}|X)\right)& \\ - \frac{1}{2}\mathbb{E}_Q\left(\mathbb{E}_{\alpha\pi|Y}(\bar{V}(X,Y|Y)) - \mathbb{E}(\bar{v}|Y)\right)&\end{aligned}$$

By denoting by:

$$\mathcal{E}(\pi,\alpha) = \mathbb{E}_P\left(\mathbb{E}_{\alpha\pi|X}(\bar{U}(X,Y|X)) - \mathbb{E}(\bar{u}|X)\right) + \mathbb{E}_Q\left(\mathbb{E}_{\alpha\pi|Y}(\bar{V}(X,Y|Y)) - \mathbb{E}(\bar{v}|Y)\right)$$

we can then remark that the social net surplus takes the form shown by Dupuy and Galichon (2014) and in Chapter II, that is:

$$\bar{\mathcal{W}}(\pi,\alpha) = \frac{1}{2}\mathbb{E}_{\alpha\pi}(\bar{S}(X,Y)) - \frac{1}{2}\mathcal{E}(\pi,\alpha) = \frac{1}{2}\int_{\mathcal{X}}\int_{\mathcal{Y}}\alpha\pi(x,y)\bar{S}(x,y)dx\,dy - \frac{1}{2}\mathcal{E}(\pi,\alpha)$$

The maximization of the social surplus $\mathcal{W}(\pi,\alpha)$ is equivalent to the maximization of the net social surplus $\bar{\mathcal{W}}(\pi,\alpha)$ over the set of feasible matchings. In the light of Galichon and Salanié (2012) and Dupuy and Galichon (2014), we have proved previously in Chapter II the following result from the first order of the maximization of the social net surplus $\bar{\mathcal{W}}(\pi, \alpha)$:

$$\bar{U}(x, y) = \sigma \ln \left(\frac{\alpha}{1 - \alpha} \frac{\pi(x, y)}{f_0(x)} \right) \tag{23}$$

and

$$\bar{V}(x, y) = \sigma \ln \left(\frac{\alpha}{1 - \alpha} \frac{\pi(x, y)}{g_0(y)} \right) \tag{24}$$

In other terms, we have:

$$U(x, y) = A^0 + \sigma \ln \left(\frac{\alpha}{1 - \alpha} \frac{\pi(x, y)}{f_0(x)} \right) \tag{25}$$

and

$$V(x, y) = B^0 + \sigma \ln \left(\frac{\alpha}{1 - \alpha} \frac{\pi(x, y)}{g_0(y)} \right) \tag{26}$$

Actually, these expressions are the forms that take the results found in the second chapter when we consider an equal proportion of men and women in the population. These expressions show how the individual surpluses are linked to the joint matching density function. From these expressions of $U(x, y)$ and $V(x, y)$, we have, by rewriting:

$$\frac{1 - \alpha}{\alpha} f_0(x) \exp \left(\frac{U(x, y)}{\sigma} \right) = \exp \left(\frac{A^0}{\sigma} \right) \pi(x, y)$$

and

$$\frac{1 - \alpha}{\alpha} g_0(y) \exp \left(\frac{V(x, y)}{\sigma} \right) = \exp \left(\frac{B^0}{\sigma} \right) \pi(x, y)$$

The function π is a density function on $\mathcal{X} \times \mathcal{Y}$ so:

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} \pi(x, y) dx\, dy = 1$$

This leads to the following lines:

$$\exp \left(\frac{A^0}{\sigma} \right) = \frac{1 - \alpha}{\alpha} \int_{\mathcal{X}} \int_{\mathcal{Y}} f_0(x) \exp \left(\frac{U(x, y)}{\sigma} \right) dx\, dy$$

and

$$\exp\left(\frac{B^0}{\sigma}\right) = \frac{1-\alpha}{\alpha}\int_{\mathcal{X}}\int_{\mathcal{Y}} g_0(y)\exp\left(\frac{V(x,y)}{\sigma}\right)dx\,dy$$

In other terms, we can write:

$$A^0 = \sigma\ln\left(\frac{1-\alpha}{\alpha}\right) + \sigma\ln\left(\int_{\mathcal{X}}\int_{\mathcal{Y}} f_0(x)\exp\left(\frac{U(x,y)}{\sigma}\right)dx\,dy\right)$$

and

$$B^0 = \sigma\ln\left(\frac{1-\alpha}{\alpha}\right) + \sigma\ln\left(\int_{\mathcal{X}}\int_{\mathcal{Y}} g_0(y)\exp\left(\frac{V(x,y)}{\sigma}\right)dx\,dy\right)$$

We denote:

$$\delta_1 := \ln\left(\int_{\mathcal{X}}\int_{\mathcal{Y}} f_0(x)\exp\left(\frac{U(x,y)}{\sigma}\right)dx\,dy\right) \tag{27}$$

and

$$\delta_2 := \ln\left(\int_{\mathcal{X}}\int_{\mathcal{Y}} g_0(y)\exp\left(\frac{V(x,y)}{\sigma}\right)dx\,dy\right). \tag{28}$$

The parameters A^0 and B^0 can then be written:

$$A^0 = \sigma\ln\left(\frac{1-\alpha}{\alpha}\right) + \sigma\delta_1 \tag{29}$$

and

$$B^0 = \sigma\ln\left(\frac{1-\alpha}{\alpha}\right) + \sigma\delta_2 \tag{30}$$

This is not an identification of A^0 and B^0. Actually the equations (27) and (29) are equivalent and the equations (28) and (30) are equivalent as well. To prove that, using the relations $U = \bar{U} + A^0$ and $V = \bar{V} + B^0$, we can rewrite δ_1 and δ_2 at the forms:

$$\delta_1 = \ln\left(\int_{\mathcal{X}}\int_{\mathcal{Y}} f_0(x)\exp\left(\frac{\bar{U}(x,y)+A}{\sigma}\right)dx\,dy\right) = \frac{A^0}{\sigma} + \ln\left(\int_{\mathcal{X}}\int_{\mathcal{Y}} f_0(x)\exp\left(\frac{\bar{U}(x,y)}{\sigma}\right)dx\,dy\right)$$

and

$$\delta_2 = \ln\left(\int_{\mathcal{X}}\int_{\mathcal{Y}} g_0(y)\exp\left(\frac{\bar{V}(x,y)+B}{\sigma}\right)dx\,dy\right) = \frac{B^0}{\sigma} + \ln\left(\int_{\mathcal{X}}\int_{\mathcal{Y}} g_0(y)\exp\left(\frac{\bar{V}(x,y)}{\sigma}\right)dx\,dy\right)$$

We recall from the theorem given in the previous chapter, the expressions of α, f_0 and g_0 below in view to rewrite the expressions of δ_1 and δ_2:

$$\alpha = \int_{\mathcal{X}} \int_{\mathcal{Y}} \frac{e^{\frac{\bar{U}(x,y)}{\sigma}}}{1 + \int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy} f(x) dx\, dy \tag{31}$$

Since the probability of marriage in the set of men and the probability of marriage in the set of women are equal from the fact the two sides of the market have the same weight $\frac{1}{2}$, we then also have:

$$\alpha = \int_{\mathcal{X}} \int_{\mathcal{Y}} \frac{e^{\frac{\bar{V}(x,y)}{\sigma}}}{1 + \int_{\mathcal{X}} e^{\frac{\bar{V}(x,y)}{\sigma}} dx} g(y) dx\, dy \tag{32}$$

The functions f_0 and g_0 are given by:

$$f_0(x) = \frac{1}{(1-\alpha)\left(1 + \int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy\right)} f(x) \tag{33}$$

and

$$g_0(y) = \frac{1}{(1-\alpha)\left(1 + \int_{\mathcal{X}} e^{\frac{\bar{V}(x,y)}{\sigma}} dx\right)} g(y) \tag{34}$$

By replacing (33) in the rewritten expression of δ_1, we find that:

$$\delta_1 = \frac{A^0}{\sigma} - \ln(1-\alpha) + \ln\left(\int_{\mathcal{X}} \int_{\mathcal{Y}} \frac{e^{\frac{\bar{U}(x,y)}{\sigma}}}{1 + \int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}} dy} f(x)\right)$$

and analogous, by replacing (34) in the rewritten expression of δ_2, we have:

$$\delta_2 = \frac{B^0}{\sigma} - \ln(1-\alpha) + \ln\left(\int_{\mathcal{X}} \int_{\mathcal{Y}} \frac{e^{\frac{\bar{V}(x,y)}{\sigma}}}{1 + \int_{\mathcal{Y}} e^{\frac{\bar{V}(x,y)}{\sigma}} dy} g(y)\right)$$

The integration of (31) and (32) respectively in these expressions gives:

$$\delta_1 = \frac{A^0}{\sigma} - \ln\left(\frac{1-\alpha}{\alpha}\right) \quad \text{and} \quad \delta_2 = \frac{B^0}{\sigma} - \ln\left(\frac{1-\alpha}{\alpha}\right)$$

And these expressions are respectively equivalent to (29) and (30). What do the equivalence between (29) and (27) and the equivalence between (30) and (28) mean ? These equivalences mean that one cannot exploit separately (29) and (27) nor exploit separately (30) and (28) since the exploitation of one is equivalent to the exploitation of the second. Concretely, it means that we know δ_1 from (27), then we can no more use (29) to derive the value of A^0 and reversely and same of δ_2 and B^0. But we can fix A^0 and B^0 and then derive δ_1 and δ_2. The value of δ_1 and the value of δ_2 cannot be chosen arbitrarily because they depend actually on f_0 and $\bar{U}$ and on g_0 and $\bar{V}$ given by the equilibrium of the market.

To ensure that (29) and (30) are not an identification, let's verify if it is possible to identify σ. As we know that only the individual net surpluses $\bar{U}$ and $\bar{V}$ are identified, we can, without losing in generality, assume $A^0 = B^0 = 0$. we have then $U = \bar{U}$ and $V = \bar{V}$. As we have shown that:

$$\delta_1 = \frac{A^0}{\sigma} - \ln\left(\frac{1-\alpha}{\alpha}\right) \quad \text{and} \quad \delta_2 = \frac{B^0}{\sigma} - \ln\left(\frac{1-\alpha}{\alpha}\right)$$

we deduce automatically that:

$$\delta_1 = \delta_2 = -\ln\left(\frac{1-\alpha}{\alpha}\right)$$

Then by replacing these expressions respectively in (29) and (30), we obtain:

$$0 = \sigma \times 0$$

The term σ can then be set to any positive and finite value. In conclusion, when $A^0 = B^0 = 0$, we cannot identify σ. In other terms, we cannot identify σ. Even if (29) and (30) are not an identification, we can nevertheless give an interpretation for these relations. Consider two different stable markets 1 and 2 with the same joint function, the same distributions P and Q, and the same σ and on which the number of men and the number of women are equal and such that at the equilibrium the conditional probability of marriage α_1 on the male population in the market 1 is lower than the conditional probability of marriage α_2 on the male population in the market 2. Single men on the market 1 are expected to be happier than single men on the market 2 i.e

$A_1 \geq A_2$ where A_1 is the deterministic part of the utility of single men on the market 1 and A_2 is the deterministic part of the utility of single men on the market 2. The intuition is that the utility of singles is expected to be relatively higher when marriage is rare and reversely, it is lower when marriage is very frequent. And we can remark that:

$$\lim_{\alpha \to 1^-} \ln\left(\frac{1-\alpha}{\alpha}\right) = -\infty \text{ and } \lim_{\alpha \to 0^+} \ln\left(\frac{1-\alpha}{\alpha}\right) = +\infty$$

In the following development, we will rewrite in this setting, the theorem of the previous chapter. This is crucial to extend the model to the case $\alpha \in \{0, 1\}$. Before going further, we can make the following assumption without loosing in generality, to simplify the setting.

Assumption 4.4. *The utility of a single man m is:*

$$u_m^0 = \Phi^0 + \sigma\varepsilon_m^0$$

and the utility of a single woman w is:

$$v_w^0 = \Phi^0 + \sigma\eta_w^0,$$

where Φ^0 is general parameter common to all single individuals, σ is a positive constant, ε_m^0 a stochastic term specific to man m and η_w a stochastic term specific to woman w.

This assumption states that:

$$A^0 = B^0 = \Phi^0$$

i.e all the single individuals have the same deterministic part denoted Φ^0. It is legitimate to assume that since we have explained above that actually we do not identify A^0 and B^0. As only the individual net surpluses $\bar{U}$ and $\bar{V}$ are identified, the parameters A^0 and B^0 are actually simply respectively a reference point for married men and a reference point for married women. We can then assume without loosing in generality the equality between these two reference points A^0 and B^0. One advantage of this assumption is the equivalence of the comparison between U and V and the comparison between $\bar{U}$ and $\bar{V}$ because we will have then $U - V = \bar{U} - \bar{V}$. By now, we fill consider then single men and single women have the same constant deterministic part of utility that is denoted Φ^0. The reader may wonder why we do not simply set this parameter to 0. Of course, we could, but for technical usefulness, we prefer to keep it as a parameter even if we are not going to identify it. The interesting trick will be to vanish it and that will turn out to be very simplifying.

From the relations (29) and (30) combined to the assumption stated above, we have $\delta_1 = \delta_2$. We will then denote

$$\delta = \delta_1 = \delta_2$$

The relations (29) and (30) take then both the form:

$$\Phi^0 = \sigma \ln \left(\frac{1-\alpha}{\alpha} \right) + \sigma \delta \tag{35}$$

The relations (25) and (25) become respectively:

$$U(x,y) = \Phi^0 + \sigma \ln \left(\frac{\alpha}{1-\alpha} \frac{\pi(x,y)}{f_0(x)} \right) \tag{36}$$

and

$$V(x,y) = \Phi^0 + \sigma \ln \left(\frac{\alpha}{1-\alpha} \frac{\pi(x,y)}{g_0(y)} \right) \tag{37}$$

The combination of (36) with (35) and the combination of (37) with (35) give respectively:

$$U(x,y) = \sigma \delta + \sigma \ln \left(\frac{\pi(x,y)}{f_0(x)} \right) \tag{38}$$

and

$$V(x,y) = \sigma \delta + \sigma \ln \left(\frac{\pi(x,y)}{g_0(y)} \right) \tag{39}$$

From (38) and (39), and using the fact that $U(x,y) + V(x,y) = S(x,y)$, we can state the following proposition.

Proposition 4.1. *Assume a stable matching market. Under Assumption 4.1, Assumption 4.3 and Assumption 4.4, we have:*

1. *for any* $x \in \mathcal{X}$, $y \in \mathcal{Y}$,

$$S(x,y) = 2\sigma\delta + \sigma \ln\left(\frac{(\pi(x,y))^2}{f_0(x)\, g_0(y)}\right)$$

2. *The systematic surplus of a man of attributes* x *from a matching with a woman of attributes* y *is such as:*

$$U(x,y) = \frac{1}{2}\left(S(x,y) - \sigma \ln\left(f_0(x)\right) + \sigma \ln\left(g_0(y)\right)\right)$$

3. *The systematic surplus of a woman of attributes* y *from a matching with a man of attributes* x *is such as:*

$$V(x,y) = \frac{1}{2}\left(S(x,y) + \sigma \ln\left(f_0(x)\right) - \sigma \ln\left(g_0(y)\right)\right)$$

4. *for any* $x \in \mathcal{X}$,

$$f_0(x) = \frac{e^{\delta}}{\left((1-\alpha)e^{\delta} + \alpha \int_{\mathcal{Y}} e^{\frac{U(x,y)}{\sigma}} dy\right)} f(x) \quad \textit{and} \quad f_1(x) = \frac{\int_{\mathcal{Y}} e^{\frac{U(x,y)}{\sigma}} dy}{\left((1-\alpha)e^{\delta} + \alpha \int_{\mathcal{Y}} e^{\frac{U(x,y)}{\sigma}} dy\right)} f(x)$$

5. *for any* $y \in \mathcal{Y}$,

$$g_0(y) = \frac{e^{\delta}}{\left((1-\alpha)e^{\delta} + \alpha \int_{\mathcal{X}} e^{\frac{V(x,y)}{\sigma}} dx\right)} g(y) \quad \textit{and} \quad g_1(y) = \frac{\int_{\mathcal{X}} e^{\frac{V(x,y)}{\sigma}} dx}{\left((1-\alpha)e^{\delta} + \alpha \int_{\mathcal{X}} e^{\frac{V(x,y)}{\sigma}} dx\right)} g(y)$$

6.
$$\delta = \ln\left(\int_{\mathcal{X}} \int_{\mathcal{Y}} (f_0(x))^{\frac{1}{2}} (g_0(y))^{\frac{1}{2}} \exp\left(\frac{S(x,y)}{2\sigma}\right) dx\, dy\right)$$

Proof. 1. The combination of (38) and (39) with the relation $U(x,y) + V(x,y) = S(x,y)$ leads directly to:

$$S(x,y) = 2\sigma\delta + \sigma \ln\left(\frac{(\pi(x,y))^2}{f_0(x)\, g_0(y)}\right)$$

2. From the relation above, we can deduce:
$$\ln(\pi(x,y)) = \frac{S(x,y) + \sigma\ln(f_0(x)) + \sigma\ln(g_0(y)) - (\delta_1+\delta_2)\sigma}{2\sigma}$$
We use this expression of $\ln(\pi(x,y))$ in (38) and we deduce the stated expression in the proposition for $U(x,y)$.

3. Analogously, we replace the expression of $\ln(\pi(x,y))$ in (39) and we obtain the expression of $V(x,y)$.

4. From the relation
$$U(x,y) = \sigma\delta + \sigma\ln\left(\frac{\pi(x,y)}{f_0(x)}\right)$$
we can deduce:
$$e^{\frac{U(x,y)}{\sigma}} f_0(x) = e^{\delta}\pi(x,y)$$
The feasibility of the optimal matching implies that:
$$\int_{\mathcal{Y}} \pi(x,y)dy = f_1(x)$$
So, we have:
$$f_0(x)\int_{\mathcal{Y}} e^{\frac{U(x,y)}{\sigma}} dy = e^{\delta} f_1(x)$$
We then use the relation:
$$f(x) = \alpha f_1(x) + (1-\alpha) f_0(x)$$
and we get the equation for $f_0(x)$:
$$f_0(x)\int_{\mathcal{Y}} e^{\frac{U(x,y)}{\sigma}} dy = e^{\delta}\left(\frac{f(x)-(1-\alpha)f_0(x)}{\alpha}\right)$$
and the solution is given by:
$$f_0(x) = \frac{e^{\delta}}{\left((1-\alpha)e^{\delta} + \alpha\int_{\mathcal{Y}} e^{\frac{U(x,y)}{\sigma}} dy\right)} f(x)$$
We deduce the expression of $f_1(x)$ by replacing the found expression of $f_0(x)$ in:
$$f_1(x) = \frac{f(x)-(1-\alpha)f_0(x)}{\alpha}$$

5. We do analogous reasoning for $g_0(y)$ and $g_1(y)$ as above.

6. We use the relation:

$$\ln(\pi(x,y)) = \frac{S(x,y) + \sigma\ln(f_0(x)) + \sigma\ln(g_0(y)) - 2\sigma\delta}{2\sigma}$$

   i.e

$$\pi(x,y)e^{\delta} = e^{\frac{S(x,y)}{2\sigma}}(f_0(x))^{\frac{1}{2}}(g_0(y))^{\frac{1}{2}}$$

   We then use the relation:

$$\int_{\mathcal{X}}\int_{\mathcal{Y}} \pi(x,y)dx\,dy = 1$$

   and we deduce directly δ.

■ Remark: We treat in Appendix, the case the deterministic part of the utility of single men A^0 and the deterministic part of the utility of single women B^0 are still constant but may be different. That is traduced by Assumption 4.2.

This proposition is actually an alternative version of the theorem stated in the previous chapter. But this version here has the advantage to clearly allow an extension to $\alpha \in \{0,1\}$. In the following development, we will make a quick remark about the case of nearly scarce matching i.e $\alpha = 0^+$ and will make a link with the approximation proposed in Chapter II for the surpluses U and V, then we will entirely and particularly focus on the interesting case of nearly full matching i.e $\alpha = 1^-$.

In Chapter II, we have that the probability of matching α at the equilibrium is:

$$\alpha = \int_{\mathcal{X}}\int_{\mathcal{Y}} \frac{e^{\frac{\bar{U}(x,y)}{\sigma}}}{1+\int_{\mathcal{Y}} e^{\frac{\bar{U}(x,y)}{\sigma}}dy} f(x)\,dx\,dy = \int_{\mathcal{X}}\int_{\mathcal{Y}} \frac{e^{\frac{\bar{V}(x,y)}{\sigma}}}{1+\int_{\mathcal{X}} e^{\frac{\bar{V}(x,y)}{\sigma}}dx} g(y)\,dx\,dy$$

That implies $0 < \alpha < 1$. This is why we will prefer the terminology of nearly scarce matching ($\alpha = 0^+$) and of nearly full matching ($\alpha = 1^-$). And as we have:

$$f(x) = \alpha f_1(x) + (1-\alpha)f_0(x) \quad \text{and} \quad g(y) = \alpha g_1(y) + (1-\alpha)g_0(y)$$

where $f(x)$ and $g(y)$ are respectively the density functions of X and Y, then the feasibility constraints will always be satisfied and in particular:

$$\int_{\mathcal{X}} f_0(x)dx = \int_{\mathcal{X}} f_1(x)dx = \int_{\mathcal{Y}} g_0(y)dy = \int_{\mathcal{Y}} g_1(y)dy = 1$$

This remark will be crucial particularly in the case of nearly scarce matching and nearly full matching.

Equilibrium in Nearly Scarce Matching: $\alpha = 0^+$

In the case at the equilibrium the probability of matching α is close to 0, the deterministic part of the utility of singles Φ^0 tends to $+\infty$ and so the individuals prefer to remain single and they do not participate to the marriage market. The direct consequence is the absolute lack of matching. In other words, in this situation, the social planer cannot increase anymore the social surplus by matching. It is optimal for all the individuals to remain single. So there is no participation to the marriage market and it occurs then on the market a general celibacy. But note that, as the joint surplus is assumed to be exogenous to the model, then, even if there is no formation of couples, it is possible to derive the theoretical surpluses $U(x, y)$ and $V(x, y)$. These functions represent the eventual surpluses partners in isolated cases of matching have at the equilibrium of this market. We may observe actually some isolated cases of matching but the measure of their set will be 0. To precise our idea, considering the market is infinite, then the number of matchings in this particular case of scarce matching will be finite, i.e the probability of marriage is 0. If the market is finite then there will be no marriage. As we assume the possibility of infinitude of the market, then marriage can occur on this market discontinuously and finitely. The optimal matching function is simply the density function limit to which $\pi(x, y)$ tends to when α is close 0. It will still be a density function in this case of scarce matching respecting feasibility constraints. From Proposition 4.1, we can deduce for $\alpha = 0^+$ at the equilibrium, we have:

1. For any $x \in \mathcal{X}$, $y \in \mathcal{Y}$,

$$\pi(x, y) = e^{-\delta}(f(x))^{\frac{1}{2}} \, (g(y))^{\frac{1}{2}} \, e^{\frac{S(x,y)}{2\sigma}}$$

2. The systematic surplus of a man of attributes x from a matching with a woman of attributes y is such as:

$$U(x,y) = \frac{1}{2}\left(S(x,y) - \sigma \ln\left(f(x)\right) + \sigma \ln\left(g(y)\right)\right)$$

3. The systematic surplus of a woman of attributes y from a matching with a man of attributes x is such as:

$$V(x,y) = \frac{1}{2}\left(S(x,y) + \sigma \ln\left(f(x)\right) - \sigma \ln\left(g(y)\right)\right)$$

4. for any $x \in \mathcal{X}$,

$$f_0(x) = f(x) \quad \text{and} \quad f_1(x) = e^{-\delta}\left(f(x)\right)^{\frac{1}{2}} \int_{\mathcal{Y}} \left(g(y)\right)^{\frac{1}{2}} e^{\frac{S(x,y)}{2\sigma}} dy$$

5. for any $y \in \mathcal{Y}$,

$$g_0(y) = g(y) \quad \text{and} \quad g_1(y) = e^{-\delta}\left(f(x)\right)^{\frac{1}{2}} \int_{\mathcal{X}} \left(g(y)\right)^{\frac{1}{2}} e^{\frac{S(x,y)}{2\sigma}} dx$$

6.

$$\delta = \ln\left(\int_{\mathcal{X}}\int_{\mathcal{Y}} \left(f(x)\right)^{\frac{1}{2}} \left(g(y)\right)^{\frac{1}{2}} e^{\frac{S(x,y)}{2\sigma}} dx\, dy\right)$$

Proof. We simply replace $\alpha = 0^+$ in Proposition 4.1. ■

In this particular state of the market all the unknown of the model are completely analytically determined. We can also remark that even there is no matching observed, the matching function $\pi(x,y)$ given by this corollary still satisfies the feasibility constraints. In fact we have:

$$\int_{\mathcal{Y}} \pi(x,y)dy = e^{-\delta}\left(f(x)\right)^{\frac{1}{2}} \int_{\mathcal{Y}} \left(g(y)\right)^{\frac{1}{2}} e^{\frac{S(x,y)}{2\sigma}} dy = f_1(x)$$

and

$$\int_{\mathcal{X}} \pi(x,y)dx = e^{-\delta}\left(g(y)\right)^{\frac{1}{2}} \int_{\mathcal{X}} \left(f(x)\right)^{\frac{1}{2}} e^{\frac{S(x,y)}{2\sigma}} dx = g_1(y)$$

The expression of δ guarantees the satisfaction of the constraint

$$\int_{\mathcal{X}}\int_{\mathcal{Y}} \pi(x,y)dx\,dy = 1.$$

One can also remark that the expressions of the surpluses $U(x, y)$ and $V(x, y)$ in the case of scarce matching are exactly the first order of approximation of the individual surpluses shown in the previous chapter in the general case.

In what follows, we will focus on a particular matching market in which nearly all the individuals are matched.

4.3 Nearly Full Matching and Quadratic Specification

4.3.1 General Equilibrium and Comparison with the Dupuy-Galichon Model

We consider an equally bipartite matching market verifying Assumptions 4.1, 4.3 and 4.4. In addition, we assume that nearly all the individuals are matched at the equilibrium. So here, the probability α is equal to 1^-. This has been treated by Dupuy and Galichon (2014). The equilibrium is given by the maximization of the social surplus. Proposition 4.1 gives the equilibrium and moreover we have:

1. for any $x \in \mathcal{X}$,

$$f_0(x) = \frac{e^{\delta}}{\int_{\mathcal{Y}} e^{\frac{U(x,y)}{\sigma}} dy} f(x) \quad \text{and} \quad f_1(x) = f(x)$$

2. for any $y \in \mathcal{Y}$,

$$g_0(y) = \frac{e^{\delta}}{\int_{\mathcal{X}} e^{\frac{V(x,y)}{\sigma}} dx} g(y) \quad \text{and} \quad g_1(y) = g(y)$$

Proof. We obtain these results by replacing $\alpha = 1^-$ in Proposition 4.1. ■

The expressions of the surpluses $\pi(x, y)$, $U(x, y)$, $V(x, y)$ and δ are given by Proposition 4.1. In what follows, we will proceed to a correspondence betwenn this model and the model of Dupuy-Galichon to prove that the model of the chapter 2 in the particular case of full matching coincides with the Dupuy-Galichon model.

Correspondence with the Dupuy-Galichon model: The following tables gives at left our results and at right the Dupuy-Galichon results in the case of nearly full matching.

This model	The Dupuy-Galichon model
$S(x,y)$	$\Phi(x,y)$
$\ln(f_0(x)) = \delta + \ln\left(\frac{f(x)}{\int_{\mathcal{Y}} e^{U(x,y)/\sigma} dy}\right)$	$a(x) = \sigma \ln\left(\int_{\mathcal{Y}} \frac{e^{U(x,y)/\sigma}}{f(x)} dy\right)$
$\ln(g_0(y)) = \delta + \ln\left(\frac{g(y)}{\int_{\mathcal{X}} e^{V(x,y)/\sigma} dx}\right)$	$b(y) = \sigma \ln\left(\int_{\mathcal{X}} \frac{e^{V(x,y)/\sigma}}{g(y)} dx\right)$
$\ln \pi(x,y) = \frac{1}{2\sigma}\left[S(x,y) + \sigma \ln(f_0(x)) + \sigma \ln(g_0(y))\right] - \delta$	$\ln \pi(x,y) = \frac{1}{2\sigma}\left[\Phi(x,y) - a(x) - b(y)\right]$
$U(x,y) = \frac{1}{2}\left[S(x,y) - \sigma \ln(f_0(x)) + \sigma \ln(g_0(y))\right]$	$U(x,y) = \frac{1}{2}\left[\Phi(x,y) + a(x) - b(y)\right]$
$V(x,y) = \frac{1}{2}\left[S(x,y) + \sigma \ln(f_0(x)) - \sigma \ln(g_0(y))\right]$	$V(x,y) = \frac{1}{2}\left[\Phi(x,y) - a(x) + b(y)\right]$

Where:

$$\delta = \ln\left(\int_{\mathcal{X}} \int_{\mathcal{Y}} (f_0(x))^{\frac{1}{2}} (g_0(y))^{\frac{1}{2}} \exp\left(\frac{S(x,y)}{2\sigma}\right)\right)$$

The link between the two models is given by the relations:

$$-\sigma \ln(f_0(x)) + \sigma\delta = a(x) \quad \text{and} \quad -\sigma \ln(g_0(y)) + \sigma\delta = b(y)$$

When we use these in the expressions at the left of the table we derive automatically the expressions at the right and reversely.

4.3.2 Parametric Inference

We specify parametrically the joint surplus $S(x,y)$ with a parameter θ belonging to a subset Θ of a real vectors subspace. We denote by $\mathcal{C} \subset \mathcal{H} \times \mathcal{F}$ the set of the couples. There exists no single on this market. We denote by π^θ the equilibrium matching density function corresponding to each particular θ. Considering a man m with characteristics x_m and a woman w with characteristics y_w, the likelihood of the union (m, w) is the product of the conditional probability the man m chooses the woman w with respect to $X = x_m$ with the conditional probability the woman w chooses the man m with respect to $Y = y_w$. That is:

$$\pi^\theta(y_w|x_m).\pi^\theta(x_m|y_w)$$

with

$$\pi^\theta(y_w|x_m) = \frac{\pi^\theta(x_m, y_w)}{f(x_m)} \quad \text{and} \quad \pi^\theta(x_m|y_w) = \frac{\pi^\theta(x_m, y_w)}{f(y_w)}$$

The likelihood L of all the unions is the product of all the product all the likelihoods of the unions. So we have:

$$\forall\, \theta \in \Theta,\ \ \mathrm{Log}\, L(\theta) = \sum_{(m,w)\in\mathcal{C}} \left[2\ln(\pi^\theta(x_m, y_w)) - \ln(f(x_m)) - \ln(g(y_w))\right]$$

We maximize the log-likelihood over the set Θ. Note the density functions f and g are assumed to be completely known. So we can withdraw them from the log-likelihood. And we then obtain a reduced log-likelihood which is:

$$\forall\, \theta \in \Theta,\ \ \mathrm{Log}\, \mathcal{L}(\theta) = \sum_{(m,w)\in\mathcal{C}} \ln \pi^\theta(x_m, y_w)$$

In the following, we will try to give some closed forms at the equilibrium of the model and the likelihood for the inference, by considering a quadratic specification for the joint surplus.

4.3.3 Analytical Solution of the Matching Equilibrium in Case of Quadratic Specification

In this subsection, we provide the analytical solution of the model in this particular case of full marriage and under some assumptions on the distributions of the characteristics. We assume X and Y to be gaussian and we specify a quadratic form for the net joint surplus $S(X, Y)$.

Assumption 4.5. *1.*

$$X = \begin{pmatrix} X^1 \\ . \\ . \\ . \\ X^p \end{pmatrix} \sim \mathcal{N}_p\left(0, \Sigma_X\right) \text{ and } Y = \begin{pmatrix} Y^1 \\ . \\ . \\ . \\ Y^q \end{pmatrix} \sim \mathcal{N}_q\left(0, \Sigma_Y\right)$$

2.

$$S(X,Y) = (X' \; Y')\, K \begin{pmatrix} X \\ Y \end{pmatrix}$$

where:

$$K = \begin{pmatrix} A & C' \\ C & B \end{pmatrix} \in \mathcal{M}_{n,n}\left(\mathbb{R}\right)$$

with: $n = p + q$, $A \in \mathcal{M}_{p,p}\left(\mathbb{R}\right)$, $B \in \mathcal{M}_{q,q}\left(\mathbb{R}\right)$ *and* $C \in \mathcal{M}_{q,p}\left(\mathbb{R}\right)$.

The following theorem describes analytically the equilibrium of this market.

Theorem 4.1. *Assume a stable nearly full matching verifying Assumption 4.1, Assumption 4.3, Assumption 4.4 and Assumption 4.5. Then,*

1. *the optimal matching* π *is the density function of a* $\mathcal{N}_{p+q}\left(0, \Sigma\right)$ *such that:*

   (a)

$$\Sigma^{-1} = \begin{pmatrix} \left(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)^{-1} & -\Sigma_X^{-1}\Sigma_{XY}\left(\Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right)^{-1} \\ -\Sigma_Y^{-1}\Sigma_{YX}\left(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)^{-1} & \left(\Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right)^{-1} \end{pmatrix}$$

   (b) *we also have:*

$$\begin{aligned} \Sigma^{-1} &= \begin{pmatrix} \Gamma_X & -\frac{1}{\sigma}C' \\ -\frac{1}{\sigma}C & \Gamma_Y \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2}\Sigma_X^{-1} + \left(\frac{1}{4}\Sigma_X^{-1} + \frac{1}{\sigma^2}C'\Sigma_Y C\right)^{\frac{1}{2}}\left(\Sigma_X^{\frac{1}{2}}\right)^{-1} & -\frac{1}{\sigma}C' \\ -\frac{1}{\sigma}C & \frac{1}{2}\Sigma_Y^{-1} + \left(\frac{1}{4}\Sigma_Y^{-1} + \frac{1}{\sigma^2}C\Sigma_X C'\right)^{\frac{1}{2}}\left(\Sigma_Y^{\frac{1}{2}}\right)^{-1} \end{pmatrix} \end{aligned}$$

(c) *and moreover*

$$\Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix} = \begin{pmatrix} \Sigma_X & \frac{1}{\sigma}\Gamma_X^{-1}C'\Sigma_Y \\ \frac{1}{\sigma}\Sigma_Y C \Gamma_X^{-1} & \Sigma_Y \end{pmatrix} = \begin{pmatrix} \Sigma_X & \frac{1}{\sigma}\Sigma_X C'\Gamma_Y^{-1} \\ \frac{1}{\sigma}\Gamma_Y^{-1} C \Sigma_X & \Sigma_Y \end{pmatrix}$$

2.

$$f_0 \textit{ is the density function of a Gaussian } \mathcal{N}_p\left(0, \left(\frac{2A + 2\sigma\Gamma_X}{\sigma}\right)^{-1}\right)$$

and

$$g_0 \textit{ is the density function of a Gaussian } \mathcal{N}_q\left(0, \left(\frac{2B + 2\sigma\Gamma_Y}{\sigma}\right)^{-1}\right)$$

3. $\forall\,(X,Y) \in \mathbb{R}^p \times \mathbb{R}^q$,

$$U(X,Y) = \frac{1}{2}X'(2A + \sigma\Gamma_X)X - \frac{\sigma}{2}Y'\Gamma_Y Y + X'C'Y - \frac{\sigma}{4}\ln\left|\frac{2A + 2\sigma\Gamma_X}{\sigma}\right| + \frac{\sigma}{4}\ln\left|\frac{2B + 2\sigma\Gamma_Y}{\sigma}\right| - \frac{\sigma(p-q)}{4}\ln(2\pi)$$

and

$$V(X,Y) = \frac{1}{2}Y'(2B + \sigma\Gamma_Y)Y - \frac{\sigma}{2}X'\Gamma_X X + Y'CX - \frac{\sigma}{4}\ln\left|\frac{2B + 2\sigma\Gamma_Y}{\sigma}\right| + \frac{\sigma}{4}\ln\left|\frac{2A + 2\sigma\Gamma_X}{\sigma}\right| + \frac{\sigma(p-q)}{4}\ln(2\pi)$$

4.

$$\delta = \frac{1}{4}\left(\ln\left|\frac{2A + 2\sigma\Gamma_X}{\sigma}\right| + \ln\left|\frac{2B + 2\sigma\Gamma_Y}{\sigma}\right| - \ln\left(\frac{|\Gamma_X||\Gamma_Y|}{|\Sigma_X||\Sigma_Y|}\right)\right)$$

Proof. The proof is given in the appendix. ■

Remark: We treat in Appendix the case the joint surplus has linear terms.

From the statements 1.a and 1.b of Theorem, we can remark that we obtain and interesting result of identification:

$$C = \sigma\Sigma_Y^{-1}\Sigma_{YX}\left(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)^{-1}$$

and equivalently,

$$C' = \sigma \Sigma_X^{-1} \Sigma_{XY} \left(\Sigma_Y - \Sigma_{YX} \Sigma_X^{-1} \Sigma_{XY} \right)^{-1}$$

With this result, we find exactly the Theorem 2 of Bojilov and Galichon (2016) identifying the affinity matrix of their model. Actually, the matrix $2C$ corresponds to the affinity matrix of the framework of Bojilov and Galichon.

The statement 1.c of Theorem 4.1 precises the matrix of covariance between X and Y:

$$\Sigma_{YX} = \frac{1}{\sigma} \Sigma_Y C \Gamma_X^{-1} = \frac{1}{\sigma} \Gamma_Y^{-1} C \Sigma_X \quad \text{and} \quad \Sigma_{XY} = \frac{1}{\sigma} \Gamma_X^{-1} C' \Sigma_Y = \frac{1}{\sigma} \Sigma_X C' \Gamma_Y^{-1}$$

Note that we have considered here for simplicity that the characteristics such that $\mathbb{E}(X) = \mathbb{E}(Y) = 0$. This can of course be extended to a generalized case.

The 'Dupuy-Galichon Constant' In the Dupuy and Galichon's (2014) model, they have proved the following expressions for the individual surpluses:

$$U(x, y) = \frac{\Phi(x, y) + a(x) - b(y)}{2} \quad \text{and} \quad V(x, y) = \frac{\Phi(x, y) - a(x) + b(y)}{2}$$

where $\Phi(x, y)$ is the joint surplus function, and $a(x)$ and $b(y)$ are potentials guaranteeing the feasibility of π; these potentials $a(x)$ and $b(y)$ exist, are unique and can be determined respectively up to two constants c_1 and c_2. Theorem 4.1 and the correspondence between this model and the Dupuy and Galichon's model allow us to provide the expressions of these potentials up to their respective constants:

$$a(x) = -\sigma \ln \left(f_0(x) \right) + c_1 \quad \text{and} \quad b(y) = -\sigma \ln \left(g_0(y) \right) + c_2$$

where the functions $f_0(x)$ and $g_0(y)$ are given analytically by Theorem 4.1. We can tell more about the constants c_1 and c_2. Actually, we can identify the sum of the constants $c_1 + c_2$. Following the correspondence of the Dupuy and Galichon's model and this model, we have $c_1 + c_2 = 2\sigma\delta$ and δ is given by Theorem 4.1. So we have:

$$c_1 + c_2 = \frac{\sigma}{2} \left(\ln \left| \frac{2A + 2\sigma \Gamma_X}{\sigma} \right| + \ln \left| \frac{2B + 2\sigma \Gamma_Y}{\sigma} \right| - \ln \left(\frac{|\Gamma_X||\Gamma_Y|}{|\Sigma_X||\Sigma_Y|} \right) \right)$$

where the matrices Γ_X and Γ_Y are given by Theorem 4.1. The exact determination of the constants c_1 and c_2 remains then on the value of $c_1 - c_2$. We will denote:

$$\mathrm{Cst}^{DG} = \frac{c_1 - c_2}{2}$$

and we will call it The Dupuy-Galichon Constant for the reason that, as we know the value of the sum $c_1 + c_2 = 2\sigma\delta$, one has:

$$c_1 = \frac{(c_1 + c_2)}{2} + \frac{(c_1 - c_2)}{2} = \sigma\delta + \mathrm{Cst}^{DG}$$

and

$$c_2 = \frac{(c_1 + c_2)}{2} - \frac{(c_1 - c_2)}{2} = \sigma\delta - \mathrm{Cst}^{DG}.$$

So the constants c_1 and c_2 are actually determined up to a constant that is the Dupuy-Galichon Constant Cst^{DG}. The value of this constant remains basically on the specification of the deterministic part of the utility of single individuals at the beginning of the market even if at the equilibrium no one has remained unmatched. More precisely, we have:

$$\mathrm{Cst}^{DG} = \sigma\left(\int_{\mathcal{X}}\int_{\mathcal{Y}} f_0(x)\exp\left(\frac{U(x,y)}{\sigma}\right)dx\,dy - \int_{\mathcal{X}}\int_{\mathcal{Y}} g_0(y)\exp\left(\frac{V(x,y)}{\sigma}\right)dx\,dy\right)$$

This can also be expressed as follows:

$$\mathrm{Cst}^{DG} = \sigma\left(\ln\left(\int_{\mathcal{X}}\int_{\mathcal{Y}} \exp\left(\frac{A^0(x)}{\sigma}\right)\pi(x,y)\,dx\,dy\right) - \ln\left(\int_{\mathcal{X}}\int_{\mathcal{Y}} \exp\left(\frac{B^0(y)}{\sigma}\right)\pi(x,y)\,dx\,dy\right)\right)$$

or equivalently here in the case of full matching, we have:

$$\int_{\mathcal{Y}} \pi(x,y)dy = f_1(x) = f(x) \quad \text{and} \quad \int_{\mathcal{X}} \pi(x,y)dx = g_1(y) = g(y)$$

then we can rewrite Cst^{DG} completely with deterministic terms:

$$\mathrm{Cst}^{DG} = \sigma\left(\ln\left(\int_{\mathcal{X}} \exp\left(\frac{A^0(x)}{\sigma}\right)f(x)\,dx\right) - \ln\left(\int_{\mathcal{Y}} \exp\left(\frac{B^0(y)}{\sigma}\right)g(y)\,dy\right)\right)$$

where $A^0(x)$ and $B^0(y)$ represent respectively the deterministic part of the utility of a single man with characteristic x and the deterministic part of the utility of a single woman with characteristic y. One implication of this identity is that, the Dupuy-Galichon Constant Cst^{DG} is actually exogenous. The specification of the deterministic parts of the utility of singles determines in advance the value this constant. In the case, we do not make any specification on the deterministic part of the utility of singles, Cst^{DG} becomes undetermined. It is the case in the Dupuy and Galichon's model (2014). In the case the deterministic parts of the utility of singles are constant functions A^0 and B^0 as stated in Assumption 4.2, then we simply have $\mathrm{Cst}^{DG} = A^0 - B^0$. Assumption 4.4 eases even things by assuming $A^0 = B^0 = \Phi^0$; it implies $\mathrm{Cst}^{DG} = 0$ and therefore:

$$c_1 = c_2 = \sigma\delta = \frac{\sigma}{4}\left(\ln\left|\frac{2A + 2\sigma\Gamma_X}{\sigma}\right| + \ln\left|\frac{2B + 2\sigma\Gamma_Y}{\sigma}\right| - \ln\left(\frac{|\Gamma_X||\Gamma_Y|}{|\Sigma_X||\Sigma_Y|}\right)\right)$$

where the matrices Γ_X and Γ_Y are given by Theorem 4.1.

As we can remark, the optimal distribution π depends only on the sub-matrix C of the matrix K. In what follows, we propose two ways to estimate it.

Parametric Inference

1. **Maximum Likelihood Estimation**

   We maximize over the set of the sub-matrices $C \in \mathcal{M}_{q,p}(\mathbb{R})$, the log-likelihood:

$$\text{Log } L(C) = -\sum_{(m,w)\in\mathcal{C}}\left[(X_m', Y_w')\,\Sigma^{-1}\begin{pmatrix} X_m \\ Y_w \end{pmatrix} + \ln|\Sigma|\right]$$

   with

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{2}\Sigma_X^{-1} + \left(\frac{1}{4}\Sigma_X^{-1} + \frac{1}{\sigma^2}C'\Sigma_Y C\right)^{\frac{1}{2}}\left(\Sigma_X^{\frac{1}{2}}\right)^{-1} & -\frac{1}{\sigma}C' \\ -\frac{1}{\sigma}C & \frac{1}{2}\Sigma_Y^{-1} + \left(\frac{1}{4}\Sigma_Y^{-1} + \frac{1}{\sigma^2}C\Sigma_X C'\right)^{\frac{1}{2}}\left(\Sigma_Y^{\frac{1}{2}}\right)^{-1} \end{pmatrix}$$

   To ease the computation of the likelihood, one may simply use:

$$\Sigma^{-1} = \begin{pmatrix} \left(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)^{-1} & -\frac{1}{\sigma}C' \\ -\frac{1}{\sigma}C & \left(\Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right)^{-1} \end{pmatrix}$$

by replacing the matrix Σ_{XY} by the empirical matrix of covariance:

$$\hat{\Sigma}_{XY} = \left(\frac{1}{|\mathcal{C}|} \sum_{(m,w)\in\mathcal{C}} X^i_m . Y^j_w \right)_{1\leq i\leq p, 1\leq j\leq q}$$

2. **Moment-based Estimation**

From the identification of C, by a first approach, we can have as estimator:

$$\hat{C} = \sigma \Sigma_Y^{-1} \hat{\Sigma}_{YX} \left(\Sigma_X - \hat{\Sigma}_{XY} \Sigma_Y^{-1} \hat{\Sigma}_{YX} \right)^{-1}$$

or

$$\hat{C}' = \sigma \Sigma_X^{-1} \hat{\Sigma}_{XY} \left(\Sigma_Y - \hat{\Sigma}_{YX} \Sigma_X^{-1} \hat{\Sigma}_{XY} \right)^{-1}$$

But note the matrix Σ_{XY} or Σ_{YX} depend actually on the parameter C. In fact, from the lemma, we have the relation:

$$\Sigma_{YX} = \frac{1}{\sigma} \Gamma_Y^{-1} C \Sigma_X \quad \text{and} \quad \Sigma_{XY} = \frac{1}{\sigma} \Gamma_X^{-1} C' \Sigma_Y$$

and the proposition precises:

$$\Gamma_X = \frac{1}{2}\Sigma_X^{-1} + \left(\frac{1}{4}\Sigma_X^{-1} + \frac{1}{\sigma^2} C' \Sigma_Y C \right)^{\frac{1}{2}} \left(\Sigma_X^{\frac{1}{2}} \right)^{-1}$$

and

$$\Gamma_Y = \frac{1}{2}\Sigma_Y^{-1} + \left(\frac{1}{4}\Sigma_Y^{-1} + \frac{1}{\sigma^2} C \Sigma_X C' \right)^{\frac{1}{2}} \left(\Sigma_Y^{\frac{1}{2}} \right)^{-1}$$

So we have:

$$\Sigma_{YX} = \frac{1}{\sigma} \left[\frac{1}{2}\Sigma_Y^{-1} + \left(\frac{1}{4}\Sigma_Y^{-1} + \frac{1}{\sigma^2} C \Sigma_X C' \right)^{\frac{1}{2}} \left(\Sigma_Y^{\frac{1}{2}} \right)^{-1} \right]^{-1} C \Sigma_X$$

and

$$\Sigma_{XY} = \frac{1}{\sigma} \left[\frac{1}{2}\Sigma_X^{-1} + \left(\frac{1}{4}\Sigma_X^{-1} + \frac{1}{\sigma^2} C' \Sigma_Y C \right)^{\frac{1}{2}} \left(\Sigma_X^{\frac{1}{2}} \right)^{-1} \right]^{-1} C' \Sigma_Y$$

We then estimate the matrix C by solving the equation:

$$\hat{\Sigma}_{YX} = \frac{1}{\sigma}\left[\frac{1}{2}\Sigma_Y^{-1} + \left(\frac{1}{4}\Sigma_Y^{-1} + \frac{1}{\sigma^2}\hat{C}\Sigma_X\hat{C}'\right)^{\frac{1}{2}}\left(\Sigma_Y^{\frac{1}{2}}\right)^{-1}\right]^{-1}\hat{C}\Sigma_X$$

or

$$\hat{\Sigma}_{XY} = \frac{1}{\sigma}\left[\frac{1}{2}\Sigma_X^{-1} + \left(\frac{1}{4}\Sigma_X^{-1} + \frac{1}{\sigma^2}\hat{C}'\Sigma_Y\hat{C}\right)^{\frac{1}{2}}\left(\Sigma_X^{\frac{1}{2}}\right)^{-1}\right]^{-1}\hat{C}'\Sigma_Y$$

A Particular but General Case of Specification

We assume that the joint surplus function is specified as follows:

$$S(X,Y) = X'.C.Y = \frac{1}{2}(X' \ \ Y')\begin{pmatrix} 0 & C' \\ C & 0 \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}, \quad \text{where} \ \ C \in \mathcal{M}_{q,p}(\mathbb{R})$$

where

$$X = \begin{pmatrix} X^1 \\ . \\ . \\ . \\ X^p \end{pmatrix} \sim \mathcal{N}_p\left(0, \Sigma_X\right) \ \text{ and } \ Y = \begin{pmatrix} Y^1 \\ . \\ . \\ . \\ Y^q \end{pmatrix} \sim \mathcal{N}_q\left(0, \Sigma_Y\right)$$

We can include in the components of X polynomials of the attributes of men, and in the components of Y polynomials of the attributes of women. In a general case where the attributes X and Y are not centered, i.e $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ can possibly be different from 0, we can consider for instance $X = (1\ x\ x^2)'$ and $Y = (1\ y\ y^2)'$, and then we have:

$$S(x,y) = a\,x^2 \, + \, b\,y^2 \, + \, c\,x\,y \, + \, a_1\,x \, + \, b_1\,y$$

We can write $S(x,y)$ at the form:

$$S(x,y) = (1\ x\ x^2)\begin{pmatrix} 0 & b_1 & b \\ a_1 & c & 0 \\ a & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ y \\ y^2 \end{pmatrix} = X'\begin{pmatrix} 0 & 0 & b \\ 0 & c & 0 \\ a & 0 & 0 \end{pmatrix}Y$$

So the specification we are considering here is a particular case of the general case we have presented above because it is its restriction with $A = 0$ and $B = 0$, but nevertheless this particular restriction is equivalent to the initial specification considered.

The optimal matching π is the density function of a Gaussian $\mathcal{N}_{p+q}(0, \Sigma)$ where

$$\Sigma^{-1} = \begin{pmatrix} \Gamma_X & -\frac{1}{2\sigma}C' \\ -\frac{1}{2\sigma}C & \Gamma_Y \end{pmatrix}$$

The individual surpluses are given by:

$$U(X,Y) = \frac{\sigma}{2}X'\Gamma_X X - \frac{\sigma}{2}Y'\Gamma_Y Y + \frac{1}{2}X'C'Y - \frac{\sigma}{4}\ln\left[\frac{|\Gamma_X|}{|\Gamma_Y|}\right] - \frac{\sigma(p-q)}{4}\ln(4\pi)$$

and

$$V(X,Y) = \frac{\sigma}{2}Y'\Gamma_Y Y - \frac{\sigma}{2}X'\Gamma_X X + \frac{1}{2}Y'CX + \frac{\sigma}{4}\ln\left[\frac{|\Gamma_X|}{|\Gamma_Y|}\right] + \frac{\sigma(p-q)}{4}\ln(4\pi)$$

where,

$$\Gamma_X = \frac{1}{2}\Sigma_X^{-1} + \left(\frac{1}{4}\Sigma_X^{-1} + \frac{1}{4\sigma^2}C'\Sigma_Y C\right)^{\frac{1}{2}}\left(\Sigma_X^{\frac{1}{2}}\right)^{-1} = \left(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)^{-1}$$

and

$$\Gamma_Y = \frac{1}{2}\Sigma_Y^{-1} + \left(\frac{1}{4}\Sigma_Y^{-1} + \frac{1}{4\sigma^2}C\Sigma_X C'\right)^{\frac{1}{2}}\left(\Sigma_Y^{\frac{1}{2}}\right)^{-1} = \left(\Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right)^{-1}$$

The matrix C is called the affinity matrix by Dupuy-Galichon. We can estimate this matrix by maximization of the log-likelihood and we have:

$$\hat{C}^{MLE} = Arg \max_{C \in \mathcal{M}_{q,p}(\mathbb{R})} - \sum_{(m,w)\in\mathcal{C}} \left[\left(X'_m \;\; Y'_w\right)\Sigma^{-1}\begin{pmatrix} X_m \\ Y_w \end{pmatrix} + \ln|\Sigma| \right]$$

with

$$\Sigma^{-1} = \begin{pmatrix} \left(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)^{-1} & -\frac{1}{2\sigma}C' \\ -\frac{1}{2\sigma}C & \left(\Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right)^{-1} \end{pmatrix}$$

We can also estimate by:

$$\hat{C} = \sigma \Sigma_Y^{-1} \hat{\Sigma}_{YX} \left(\Sigma_X - \hat{\Sigma}_{XY} \Sigma_Y^{-1} \hat{\Sigma}_{YX} \right)^{-1}$$

or

$$\hat{C}' = \sigma \Sigma_X^{-1} \hat{\Sigma}_{XY} \left(\Sigma_Y - \hat{\Sigma}_{YX} \Sigma_X^{-1} \hat{\Sigma}_{XY} \right)^{-1}$$

Uni-dimensional Case

We consider here that the characteristics follow standard gaussian distributions. Also for simplicity, we assume $\sigma = 1$. We obtain as results in this particular case the following results.

$$S(x,y) = (x,y) \begin{pmatrix} a & c \\ c & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

The optimal matching π that is the density function of (X,Y) and:

$$\pi(x,y) = \frac{1}{2\pi} \sqrt{\frac{1+\sqrt{1+4c^2}}{2}} \exp\left\{ -\frac{1}{2} (x \;\; y) \begin{pmatrix} \frac{1+\sqrt{1+4c^2}}{2} & -c \\ -c & \frac{1+\sqrt{1+4c^2}}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right\}$$

The individual surpluses are given by:

$$U(x,y) = \frac{1+4a+\sqrt{1+4c^2}}{4} x^2 - \frac{1+\sqrt{1+4c^2}}{4} y^2 + c\,x\,y - \frac{1}{4} \ln \left(\frac{|1+2a+\sqrt{1+4c^2}|}{|1+2b+\sqrt{1+4c^2}|} \right)$$

and

$$V(x,y) = \frac{1+4b+\sqrt{1+4c^2}}{4} y^2 - \frac{1+\sqrt{1+4c^2}}{4} x^2 + c\,x\,y + \frac{1}{4} \ln \left(\frac{|1+2a+\sqrt{1+4c^2}|}{|1+2b+\sqrt{1+4c^2}|} \right)$$

The functions $f_0(x)$ and $g_0(y)$ are given by:

$$f_0(x) = \sqrt{\frac{1+2a+\sqrt{1+4c^2}}{2\pi}} \exp\left\{ -\frac{1+2a+\sqrt{1+4c^2}}{2} x^2 \right\}$$

and

$$g_0(y) = \sqrt{\frac{1+2b+\sqrt{1+4c^2}}{2\pi}} \exp\left\{-\frac{1+2b+\sqrt{1+4c^2}}{2}y^2\right\}$$

The constant δ is given by:

$$\delta = \frac{1}{4}\ln\left(\left|1+2a+\sqrt{1+4c^2}\right|\right) + \frac{1}{4}\ln\left(\left|1+2b+\sqrt{1+4c^2}\right|\right) - \frac{1}{2}\ln\left(\left|\frac{1+\sqrt{1+4c^2}}{2}\right|\right)$$

A Particular Application: Golden Section We consider the following simple case. We assume $\sigma = 1$, and X and Y are assumed to be standard Gaussian variables, and the joint surplus is defined as:

$$S(x,y) = -(x-y)^2 = (x\ \ y)\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

Then the optimal matching joint density function $\pi(x,y)$ is the density function of the following two-dimensional gaussian:

$$\mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \varphi^{-1} \\ \varphi^{-1} & 1 \end{pmatrix}\right)$$

where

$$\varphi = \frac{1+\sqrt{5}}{2}$$

φ is called the golden section. This natural constant was considered by Greeks as the absolute perfect proportion. The optimal matching is given by:

$$\pi(x,y) = \frac{\sqrt{\varphi}}{2\pi}\exp\left\{-\frac{1}{2}(x\ \ y)\begin{pmatrix} \varphi & -1 \\ -1 & \varphi \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}\right\}$$

Moreover, the functions $f_0(x)$ and $g_0(y)$ are equal to the density function of a gaussian:

$$\mathcal{N}\left(0, \frac{1}{2}\varphi\right)$$

In other words:

$$f_0(x) = \frac{1}{\sqrt{\pi\varphi}} \exp\left\{-\frac{1}{\varphi}x^2\right\} = \sqrt{\frac{-1+\sqrt{5}}{2\pi}} \exp\left\{\frac{1-\sqrt{5}}{2}x^2\right\}$$

and

$$g_0(y) = \frac{1}{\sqrt{\pi\varphi}} \exp\left\{-\frac{1}{\varphi}y^2\right\} = \sqrt{\frac{-1+\sqrt{5}}{2\pi}} \exp\left\{\frac{1-\sqrt{5}}{2}y^2\right\}$$

The individual surpluses functions $U(x,y)$ and $V(x,y)$ are:

$$U(x,y) = -\frac{1}{2}(2-\varphi)x^2 - \frac{1}{2}\varphi y^2 + xy = -\frac{1}{4}(3-\sqrt{5})x^2 - \frac{1}{4}(1+\sqrt{5})y^2 + xy$$

and

$$V(x,y) = -\frac{1}{2}(2-\varphi)y^2 - \frac{1}{2}\varphi x^2 + xy = -\frac{1}{4}(3-\sqrt{5})y^2 - \frac{1}{4}(1+\sqrt{5})x^2 + xy$$

And finally, the constant δ from its expression in Theorem 4.1 can be computed and is equal here to:

$$\delta = \frac{1}{2}\ln\left(3-\sqrt{5}\right)$$

4.4 Conclusion

In this paper, we have considered a bipartite matching market with transferable utility, on which the two sides have the same weight. We use the framework developed in the chapter 2. We rewrite the results of the chapter 2 in view to extend explicitly the model to the case $\alpha \in \{0,1\}$ where α is the probability of matching on each of the two sides. The model even allows to solve analytically the equilibrium of the market in the case of scarce matching. But the main goal of the paper constitutes into investigating on closed for the equilibrium in a particular case of full matching. For that we have restricted our framework to a matching market without singles, with a quadratic specification of the joint surplus and Gaussian distributions of the observable characteristics. This question has been treated by Bojilov and Galichon (2016); on the basis of Dupuy and Galichon setting, they have proposed the analytical expression of the optimal matching and have identified the affinity matrix of their model for this particular matching market. We find exactly the same identification of the affinity matrix with Bojilov and Galichon; we also provide closed forms for the optimal matching that are as well similar to the ones found by Bojilov and Galichon. But at the difference of Bojilov and Galichon, in addition we provide the analytical expressions of the individual surpluses. Dupuy and Galichon (2014) have actually proved that the individual surpluses $U(x, y)$ and $V(x, y)$ are identified respectively up to a function $a(x)$ and up to a function $b(y)$ and these functions exist, are unique and can be determined up to a constant. In the approach we suggest here, we can entirely retrieve the functions $U(x, y)$ and $V(x, y)$. This result is an interesting novelty in the sense that for such models, until now no closed forms are proposed for the individual surpluses to our knowledge. In fact, Bojilov and Galichon (2016) have given the conditional distributions $X|Y$ and $Y|X$ but have not given the expressions of the functional constants $a(x)$ and $b(y)$ crucial to recover the entirely the expressions of the individual surpluses in their framework. We finally provide two approaches to estimate the model: by maximum likelihood and by moments based estimation. The framework we propose here is based on transferable utility; but these results can be applied in the case of full matching with quadratic specification of the surplus in case of non transferable utility since following Menzel (2015), matching equilibrium in such circumstances of Gaussian distributions and quadratic specification and with transferable utility is identical to the equilibrium in these same circumstances with non transferable utility.

References

[1] Bojilov & Galichon. 2016, Matching in Closed-Form: Equilibrium, identification, and comparative statics, Economic Theory, 61(4), 587-609. https://doi.org/10.1007/s00199-016-0961-8

[2] Chiappori, Salanié & Weiss. 2008, Assortative matching on the marriage market: a structural investigation, ADRES

[3] Choo, Eugene & Aloysius Siow. 2006, Who marries whom and why, Journal of Political Economy, Vol. 114, No. 1, pp. 175-201

[4] Dupuy, Arnaud & Alfred Galichon. 2014, Personality Traits and the Marriage Market, Journal of Political Economy, University of Chicago Press, vol. 122(6), p 1271-1319

[5] Gale & Shapley. 1962, College Admissions and the Stability of Marriage, Mathematical Association of America, Vol. 69, No. 1, pp. 9-15

[6] Galichon & Salanié. 2012, Cupid's invisible hand: social surplus and identification in matching models, Institut d'études politiques

[7] Lindenlaub. 2017, Sorting Multidimensional Types: Theory and Application, Review of Economic Studies, Oxford University Press, vol. 84(2), pages 718-789

[8] Menzel. 2015, Large Matching Markets as Two-Sided Demand Systems, Econometrica, Pages 897-941

[9] Shapley & Shubik. 1971, The assignment game I: The core, International Journal of Game Theory volume 1, p 111-130

[10] Volker Strassen. 1969, Gaussian Elimination is not Optimal, Numerische Mathematik, vol. 13, p. 354-356.

4.5 Appendix

4.5.1 Proof of Theorem 4.1

Let:

$$X = \begin{pmatrix} X^1 \\ . \\ . \\ . \\ X^p \end{pmatrix} \sim \mathcal{N}\left(0, \Sigma_X\right) \text{ and } Y = \begin{pmatrix} Y^1 \\ . \\ . \\ . \\ Y^q \end{pmatrix} \sim \mathcal{N}\left(0, \Sigma_Y\right)$$

and

$$S(X,Y) = (X',Y') \begin{pmatrix} A & C' \\ C & B \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

where

$$A \in \mathcal{M}_{p,p}(\mathbb{R}), \; B \in \mathcal{M}_{q,q}(\mathbb{R}) \text{ and } C \in \mathcal{M}_{q,p}(\mathbb{R})$$

Let us denote:

$$K = \begin{pmatrix} A & C' \\ C & B \end{pmatrix}$$

From Proposition 4.1, we have:

$$S(X,Y) = 2\sigma\delta + 2\sigma \ln(\pi(X,Y)) - \sigma \ln(f_0(X)) - \sigma \ln(g_0(Y))$$

Then we deduce:

$$\mathrm{Hess}(S)(X,Y) = 2K = 2\sigma\mathrm{Hess}(\ln(\pi))(X,Y) - \begin{pmatrix} \sigma\mathrm{Hess}(\ln(f_0))(X) & 0 \\ 0 & \sigma\mathrm{Hess}(\ln(g_0))(Y) \end{pmatrix}$$

i.e

$$2\begin{pmatrix} A & C' \\ C & B \end{pmatrix} = 2\sigma\mathrm{Hess}(\ln(\pi))(X,Y) - \begin{pmatrix} \sigma\mathrm{Hess}(\ln(f_0))(X) & 0 \\ 0 & \sigma\mathrm{Hess}(\ln(g_0))(Y) \end{pmatrix}$$

As the Hessian matrix of S is symmetric so is the Hessian matrix of $\ln(\pi)$. Moreover, X and Y are Gaussian and as the matching is assumed to be full, then the distributions of X and Y are exactly their conditional distributions with respect to marriage $M = 1$. So the feasibility constraints are traduced by:

$$\int_{\mathbb{R}^q} \pi(X,Y)dY = f(X) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma_X|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}X'\Sigma_X^{-1}X\right\}$$

and

$$\int_{\mathbb{R}^p} \pi(X,Y)dX = g(Y) = \frac{1}{(2\pi)^{\frac{q}{2}}|\Sigma_Y|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}Y'\Sigma_Y^{-1}Y\right\}$$

These feasibility constraints require the Hessian matrix of $\ln(\pi)$ to be constant.

The Hessian matrix of $\ln(\pi)$ is symmetric and constant so we can let it be at the form:

$$\mathrm{Hess}\left(\ln(\pi)\right)(X,Y) = -\begin{pmatrix} \Gamma_X & \Lambda' \\ \Lambda & \Gamma_Y \end{pmatrix}$$

where $\Gamma_X \in \mathcal{M}_{p,p}(\mathbb{R})$, $\Gamma_Y \in \mathcal{M}_{q,q}(\mathbb{R})$ and $\Lambda \in \mathcal{M}_{q,p}(\mathbb{R})$. We then have:

$$2\begin{pmatrix} A & C' \\ C & B \end{pmatrix} = -2\sigma\begin{pmatrix} \Gamma_X & \Lambda' \\ \Lambda & \Gamma_Y \end{pmatrix} - \begin{pmatrix} \sigma\mathrm{Hess}\left(\ln(f_0)\right)(X) & 0 \\ 0 & \sigma\mathrm{Hess}\left(\ln(g_0)\right)(Y) \end{pmatrix}$$

From this equation, we identify immediately:

$$\Lambda = -\frac{1}{\sigma}C$$

As the Hessian matrix $\ln(\pi)$ is constant, then π is a multidimensional Gaussian density function. We can then denote:

$$\mathrm{Hess}(\ln(\pi))(X,Y) = -\Sigma^{-1}$$

We deduce:

$$\pi(X,Y) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(X' \ \ Y')\Sigma^{-1}\begin{pmatrix} X \\ Y \end{pmatrix}\right\}$$

We then exploit the feasibility constraints and we have:

$$\int_{\mathbb{R}^q}\pi(X,Y)dY = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma_X|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}X'\Sigma_X^{-1}X\right\}$$

and

$$\int_{\mathbb{R}^p}\pi(X,Y)dX = \frac{1}{(2\pi)^{\frac{q}{2}}|\Sigma_Y|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}Y'\Sigma_Y^{-1}Y\right\}$$

Remark that:

$$(X' \ \ Y')\Sigma^{-1}\begin{pmatrix} X \\ Y \end{pmatrix} = X'\Gamma_X X + 2X'\Lambda' Y + Y'\Gamma_Y Y$$

$$= X'\left(\Gamma_X - \Lambda'\Gamma_Y^{-1}\Lambda\right)X + \left(Y + \Gamma_Y^{-1}\Lambda X\right)'\Gamma_Y\left(Y + \Gamma_Y^{-1}\Lambda X\right)$$

$$= Y'\left(\Gamma_Y - \Lambda\Gamma_X^{-1}\Lambda'\right)Y + \left(X + \Gamma_X^{-1}\Lambda'Y\right)'\Gamma_X\left(X + \Gamma_X^{-1}\Lambda'Y\right)$$

We then have by computation,

$$\int_{\mathbb{R}^q} \pi(X,Y)dY = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}|\Gamma_X|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}X'\left(\Gamma_X - \Lambda'\Gamma_Y^{-1}\Lambda\right)X\right\}$$

and

$$\int_{\mathbb{R}^p} \pi(X,Y)dX = \frac{1}{(2\pi)^{\frac{q}{2}}|\Sigma|^{\frac{1}{2}}|\Gamma_Y|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}Y'\left(\Gamma_Y - \Lambda\Gamma_X^{-1}\Lambda'\right)Y\right\}$$

Thus the feasibility constraints are traduced by:

$$\frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma_X|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}X'\Sigma_X^{-1}X\right\} = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}|\Gamma_X|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}X'\left(\Gamma_X - \Lambda'\Gamma_Y^{-1}\Lambda\right)X\right\}$$

and

$$\frac{1}{(2\pi)^{\frac{q}{2}}|\Sigma_Y|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}Y'\Sigma_Y^{-1}Y\right\} = \frac{1}{(2\pi)^{\frac{q}{2}}|\Sigma|^{\frac{1}{2}}|\Gamma_Y|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}Y'\left(\Gamma_Y - \Lambda\Gamma_X^{-1}\Lambda'\right)Y\right\}$$

This is equivalent to:

$$\begin{cases} |\Sigma||\Gamma_Y| = |\Sigma_X| \\ \Gamma_X - \Lambda'\Gamma_Y^{-1}\Lambda = \Sigma_X^{-1} \\ |\Sigma||\Gamma_X| = |\Sigma_Y| \\ \Gamma_Y - \Lambda\Gamma_X^{-1}\Lambda' = \Sigma_Y^{-1} \end{cases}$$

We can then deduce from this system the following equations:

$$\begin{pmatrix} (\Gamma_X - \Lambda'\Gamma_Y^{-1}\Lambda)^{-1} & -\Gamma_X^{-1}\Lambda'(\Gamma_Y - \Lambda\Gamma_X^{-1}\Lambda')^{-1} \\ -(\Gamma_Y - \Lambda\Gamma_X^{-1}\Lambda')^{-1}\Lambda\Gamma_X^{-1} & (\Gamma_Y - \Lambda\Gamma_X^{-1}\Lambda')^{-1} \end{pmatrix} = \begin{pmatrix} \Sigma_X & -\Gamma_X^{-1}\Lambda'\Sigma_Y \\ -\Sigma_Y\Lambda\Gamma_X^{-1} & \Sigma_Y \end{pmatrix}$$

and

$$\begin{pmatrix} (\Gamma_X - \Lambda'\Gamma_Y^{-1}\Lambda)^{-1} & -(\Gamma_X - \Lambda'\Gamma_Y^{-1}\Lambda)^{-1}\Lambda'\Gamma_Y^{-1} \\ -\Gamma_Y^{-1}\Lambda(\Gamma_X - \Lambda'\Gamma_Y^{-1}\Lambda)^{-1} & (\Gamma_Y - \Lambda\Gamma_X^{-1}\Lambda')^{-1} \end{pmatrix} = \begin{pmatrix} \Sigma_X & -\Sigma_X\Lambda'\Gamma_Y^{-1} \\ -\Gamma_Y^{-1}\Lambda\Sigma_X & \Sigma_Y \end{pmatrix}$$

And the two matrices in left members of the equations are exactly the inverse of the matrix Σ^{-1} in the matrix inversion algorithm developed by Strassen (1969). In other words, all these matrices are equal to Σ. This is traduced by:

$$\Sigma = \begin{pmatrix} \Gamma_X & \Lambda' \\ \Lambda & \Gamma_Y \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma_X & -\Gamma_X^{-1}\Lambda'\Sigma_Y \\ -\Sigma_Y\Lambda\Gamma_X^{-1} & \Sigma_Y \end{pmatrix} = \begin{pmatrix} \Sigma_X & -\Sigma_X\Lambda'\Gamma_Y^{-1} \\ -\Gamma_Y^{-1}\Lambda\Sigma_X & \Sigma_Y \end{pmatrix}$$

This proves the statement 1.c of the theorem , and this is equivalent to the result:

$$\Sigma_Y\Lambda\Gamma_X^{-1} = \Gamma_Y^{-1}\Lambda\Sigma_X$$

As we have:

$$\Lambda = -\frac{1}{\sigma}C$$

we obtain the equation:

$$\Sigma_Y C\Gamma_X^{-1} = \Gamma_Y^{-1}C\Sigma_X$$

We summary the main elementary results necessary for the proof in the following lemma.

Lemma 4.1. *Assume a stable full matching verifying Assumption 4.1, Assumption 4.3, Assumption 4.4 and Assumption 4.5. Then there exist an invertible $p \times p$ squared matrix Γ_X and an invertible $q \times q$ squared matrix Γ_Y such that the optimal matching joint density function $\pi(X,Y)$ verifies:*

$$\pi(X,Y) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(X' \;\; Y')\Sigma^{-1}\begin{pmatrix} X \\ Y \end{pmatrix}\right\}$$

where:

$$\Sigma = \begin{pmatrix} \Gamma_X & -\frac{1}{\sigma}C' \\ -\frac{1}{\sigma}C & \Gamma_Y \end{pmatrix}^{-1}$$

and

$$\begin{cases} \Gamma_X - \frac{1}{\sigma^2}C'\Gamma_Y^{-1}C = \Sigma_X^{-1} \\ \Gamma_Y - \frac{1}{\sigma^2}C\Gamma_X^{-1}C' = \Sigma_Y^{-1} \\ \Sigma_Y C\Gamma_X^{-1} = \Gamma_Y^{-1}C\Sigma_X \\ |\Sigma||\Gamma_Y| = |\Sigma_X| \\ |\Sigma||\Gamma_X| = |\Sigma_Y| \end{cases},$$

Proof. The proof is immediate from the development and the following remark. We have:

$$\left\{\begin{array}{l} \Gamma_X - \frac{1}{\sigma^2} C'\Gamma_Y^{-1}C = \Sigma_X^{-1} \\ \Gamma_Y - \frac{1}{\sigma^2} C\Gamma_X^{-1}C' = \Sigma_Y^{-1} \\ \int_{\mathbb{R}^p}\int_{\mathbb{R}^q} \pi(X,Y)dX\,dY = \int_{\mathbb{R}^q}\int_{\mathbb{R}^p} \pi(X,Y)dY\,dX = 1 \end{array}\right. \Longrightarrow \left\{\begin{array}{l} |\Sigma||\Gamma_Y| = |\Sigma_X| \\ |\Sigma||\Gamma_X| = |\Sigma_Y| \end{array}\right.$$

■

By equivalence, we have:

$$\left\{\begin{array}{l} \Gamma_X - \frac{1}{\sigma^2} C'\Gamma_Y^{-1}C = \Sigma_X^{-1} \\ \Gamma_Y - \frac{1}{\sigma^2} C\Gamma_X^{-1}C' = \Sigma_Y^{-1} \end{array}\right. \Longleftrightarrow \left\{\begin{array}{l} \Gamma_X - \frac{1}{\sigma^2} C'\Gamma_Y^{-1}C = \Sigma_X^{-1} \\ \Gamma_Y - \frac{1}{\sigma^2} C\Gamma_X^{-1}C' = \Sigma_Y^{-1} \\ \Sigma_Y C\Gamma_X^{-1} = \Gamma_Y^{-1}C\Sigma_X \end{array}\right.$$

$$\Longleftrightarrow \left\{\begin{array}{l} \Gamma_X\Sigma_X\Gamma_X - \Gamma_X = \frac{1}{\sigma^2} C'\Sigma_Y C \\ \Gamma_Y\Sigma_Y\Gamma_Y - \Gamma_Y = \frac{1}{\sigma^2} C\Sigma_X C' \end{array}\right.$$

$$\Longleftrightarrow \left\{\begin{array}{l} \left(\Gamma_X - \frac{1}{2}\Sigma_X^{-1}\right)\Sigma_X\left(\Gamma_X - \frac{1}{2}\Sigma_X^{-1}\right) = \frac{1}{4}\Sigma_X^{-1} + \frac{1}{\sigma^2} C'\Sigma_Y C \\ \left(\Gamma_Y - \frac{1}{2}\Sigma_Y^{-1}\right)\Sigma_Y\left(\Gamma_Y - \frac{1}{2}\Sigma_Y^{-1}\right) = \frac{1}{4}\Sigma_Y^{-1} + \frac{1}{\sigma^2} C\Sigma_X C' \end{array}\right.$$

The matrices Γ_X and Γ_Y are defined positive. We have:

$$\Gamma_X = \frac{1}{2}\Sigma_X^{-1} + \left(\frac{1}{4}\Sigma_X^{-1} + \frac{1}{\sigma^2}C'\Sigma_Y C\right)^{\frac{1}{2}}\left(\Sigma_X^{\frac{1}{2}}\right)^{-1}$$

and

$$\Gamma_Y = \frac{1}{2}\Sigma_Y^{-1} + \left(\frac{1}{4}\Sigma_Y^{-1} + \frac{1}{\sigma^2}C\Sigma_X C'\right)^{\frac{1}{2}}\left(\Sigma_Y^{\frac{1}{2}}\right)^{-1}$$

This proves the first point of the theorem as:

$$\Sigma^{-1} = \begin{pmatrix} \Gamma_X & -\frac{1}{\sigma}C' \\ -\frac{1}{\sigma}C & \Gamma_Y \end{pmatrix}$$

and with the Strassen algorithm, we have:

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \left(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)^{-1} & -\Sigma_X^{-1}\Sigma_{XY}\left(\Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right)^{-1} \\ -\Sigma_Y^{-1}\Sigma_{YX}\left(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)^{-1} & \left(\Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right)^{-1} \end{pmatrix}$$

In the next step of the proof, we will identify $f_0(X)$ and $g_0(Y)$. We have:

$$\begin{cases} 2A = -2\sigma\Gamma_X - \sigma \text{Hess} \ln\left(f_0\right)(X) \\ 2B = -2\sigma\Gamma_Y - \sigma \text{Hess} \ln\left(g_0\right)(Y) \end{cases} \iff \begin{cases} \text{Hess} \ln\left(f_0\right)(X) = \frac{-2A-2\sigma\Gamma_X}{\sigma} \\ \text{Hess} \ln\left(g_0\right)(Y) = \frac{-2B-2\sigma\Gamma_Y}{\sigma} \end{cases}$$

As we are in nearly full matching, we still have:

$$\int_{\mathbb{R}^p} f_0(X)dX = 1 \text{ and } \int_{\mathbb{R}^q} g_0(Y)dY = 1$$

and we deduce:

$$f_0(X) = \left|\frac{A+\sigma\Gamma_X}{\pi\sigma}\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}X'\left(\frac{2A+2\sigma\Gamma_X}{\sigma}\right)X\right\}$$

and

$$g_0(Y) = \left|\frac{B+\sigma\Gamma_Y}{\pi\sigma}\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}Y'\left(\frac{2B+2\sigma\Gamma_Y}{\sigma}\right)Y\right\}$$

We then have:

$$\ln\left(f_0(X)\right) = -\frac{1}{2}\left(X'\left(\frac{2A+2\sigma\Gamma_X}{\sigma}\right)X - \ln\left|\frac{A+\sigma\Gamma_X}{\pi\sigma}\right|\right)$$

and

$$\ln\left(g_0(Y)\right) = -\frac{1}{2}\left(Y'\left(\frac{2B+2\sigma\Gamma_Y}{\sigma}\right)Y - \ln\left|\frac{B+\sigma\Gamma_Y}{\pi\sigma}\right|\right)$$

From Proposition 4.1, we have:

$$U(X,Y) = \frac{1}{2}\left(S(X,Y) - \sigma\ln(f_0(X)) + \sigma\ln(g_0(Y))\right)$$

and

$$V(X,Y) = \frac{1}{2}\left(S(X,Y) + \sigma\ln(f_0(X)) - \sigma\ln(g_0(Y))\right)$$

The constant δ is given by:

$$\delta = \ln\left(\int_{\mathbb{R}^p}\int_{\mathbb{R}^q} e^{\frac{S(X,Y)}{2\sigma}}\left(f_0(X)\right)^{\frac{1}{2}}\left(g_0(Y)\right)^{\frac{1}{2}} dX\, dY\right)$$

With these formulas we derive the expressions of $U(X,Y)$, $V(X,Y)$ and δ.

4.5.2 General Case with $A^0 \neq B^0$

We will consider here that the deterministic part of the utility of single men is a constant A^0 and the deterministic part of utility of single women is a constant B^0. We assume that A^0 may be different from B^0.

Proposition 4.2. *Assume a stable matching market. Under Assumption 4.1, Assumption 4.2, and Assumption 4.3 we have:*

1. *for any $x \in \mathcal{X}$, $y \in \mathcal{Y}$,*

$$S(x,y) = \sigma(\delta_1 + \delta_2) + \sigma \ln \left(\frac{(\pi(x,y))^2}{f_0(x)\, g_0(y)} \right)$$

2. *The systematic surplus of a man of attributes x from a matching with a woman of attributes y is such as:*

$$U(x,y) = \frac{1}{2} \left(S(x,y) - \sigma \ln \left(f_0(x) \right) + \sigma \ln \left(g_0(y) \right) + \sigma(\delta_1 - \delta_2) \right)$$

3. *The systematic surplus of a woman of attributes y from a matching with a man of attributes x is such as:*

$$V(x,y) = \frac{1}{2} \left(S(x,y) + \sigma \ln \left(f_0(x) \right) - \sigma \ln \left(g_0(y) - \sigma(\delta_1 - \delta_2) \right) \right)$$

4. *for any $x \in \mathcal{X}$,*

$$f_0(x) = \frac{e^{\delta}}{\left((1-\alpha) e^{\delta} + \alpha \int_{\mathcal{Y}} e^{\frac{U(x,y)}{\sigma}} dy \right)} f(x) \quad \textit{and} \quad f_1(x) = \frac{\int_{\mathcal{Y}} e^{\frac{U(x,y)}{\sigma}} dy}{\left((1-\alpha) e^{\delta} + \alpha \int_{\mathcal{Y}} e^{\frac{U(x,y)}{\sigma}} dy \right)} f(x)$$

5. *for any $y \in \mathcal{Y}$,*

$$g_0(y) = \frac{e^{\delta}}{\left((1-\alpha) e^{\delta} + \alpha \int_{\mathcal{X}} e^{\frac{V(x,y)}{\sigma}} dx \right)} g(y) \quad \textit{and} \quad g_1(y) = \frac{\int_{\mathcal{X}} e^{\frac{V(x,y)}{\sigma}} dx}{\left((1-\alpha) e^{\delta} + \alpha \int_{\mathcal{X}} e^{\frac{V(x,y)}{\sigma}} dx \right)} g(y)$$

6.

$$\delta_1 + \delta_2 = 2 \ln \left(\int_{\mathcal{X}} \int_{\mathcal{Y}} (f_0(x))^{\frac{1}{2}} (g_0(y))^{\frac{1}{2}} \exp \left(\frac{S(x,y)}{2\sigma} \right) dx\, dy \right)$$

The proof is analogous to the proof of Proposition 4.1 with:

$$\delta_1 = \ln \left(\int_{\mathcal{X}} \int_{\mathcal{Y}} f_0(x) \exp \left(\frac{U(x,y)}{\sigma} \right) dx\, dy \right)$$

and

$$\delta_2 = \ln \left(\int_{\mathcal{X}} \int_{\mathcal{Y}} g_0(y) \exp \left(\frac{V(x,y)}{\sigma} \right) dx\, dy \right)$$

4.5.3 Nearly Full Matching and Quadratic Specification of the joint surplus including linear terms

We make the following assumption.

Assumption 4.6. *1.*

$$X = \begin{pmatrix} X^1 \\ . \\ . \\ . \\ X^p \end{pmatrix} \sim \mathcal{N}_p \left(0, \Sigma_X\right) \quad \textit{and} \;\; Y = \begin{pmatrix} Y^1 \\ . \\ . \\ . \\ Y^q \end{pmatrix} \sim \mathcal{N}_q \left(0, \Sigma_Y\right)$$

2.

$$S(X,Y) = (X' \; Y')\, K \begin{pmatrix} X \\ Y \end{pmatrix} + X'a + Y'b$$

where:

$$K = \begin{pmatrix} A & C' \\ C & B \end{pmatrix} \in \mathcal{M}_{n,n}\left(\mathbb{R}\right)$$

with: $n = p + q$, $A \in \mathcal{M}_{p,p}\left(\mathbb{R}\right)$, $B \in \mathcal{M}_{q,q}\left(\mathbb{R}\right)$ *and* $C \in \mathcal{M}_{q,p}\left(\mathbb{R}\right)$, *and* $a \in \mathbb{R}^p$, $b \in \mathbb{R}^q$.

Proposition 4.3. *Assume a stable nearly full matching. Under the Assumption 4.1, Assumption 4.3, Assumption 4.4 and Assumption 4.6, we have:*

1. *the optimal matching π is the density function of a $\mathcal{N}_{p+q}(0,\Sigma)$ such that:*

   (a)

$$\Sigma^{-1}=\begin{pmatrix} \left(\Sigma_X-\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)^{-1} & -\Sigma_X^{-1}\Sigma_{XY}\left(\Sigma_Y-\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right)^{-1} \\ -\Sigma_Y^{-1}\Sigma_{YX}\left(\Sigma_X-\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\right)^{-1} & \left(\Sigma_Y-\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right)^{-1} \end{pmatrix}$$

   (b) *we also have:*

$$\begin{aligned}\Sigma^{-1} &= \begin{pmatrix} \Gamma_X & -\frac{1}{\sigma}C' \\ -\frac{1}{\sigma}C & \Gamma_Y \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2}\Sigma_X^{-1}+\left(\frac{1}{4}\Sigma_X^{-1}+\frac{1}{\sigma^2}C'\Sigma_Y C\right)^{\frac{1}{2}}\left(\Sigma_X^{\frac{1}{2}}\right)^{-1} & -\frac{1}{\sigma}C' \\ -\frac{1}{\sigma}C & \frac{1}{2}\Sigma_Y^{-1}+\left(\frac{1}{4}\Sigma_Y^{-1}+\frac{1}{\sigma^2}C\Sigma_X C'\right)^{\frac{1}{2}}\left(\Sigma_Y^{\frac{1}{2}}\right)^{-1} \end{pmatrix}\end{aligned}$$

   (c) *and moreover*

$$\Sigma=\begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}=\begin{pmatrix} \Sigma_X & \frac{1}{\sigma}\Gamma_X^{-1}C'\Sigma_Y \\ \frac{1}{\sigma}\Sigma_Y C\Gamma_X^{-1} & \Sigma_Y \end{pmatrix}=\begin{pmatrix} \Sigma_X & \frac{1}{\sigma}\Sigma_X C'\Gamma_Y^{-1} \\ \frac{1}{\sigma}\Gamma_Y^{-1}C\Sigma_X & \Sigma_Y \end{pmatrix}$$

2.

$$f_0 \textit{ is the density function of a Gaussian } \mathcal{N}_p\left(\mu_X^0,\left(\frac{2A+2\sigma\Gamma_X}{\sigma}\right)^{-1}\right) \textit{ with}$$

$$\mu_X^0=-\left(2A+2\sigma\Gamma_X\right)^{-1}a$$

and

$$g_0 \textit{ is the density function of a Gaussian } \mathcal{N}_q\left(\mu_Y^0,\left(\frac{2B+2\sigma\Gamma_Y}{\sigma}\right)^{-1}\right) \textit{ with}$$

$$\mu_Y^0=-\left(2B+2\sigma\Gamma_Y\right)^{-1}b$$

3. $\forall\,(X,Y)\in\mathbb{R}^p\times\mathbb{R}^q$,

$$\begin{aligned}U(X,Y) &= \frac{1}{2}X'(2A+\sigma\Gamma_X)X-\frac{\sigma}{2}Y'\Gamma_Y Y+X'C'Y+X'a-\frac{\sigma}{4}\ln\left|\frac{2A+2\sigma\Gamma_X}{\sigma}\right| \\ &-\frac{1}{8}a'\left(A+\sigma\Gamma_X\right)^{-1}a+\frac{1}{8}b'\left(B+\sigma\Gamma_Y\right)^{-1}b+\frac{\sigma}{4}\ln\left|\frac{2B+2\sigma\Gamma_Y}{\sigma}\right|-\frac{\sigma(p-q)}{4}\ln(2\pi)\end{aligned}$$

and

$$V(X,Y) = \frac{1}{2}Y'(2B+\sigma\Gamma_Y)Y - \frac{\sigma}{2}X'\Gamma_X X + Y'CX + Y'b - \frac{\sigma}{4}\ln\left|\frac{2B+2\sigma\Gamma_Y}{\sigma}\right|$$
$$+\frac{1}{8}a'\left(A+\sigma\Gamma_X\right)^{-1}a - \frac{1}{8}b'\left(B+\sigma\Gamma_Y\right)^{-1}b + \frac{\sigma}{4}\ln\left|\frac{2A+2\sigma\Gamma_X}{\sigma}\right| + \frac{\sigma(p-q)}{4}\ln(2\pi)$$

4.
$$\delta = \frac{1}{4}\left(\ln\left|\frac{2A+2\sigma\Gamma_X}{\sigma}\right| + \ln\left|\frac{2B+2\sigma\Gamma_Y}{\sigma}\right| - \ln\left(\frac{|\Gamma_X||\Gamma_Y|}{|\Sigma_X||\Sigma_Y|}\right)\right)$$

Proof. The proof is then analogous to the proof of Theorem 4.1 and in addition we use the following equations:

$$\mathrm{Grad}(S)(X,Y) = \begin{pmatrix} 2AX+2C'Y+a \\ 2BY+2C'X+b \end{pmatrix} = 2\sigma\mathrm{Grad}\ln(\pi)(X,Y) - \begin{pmatrix} \sigma\mathrm{Grad}(\ln(f_0))(X) \\ \sigma\mathrm{Grad}(\ln(g_0))(Y) \end{pmatrix}$$

where
$$\mathrm{Grad}(\ln(\pi))(X,Y) = -\begin{pmatrix} \Gamma_X X + \Lambda' Y \\ \Gamma_Y Y + \Lambda X \end{pmatrix}$$

and:
$$\Lambda = -\frac{1}{\sigma}C$$

and we deduce that:

$$\sigma\mathrm{Grad}(\ln(f_0))(X) = -2\left(A+\sigma\Gamma_X\right)X - a$$

and
$$\sigma\mathrm{Grad}(\ln(g_0))(Y) = -2\left(B+\sigma\Gamma_Y\right)Y - b$$

Furthermore, we have:

$$\mathrm{Grad}(\ln(f_0))(X) = -\left(\frac{2A+2\sigma\Gamma_X}{\sigma}\right)X + \left(\frac{2A+2\sigma\Gamma_X}{\sigma}\right)\mu_X^0$$

and
$$\mathrm{Grad}(\ln(g_0))(Y) = -\left(\frac{2B+2\sigma\Gamma_Y}{\sigma}\right)Y + \left(\frac{2B+2\sigma\Gamma_Y}{\sigma}\right)\mu_Y^0$$

with

$$\mu_X^0 = -(2A + 2\sigma\Gamma_X)^{-1}a$$

and

$$\mu_Y^0 = -(2B + 2\sigma\Gamma_Y)^{-1}b$$

As we are in nearly full matching, we still have:

$$\int_{\mathbb{R}^p} f_0(X)dX = 1 \text{ and } \int_{\mathbb{R}^q} g_0(Y)dY = 1$$

And we deduce immediately the expressions of $f_0(X)$ and $g_0(Y)$. We then use Proposition 4.1 to deduce the expressions of the individual surpluses $U(X,Y)$, $V(X,y)$ and of δ. ■

5 Marriage Decision and Household Consumption Under Transferable Utility

Abstract

In this paper, we investigate on the implication of the stability of the marriage market on the household consumption. We assume transferability of utility and we consider the continuous model of marriage proposed in Chapter II with a collective approach of consumer. The model is quite simple. Individuals make a choice on the marriage market. Then, they consume alone in their household a private good and consume a public good with their eventual partner. Considering a parametric generalized quasi-linear form for consumption preferences, we derive the household demand functions for consumption from the equilibrium of the marriage market and we then estimate the parameters of the preferences on PSID data over the period 1968-2001. The estimation revealed a difference in the preferences parameters between men and women. More precisely, its showed that women have a higher taste for public good than men.

5.1 Introduction

In the present paper, we examine consumption and marriage decisions of individuals in a unified framework. To do so, we assume that the marriage market is stable according to Gale and Shapley (1962)'s definition and that the decision of the couple can be described by the collective approach, i.e, each individual has its own preference and the outcome of the decision process is Pareto efficient.

The collective approach has been introduced to study the behavior of multi-individual households by Chiappori (1988, 1992) and, since then, has rapidly gained in popularity. Bourguignon, Browning and Chiappori (2009) provide the main theoretical results for the case of constant prices. They show that the collective approach generates testable restrictions and that, if individual preferences are egoistic and consumed goods are private, the intrahousehold distribution of resources can be recovered. Browning and Chiappori (1998) incorporate prices. Blundell, Chiappori, and Meghir (2005) and Donni

(2012) extend the collective model to public goods. Chiappori and Ekeland (2007, 2009) provide a complete characterization of the collective model. Cherchye, De Rock, and Vermeulen (2011) and Cherchye et al. (2015) propose an alternative approach to the identification of the intrahousehold distribution of resources based on the theory of revealed preferences.

Empirical applications have shown that the intra-household distribution of resources depends on variables such as individual wages, prices, or distribution factors. See, for example, Bargain and Donni (2012), Bourguignon, Browning, and Chiappori (1994), Browning, Chiappori, and Lewbel (2013), Cherchye, De Rock, and Vermeulen (2012), Couprie, Peluso, and Trannoy (2010), Lewbel and Pendakur (2008), Lise and Seitz (2011) and Dunbar, Lewbel, and Pendakur (2013). However, the precise mechanism behind the intra-household distribution of resources is not clearly identified. As pointed out by Browning, Chiappori, and Weiss (2014, p. 122), the collective approach remains "agnostic" about the specific intra-household bargaining process that generates Pareto efficient outcomes. One of the rare structural models that attempt to explain how resources are shared among household members is proposed by Cherchye, Demuynck, De Rock, and Vermeulen (2017). They integrate the assumption of a stable marriage market with the collective consumption model to analyze the choice behavior of households. They then use a revealed preferences approach and identify the sharing rule of household resources, which is thus determined by the state of the marriage market.

The present paper is related to Cherchye, Demuynck, De Rock, and Vermeulen (2017)'s contribution but differs from it in at least two important aspects. Instead of using a revealed preferences approach, we construct a model of marriage market with transferable utility inspired from that of Choo and Siow (2006) and its sequels. In addition, we do not use information on household consumption but we identify consumption from the observation of matching patterns. More precisely, our approach is based on Dupuy and Galichon (2014) and its variation proposed in Chapter II. The individual utility is represented as the sum of a deterministic part and a stochastic part. The deterministic part corresponds to the consumption-based utility while the stochastic part can be seen as a "sympathy shock". From traditional results in the marriage market literature, the deterministic part is identified from the observation of matching patterns. The basic idea of the model is then simple. If we assume that the deterministic part, i.e., the consumption-based utility function of each individual, has a parametric form compatible with transferable utility, namely, a Generalized Quasi-Linear form (Bergstrom and Cornes, 1983), that depends on the consumption of a private and a public good, then the parameters of the consumption-based utility function can be recovered. Once the parameters are recovered, the consumption of the private and the public good can be identified.

The parametric specification of individual-based utility functions has at least three applications. (i) If single individuals have the same deterministic preferences as married individuals, then the consumption-based utility function of single individuals are identified from that of married individuals. This potentially allows comparing the changes in welfare of single individuals as well as that married individuals. To make a comparison with the results in Chapter II, the unidentified constant that represents the exchange rate between utility and money can be identified. (ii) The changes in endogamy observed during the second half of the 20th century can be structurally interpreted as a change in the individual taste for public goods together with a change in the distribution of individual incomes. (iii) From a more theoretical perspective, the identification of individual consumption suggests that observed matching patterns contain information on household consumption. It is a first step toward the unification of collective models and marriage models.

The theoretical model is illustrated by an empirical application using data from the PSID for the period 1968-2001. We find that, unsurprisingly, men and women have different parameters for the consumption-based utility functions. To be more precise, women have a higher taste for the public good than men. Consequently, the average gain from marriage is higher for men than for women. We also find that married individuals have a higher level of consumption and a higher level of surplus than singles. In addition, inequality is higher among single individuals than among married individuals. The study thus confirms the conclusion we drew in Chapter II, i.e., marriage reduces inequality in terms of consumption as well as surplus. Over the period that we examine, we observe a general increase in (consumption and surplus) inequality, with analogous dynamics to the incomes dynamics. Finally, we compute a parameter measuring the exchange rate between marriage surplus and money(which can be interpreted as the monetary value of marriage). This parameters varies over the examined period: it increases between 1968 and 1970, drastically decreases between 1970 and 1983, and it has been stable between 1983 and 2001. The method used to compute this value is an alternative to what we suggested in Chapter II. Although, the two measures give us very similar results in terms of dynamics over the time of the study.

The paper is structured as follows. In Section 1, we describe the marriage market and define the stability conditions in the sense of Gale and Shapley (1960). In Section 2, we identify the individual consumption-based utility of individual within households in function of the primitives of the model. In Section 3, we consider the case of the parametric Generalized Quasi-Linear utility function and we show that the parameters of the function can be retrieved. In Section 4, we present estimates of the model on the PSID data on the period 1968-2001.

5.2 Theoretical Model

In this subsection, we consider the model of marriage proposed in Chapter II and we take into account the consumption within households.

5.2.1 Marriage Market Description

The economy consists of a bipartite marriage market with transferable utility. Women constitute one side $\mathcal{W}$ and men $\mathcal{M}$ the other side. Each individual decides to match with an individual of the opposite side or remains single. The two sides of the market are possibly unequal and infinite. We denote by ν the proportion of men and the proportion of women by $1-\nu$. If a man and a woman get married, they consume individually a private good and consume together a public good and benefit from direct utility from partnership. If they remain single, they just have their own private consumption and consume alone the public good. The individuals choose their partner in function of their utility from consumption (private and public) they can potentially get from the union and depending also on their sympathy for the potential partner.

We assume that individuals are characterized by their income. The income of men is denoted X and the income of women Y. The support space of the income is $\mathbb{R}_+$. The random variables X and Y are assumed to be measurable. Let P and Q be respectively the conditional probability distribution of X with respect to $\mathcal{M}$ and the conditional probability distribution of Y with respect to $\mathcal{W}$. In addition we assume these conditional probability distributions have probability density functions with respect to the Lebesgue measure. Let f and g be respectively the conditional density function of X with respect to $\mathcal{M}$ and the conditional density function of Y with respect to $\mathcal{W}$. These functions are assumed to be completely exogenous to the model. In what follows, we will introduce some useful notations. We define the variable M that is the indication of the marriage status at the equilibrium. For any man i, M_i is equal to one if and only if the man i is matched and is equal to zero if he is single. And as well for any woman j, M_j is equal to one if and only if she is matched and 0 if not. We then denote respectively by α and β the probability to choose at the equilibrium a married man in the male population and the probability to choose at the equilibrium a married woman in the female population. As at the equilibrium, the number of matched men is equal to the number of matched women, we denote $\kappa = \alpha\nu = \beta(1-\nu)$. This is the proportion of married men in the whole population or the proportion of married women in the whole population. In the case men and women are in equal proportions in the total population, we have $\alpha = \beta$. In particular, the case $\alpha = \beta = 1$ corresponds to a full matching and this case has been treated by Dupuy and Galichon (2014).

Let f_0 be the conditional density function of X with respect to the event $M = 0$ and f_1 its conditional density function with respect to the event $M = 1$. Analogously, we define g_0 the conditional density function of Y with respect to the event $M = 0$ and g_1 its conditional density function with respect to the event $M = 1$. We define properly a feasible matching in the light of the chapter II.

Definition 5.1 (Feasible Matching)**.** *A feasible matching is a joint density function $\pi(x, y)$ defined on $\mathcal{X} \times \mathcal{Y}$ and a probability κ.*

The probability κ is the absolute probability of marriage on the market. Note that the conditional probabilities of marriage to the set of men and to the set of women are respectively:

$$\alpha = \frac{\kappa}{\nu} \text{ and } \beta = \frac{\kappa}{1-\nu}$$

In what follows, we will model the utility from marriage.

5.2.2 Marriage surplus, Utility from Consumption and Heterogeneity

The consumption decision is based on the incomes of the partners. The marriage decision of two potential partners depends on indirect consumption based utility and on the sympathy shocks. The acquaintance of a man i with income x with a woman j with income y creates for the man a sympathy shock $\varepsilon_i(y)$ and for the woman a sympathy shock $\eta_j(x)$. We will also assume that any single man i has an auto-sympathy shock ε_i^0 and any single woman j has an auto-sympathy shock η_j^0. We make the following assumptions on the utility and these will be the primitives of our model.

1. **Singles**

   We suppose that the direct utility $\tilde{u}_i^0$ of a single man i characterized by an income x, a private consumption a^0 and a public consumption G^0 is defined:

   $$\tilde{u}_i^0 = \tilde{U}(a^0, G^0) + \sigma\varepsilon_i^0$$

   where $\tilde{U}(.,.)$ is the consumption-based utility function of men. It is a Generalized Quasi Linear function with unknown general parameters including the taste for public good of men quantified by a general parameter. This GQL function will be defined later. We adopt a Generalized Quasi Linear form for the preferences to ensure the transferability of utility following Bergstrom and Cornes (1983). The term ε_i^0 is a stochastic term representing the specific gain of man i of being single and not depending on his consumption; and σ is a positive parameter. The stochastic part of the utility represents somehow an auto-sympathy term.

   Analogously, we model the direct utility of a woman j characterized by an income y and having a private consumption b^0 and a public consumption B^0 as follows:

$$\tilde{v}_j^0 = \tilde{V}(b^0, G^0) + \tau\eta_j^0$$

where $\tilde{V}(.,.)$ is the consumption based utility function of women. It is also a generalized quasi linear function with unknown parameters and also including the taste for public good of women quantified by one general parameter. The term η_j^0 is a stochastic term representing the specific gain of woman j of being single and not depending on her consumption; and τ is a positive parameter.

2. **Couples**

Now let us consider a couple (i, j) in $\mathcal{M} \times \mathcal{W}$. We denote respectively by a, b and G, the private consumption of the man and the private consumption of the woman and the public consumption within the household. We assume that the man i is characterized by x and the woman j by y. We suppose that the direct utility of the man i is:

$$\tilde{u}_i = \tilde{U}(a, G) + \sigma\varepsilon_i(y)$$

and the direct utility of the woman j is:

$$\tilde{v}_j = \tilde{V}(b, G) + \tau\eta_j(x)$$

The transferability of the utility implies that the joint utility $\tilde{\Phi}_{ij}$ of the couple is equal to the sum of the individual utilities:

$$\tilde{\Phi}_{ij} = \tilde{U}(a, G) + \tilde{V}(b, G) + \sigma\varepsilon_i(y) + \tau\eta_j(x)$$

where $\tilde{U}(a, G)$ and $\tilde{V}(b, G)$ are the direct consumption-based utilities of man i and woman j; $\varepsilon_i(y)$ and $\eta_j(x)$ are stochastic terms not depending on consumptions and representing respectively the sympathy shock of man i with the woman j with income y and the sympathy shock of woman j with man i with income x. The modeling of the joint utility from marriage is based on the separability assumption formalized by Galichon and Salanié (2015) and Chiappori, Salanié and Weiss (2017) that suggests that the random part of the joint utility does not depend on interactions between the unobserved characteristics of the partners conditionally on their characteristics. Using the terminology of Choo and Siow (2006), the consumption based utilities $\tilde{U}(a, G)$, $\tilde{V}(b, G)$ and the joint consumption based utility $\tilde{U}(a, G) + \tilde{V}(b, G)$ are respectively the systematic individual surplus of the man i and the systematic individual surplus of the woman j from this union, and the systematic joint surplus of the couple. The utilities and surpluses depend on consumption decisions. In the next subsection, we will show how the equilibrium we derive the indirect utilities that will exclusively depend on the characteristics of the individuals.

5.2.3 Equilibrium and Indirect Utilities

The equilibrium of the market is stable in the sense of Gale and Shapley (1962). A pair (i, j) of a man i and a woman j is matched to each other if and only if the woman j gets from the man i the maximum of the utility she can get on the market and a greater utility than her utility from single-hood. And conversely, the man i gets from the woman j the maximum utility he can reach on the market and greater than his utility from single-hood.

Assumption 5.1.

At the equilibrium,

1. *There is no matched person who would prefer be single.*
2. *There is no pair of persons non matched to each other who would rather form a new union.*

The consumption choices are made so that the direct utilities are maximized under the constraint budget and these choices provide the indirect utilities. Conditionally to the indirect utilities, the individuals will then make their matching choices. So, in what follows, we will first formalize the consumption decision and we will then focus on the matching equilibrium. Let us consider a single man i of income x facing the maximization problem:

$$\max_{a^0,G^0} \tilde{U}(a^0, G^0) + \sigma\varepsilon_i^0 \text{ under the constraint } a^0 + G^0 \leq x$$

The prices are normalized to 1 and relatives prices are constant. As the stochastic term does not depend on consumption, the indirect surplus of this single man i with income x is then:

$$U^0(x) = \max_{a^0+G^0 \leq x} \left(\tilde{U}(a^0, G^0)\right)$$

We derive his indirect utility as:

$$u_i^0(x) = U^0(x) + \sigma\varepsilon_i^0$$

Analogously, considering a single woman j with income y, her maximization problem is:

$$\max_{b^0,G^0} \tilde{V}(b^0, G^0) + \tau\eta_j^0 \text{ under the constraint } b^0 + G^0 \leq y$$

As the stochastic term does not depend on consumption, the indirect surplus of this single woman j with income y is then:

$$V^0(y) = \max_{b^0+G^0 \leq y} \left(\tilde{V}(b^0, G^0)\right)$$

and then her indirect utility is:

$$v_j^0(y) = V^0(y) + \tau\eta_j^0$$

Now, we focus on the consumption decision of couples. Let us consider a man i with income x matched to a woman j with income y. As the utility is assumed to be transferable, their joint utility $\tilde{\Phi}_{ij}$ is the sum the individual utilities $\tilde{u}_i$ and $\tilde{v}_j$ and the joint maximization of the individual utilities $\tilde{u}_i$ and $\tilde{v}_j$ is equivalent to the maximization of the joint utility $\tilde{\Phi}_{ij}$. So the decision of consumption within the household is given by: We denote by $S(x, y)$ the indirect joint consumption based utility that is the joint surplus.

$$S(x, y) = \max_{a,b,G} \tilde{U}(a, G) + \tilde{V}(b, G) \quad \text{under the constraint} \quad a + b + G \leq x + y$$

The indirect joint utility is then expressed as follows:

$$\Phi_{ij}(x, y) = S(x, y) + \sigma\varepsilon_i(y) + \tau\eta_j(x)$$

As formulated by Galichon and Salanié (2015) and in Chapter II, from the condition of the equilibrium on the marriage market, we can set the following proposition characterizing the decision of marriage and integrating the decision of consumption of individuals.

Proposition 5.1. *The stable matching is such that, under the assumptions stated above, for any man i of attributes x, and for a woman j of attributes y,*

$$u_i(x) = \max\left(\max_{j\in\mathcal{W}}(U(x, y_j)) + \sigma\varepsilon_i(y_j), U^0(x) + \sigma\varepsilon_i^0\right)$$

and

$$v_j(y) = \max\left(\max_{i\in\mathcal{M}}(V(x_i, y)) + \tau\eta_j(x_i), V^0(y) + \tau\eta_j^0\right)$$

for some functions $U(x, y)$ and $V(x, y)$ such that:

$$U(x, y) + V(x, y) = S(x, y)$$

Proof. For the proof, see Galichon and Salanié (2015) and Chapter II. ∎

So in the case man i with income x is matched to woman j with income y, he gets from this union an utility:

$$u_i = U(x, y) + \sigma\varepsilon_i(y)$$

and the woman gets from this union an utility:

$$v_j = V(x, y) + \tau\eta_j(y)$$

The functions $U(x, y) - U^0(x)$ and $V(x, y) - V^0(y)$ are to be identified. As soon as, we identify them, we can retrieve the household consumption vector (a, b, G) from the following three equations.

$$\tilde{U}(a, G) = U(x, y), \;\; \tilde{V}(b, G) = V(x, y) \;\; \text{and} \;\; a + b + G = x + y$$

5.2.4 Identification of the sharing rule

For simplification, we adopt the following notations:

Notation 5.1.

$$\bar{S}(x, y) = S(x, y) - U^0(x) - V^0(y),$$

$$\bar{U}(x, y) = U(x, y) - U^0(x) \;\; \textit{and} \;\; \bar{V}(x, y) = V(x, y) - V^0(y)$$

where $\bar{U}(x, y)$ and $\bar{V}(x, y)$ represents the net surplus of married men of characteristics x matched to women with characteristics y relatively to single men with their characteristics. Actually, $\bar{U}(x, y)$ is the net gain in consumption utility of a man i with characteristics x for marrying a woman with characteristics y instead of remaining single. In the case the man loses in consumption utility, then his decision of marriage is somehow very motivated by a high sympathy shock $\varepsilon_i(y)$. Analogously, $\bar{V}(x, y)$ is the net gain in consumption utility of women with characteristics y for matching with men with characteristics x than remaining single. $\bar{S}(x, y)$ is the total net gain of the couple in consumption utility. We also adopt some notations in the light of Galichon and Salanié (2015). Let us consider a man i with income x and a woman j with income y. Their respective equilibrium utilities are $u_i(x)$ and $v_j(y)$. We denote $\bar{u}_i = u_i - U^0$ and $\bar{v}_j = v_j - V^0$. We also denote:

$$G_x(\bar{U}) = \mathbb{E}\left(\bar{u}_i|x\right) \;\; \text{and} \;\; H_y(\bar{V}) = \mathbb{E}(\bar{v}_j|y).$$

Let (π, κ) be the optimal matching. Following Galichon and Salanié (2015) and Chapter II, we have:

$$\frac{\partial}{\partial \bar{U}} G_x(\bar{U}) = \frac{\alpha \pi(x,y)}{f(x)} \quad \text{and} \quad \frac{\partial}{\partial \bar{V}} H_y(\bar{V}) = \frac{\beta \pi(x,y)}{g(y)}$$

This result is given by the envelop theorem. In the light of Galichon and Salanié (2014) and Dupuy and Galichon (2014), we know that the equilibrium is ensured here by the maximization of the social surplus $\mathcal{W}$ which is the total expected utility and this is equivalent to the maximization of the net social surplus that is:

$$\bar{\mathcal{W}} = \nu \mathbb{E}[\bar{u}_i] + (1-\nu)\mathbb{E}[\bar{v}_j]$$

This can be rewritten as follows:

$$\bar{\mathcal{W}} = \nu \int_0^{+\infty} G_x(\bar{U}) f(x) dx + (1-\nu) \int_0^{+\infty} H_y(\bar{V}) g(y) dy$$

We denote by $G_x^*\left(\frac{\kappa\pi}{f}\right)$ the conjugate dual of $G_x(\bar{U})$ and by $H_y^*\left(\frac{\kappa\pi}{g}\right)$ the conjugate dual of $H_y(\bar{V})$ by the Legendre-Frenchel transformation. We have:

$$G_x^*\left(\frac{\kappa\pi}{f}\right) = \max_{\bar{U}(x,.)} \left(\int_0^{+\infty} \frac{\kappa\pi(x,y)}{f(x)} \bar{U}(x,y) dy - \nu G_x(\bar{U}) \right)$$

and

$$H_y^*\left(\frac{\kappa\pi}{g}\right) = \max_{\bar{V}(.,y)} \left(\int_0^{+\infty} \frac{\kappa\pi(x,y)}{g(y)} \bar{V}(x,y) dx - (1-\nu) H_y(\bar{V}) \right)$$

For any feasible matching (π, κ), as Galichon and Salanié (2015), we showed in Chapter II, that the corresponding net social surplus can be decomposed into two parts:

$$\bar{\mathcal{W}}(\pi,\kappa) = \kappa \int_0^{+\infty} \int_0^{+\infty} \pi(x,y) \bar{S}(x,y) dx\, dy - \mathcal{E}(\pi,\kappa)$$

where

$$\mathcal{E}(\pi,\kappa) = \int_0^{+\infty} G_x^*\left(\frac{\kappa\pi}{f}\right) f(x) dx + \int_0^{+\infty} H_y^*\left(\frac{\kappa\pi}{g}\right) g(y) dy$$

The optimal matching (π, κ) is such that $\mathcal{W}(\pi, \kappa)$ is the maximum social surplus other the set of feasible matchings. The term $\mathcal{E}(\pi, \kappa)$ is an entropy and is linked to the variation of the stochastic terms ε and η. In the case of very weak variation, the optimal matching maximizes simply the total joint surplus.

The maximization of the total social surplus lead us to the identification of $U(x, y)$ and $V(x, y)$. The main importance of this identification is the explanation of the sharing of the resources within the household. The following proposition formalizes that:

Proposition 5.2. *Assume a stable matching (π, κ). For any stable couple with characteristic (x, y), their consumptions (a, b, G) are such that:*

$$\tilde{U}(a(x, y), G(x, y)) = U^0(x) + \frac{\partial}{\partial(\kappa\pi/f)} G_x^* \left(\frac{\kappa\pi}{f}\right)(y)$$

and

$$\tilde{V}(b(x, y), G(x, y)) = V^0(y) + \frac{\partial}{\partial(\kappa\pi/g)} H_y^* \left(\frac{\kappa\pi}{g}\right)(x)$$

and

$$a(x, y) + b(x, y) + G(x, y) = x + y$$

Proof. As showed by Galichon & Salanié (2015) and in Chapter II, from the Envelop Theorem, we have:

$$\bar{U}(x, y) = \frac{\partial}{\partial(\kappa\pi/f)} G_x^* \left(\frac{\kappa\pi}{f}\right)(y)$$

and

$$\bar{V}(x, y) = \frac{\partial}{\partial(\kappa\pi/g)} H_y^* \left(\frac{\kappa\pi}{g}\right)(x)$$

We get the proposition from the fact that:

$$\tilde{U}(a, G) = \bar{U}(x, y) + U^0(x) \text{ and } \tilde{V}(b, G) = \bar{V}(x, y) + V^0(y)$$

■

We remark that this proposition gives a system of three equations with three unknowns that are the individual private consumptions a, b and the public consumption G. The computation of the surpluses $U(x, y)$ and $V(x, y)$ will give the expressions of the consumptions in function of the characteristics (x, y) of the couple. For this, we will consider a particular specification for the random processes $\varepsilon_i(y)$, ε_i^0, $\eta_j(x)$ and η_j^0. Precisely, in the following development we will consider the Dupuy and Galichon (2014) framework.

5.3 Specification of Heterogeneity and of Consumption Preferences

5.3.1 The Dupuy-Galichon Framework

We have assumed till now that from acquaintances, partners can have a sympathy term in the utility they get from their union. In what follows, we will specify this sympathy term. Considering a man i with income x, and a woman with income y, the random term ε_i^0 is assumed to have a standard Gumbel distribution conditionally to x and the random term η_j^0 is as well assumed to have standard distribution conditionally to y. Following Dupuy and Galichon (2014), we assume that each individual meets subset of the opposite side of the market. From any acquaintance k of characteristics y_k, man i gets the sympathy shock ε_i^k and the process $\left\{(y_k, \varepsilon_i^k), k \in \mathbb{N}^*\right\}$ is a Poisson process on the space $\mathbb{R}_+$ with intensity $e^{-\varepsilon-\gamma} dy\, d\varepsilon$. And in the same way, we assume that woman j gets from any acquaintance l with characteristics x_l the sympathy shock η_j^l and the process $\left\{(x_l, \eta_j^l), l \in \mathbb{N}^*\right\}$ is a Poisson process on the space $\mathbb{R}_+$ with intensity $e^{-\eta-\gamma} dx\, d\eta$. From an acquaintance $k \in \mathbb{N}^*$, a man i with characteristics x gets the utility:

$$u_i^k = U(x, y_k) + \sigma \varepsilon_i^k$$

And the woman j with characteristics y gets from an acquaintance $l \in \mathbb{N}^*$ the utility:

$$v_j^l = V(x_l, y) + \tau \eta_j^l$$

We assume that the term ε_i^0 is independent from the different acquaintances k of the man i, and we also assume that the term η_j^0 is independent from the different acquaintances l of the woman j.

With this assumption, we can derive the distribution of $u_i(x)$ and of $v_j(y)$.

Proposition 5.3. *Assume a stable matching. Then, for any man i, whose utility at the equilibrium is $u_i(x)$, and for any woman j whose utility at the equilibrium is $v_j(y)$, we have:*

$$u_i(x) \sim Gumbel\left(U^0(x) - \gamma\sigma + \sigma \ln\left(1 + \int_0^{+\infty} e^{\frac{\tilde{U}(x,y)}{\sigma}} dy\right), \sigma\right)$$

and

$$v_j(y) \sim Gumbel\left(V^0(y) - \gamma\tau + \tau\ln\left(1 + \int_0^{+\infty} e^{\frac{\bar{V}(x,y)}{\tau}} dx\right), \tau\right)$$

Proof. Let us consider a couple (i, j) with characteristics (x, y). We have:

$$u_i(x) = \max\left(\max_{k\in\mathcal{W}}\left(U(x, y_k) + \sigma\varepsilon_i^k, U^0(x) + \sigma\varepsilon_i^0\right)\right.$$

where the process (y_k, ε_i^k) is following a Poisson process of intensity $e^{\varepsilon-\gamma}d\varepsilon dy$; and

$$v_j(y) = \max\left(\max_{l\in\mathcal{M}}\left(V(x_l, y) + \tau\eta_j^l\right), V^0(y) + \tau\eta_j^0\right)$$

where the process (x_l, η_j^l) is following a Poisson process of intensity $e^{\eta-\gamma}d\eta dx$.

Using the independence of ε_i^0 from the process (y_k, ε_i^k), we obtain:

$$F_{u(x)}(t) = \exp\left(-\exp\left(\frac{1}{\sigma}\left(U^0(x) - \gamma\sigma + \sigma\ln\left(1 + \int_0^{+\infty}\exp\left(\frac{U(x,y)}{\sigma}\right)dy - t\right)\right)\right)\right)$$

and similarly, with the independence of η_j^0 from the process (x_l, η_j^l), we have:

$$F_{v(y)}(t) = \exp\left(-\exp\left(\frac{1}{\tau}\left(V^0(y) - \gamma\tau + \tau\ln\left(1 + \int_0^{+\infty}\exp\left(\frac{V(x,y)}{\tau}\right)dx - t\right)\right)\right)\right)$$

These distribution functions are Gumbel distributions.

■

The following corollary gives the relation between the optimal matching (π, κ) and the individual surpluses $U(x, y)$ and $V(x, y)$.

Corollary 5.1. *Assume a stable matching (π, κ). We have:*

$$\exp\left(\frac{\bar{U}(x,y)}{\sigma}\right) = \frac{\alpha}{1-\alpha}\frac{\pi(x,y)}{f_0(x)}$$

and

$$\exp\left(\frac{\bar{V}(x,y)}{\tau}\right) = \frac{\beta}{1-\beta}\frac{\pi(x,y)}{g_0(y)}$$

Proof. From the previous proposition, the expected value of utility obtained by a man i with characteristcs x is:

$$G_x(\bar{U}) = \sigma \ln \left(1 + \int_0^{+\infty} e^{\frac{\bar{U}(x,y)}{\sigma}} dy\right)$$

and the expected value of utility of woman j is:

$$H_y(\bar{V}) = \tau \ln \left(1 + \int_0^{+\infty} e^{\frac{\bar{V}(x,y)}{\tau}} dx\right)$$

We recall the definition of the conjugates G_x^* and H_y^*:

$$G_x^*\left(\frac{\kappa\pi}{f}\right) = \max_{\bar{U}(x,.)}\left(\int_0^{+\infty} \frac{\kappa\pi(x,y)}{f(x)}\bar{U}(x,y)dy - \nu G_x(\bar{U})\right)$$

and

$$H_y^*\left(\frac{\kappa\pi}{g}\right) = \max_{\bar{V}(.,y)}\left(\int_0^{+\infty} \frac{\kappa\pi(x,y)}{g(y)}\bar{V}(x,y)dx - (1-\nu)H_y(\bar{V})\right)$$

From the envelop theorem we have:

$$\frac{\partial}{\partial \bar{U}} G_x(\bar{U}) = \frac{\alpha\pi(x,y)}{f(x)} \quad \text{and} \quad \frac{\partial}{\partial \bar{V}} H_{(y)}(\bar{V}) = \frac{\beta\pi(x,y)}{g(y)}$$

We then have:

$$\frac{\exp(\bar{U}(x,y)/\sigma)}{1 + \int_0^{+\infty} \exp(\bar{U}(x,y)/\sigma)dy} = \frac{\alpha\pi(x,y)}{f(x)}$$

and

$$\frac{\exp(\bar{V}(x,y)/\tau)}{1 + \int_0^{+\infty} \exp(\bar{V}(x,y)/\tau)dx} = \frac{\beta\pi(x,y)}{g(y)}$$

We obtain the assertion of the corollary by combining these equations above to the feasibility of the matching that is traduced by:

$$\int_0^{+\infty} \pi(x,y)dy = f_1(x) \quad \text{and} \quad \int_0^{+\infty} \pi(x,y)dx = g_1(y)$$

As we have:

$$\alpha f_1(x) + (1-\alpha)f_0(x) = f(x) \text{ and } \beta g_1(y) + (1-\beta)g_0(y) = g(y)$$

we can write:

$$\alpha \int_0^{+\infty} \pi(x,y)dy + (1-\alpha)f_0(x) = f(x)$$

and

$$\beta \int_0^{+\infty} \pi(x,y)dx + (1-\beta)g_0(y) = g(y)$$

■

These results can be interpreted as follows: the probability density at which a man of attributes x gets married with a woman of attributes (y) is $\exp\left(\frac{\bar{U}(x,y)}{\sigma}\right)$ times greater than the probability density of remaining single conditionally to his attributes. And same for women, the probability density at which a woman of attributes z gets matched with a man of attributes x is $\exp\left(\frac{\bar{V}(x,y)}{\tau}\right)$ greater than the probability density of remaining single conditionally to her attributes. Note that single individuals are taken as the reference here. But the utility functions $U^0(x)$ and $V^0(y)$ cannot be set to 0 for the reason that they incorporate the consumption-based utility function.

5.3.2 Specification of Consumption Preferences with Generalized Quasi-Linear Functions

The prices of the private good and of the public good are set to one. To guarantee the transferability of utility, the consumption-based utility functions are supposed to be of the Generalized Quasi-Linear form. The identification is parametric. The consumption-based utility of man is:

$$\tilde{U}(a,G) = H(G) + a.h(G)$$

where $h(.)$ is an increasing function and $H(.)$ is an increasing, concave function. More specifically,

$$H(G) = \xi(G-\rho)^{\delta} \text{ and } h(G) = (G-\rho)^{\delta}$$

where $\rho < 0$ and $\delta \in]0,1[$ are general parameters common to all men, and $\xi > 0$ is a general parameter that represents unobserved men' taste for public consumption. Formally, the utility from consumption of man is:

$$\tilde{U}(a,G) = \xi(G-\rho)^{\delta} + a.(G-\rho)^{\delta}$$

The function H can be specific for each man. But for simplicity, we assume that its parameters are common for all the men. The consumption-based utility of woman is:

$$\tilde{V}(b,G) = K(G) + b.h(G)$$

where $K(.)$ is an increasing, concave function. More specifically,

$$K(G) = \zeta(G-\rho)^{\delta}$$

where $\zeta > 0$ is a general parameter that represents unobserved women' taste for public consumption. The parameters ρ and δ are assumed to be the same for men and women. Formally, we have:

$$\tilde{V}(b,G) = \zeta(G-\rho)^{\delta} + b.(G-\rho)^{\delta}$$

The function K can be specific for each man. But for simplicity, we assume that its parameters are common for all the men. It is possible to let the parameters ξ, ζ to be respectively individual for each man and each woman. This can treated in view to generalize this framework.

Singles Let's consider a single man i with characteristics $x = x$. If he is not married, he maximizes his consumption-based utility under the budget constraint:

$$a + G \leq x$$

So his direct utility is:

$$\max_{a^0,G^0} \tilde{U}(a^0,G^0) + \sigma\varepsilon_i^0 \text{ such that } a^0 + G^0 \leq x.$$

This is equivalent to maximize the consumption-based utility under the budget constraint.

$$\max_{G^0} H(G^0) + (x - G^0)h(G^0)$$

with the solution:

$$G^0(x) = \frac{\delta}{1+\delta}x + \frac{\delta\xi + \rho}{1+\delta}, \text{ and } a^0(x) = \frac{1}{1+\delta}x - \frac{\delta\xi + \rho}{1+\delta}$$

Incorporating these expressions into the direct utility functions gives, after simplifications, the indirect utility function of the man i:

$$u_i^0(x) = \frac{1}{\delta}\left(\frac{\delta}{1+\delta}\right)^{\delta+1}(x + \xi - \rho)^{\delta+1} + \sigma\varepsilon_i^0$$

The indirect consumption based utility $U^0(x)$ is the deterministic part of this utility:

$$U^0(x) = \frac{1}{\delta}\left(\frac{\delta}{1+\delta}\right)^{\delta+1}(x + \xi - \rho)^{\delta+1}$$

The results are similar for women. A single woman j has as consumptions:

$$G^0(y) = \frac{\delta}{1+\delta}y + \frac{\delta\zeta + \rho}{1+\delta} \text{ and } b^0(y) = \frac{1}{1+\delta}y - \frac{\delta\zeta + \rho}{1+\delta}$$

Her indirect utility is then:

$$v_j^0(y) = \frac{1}{\delta}\left(\frac{\delta}{1+\delta}\right)^{\delta+1}(y + \zeta - \rho)^{\delta+1} + \tau\eta_j^0$$

The indirect consumption based utility $V^0(y)$ is the deterministic part of this utility:

$$V^0(y) = \frac{1}{\delta}\left(\frac{\delta}{1+\delta}\right)^{\delta+1}(y + \zeta - \rho)^{\delta+1}$$

Let us now focus on utility of married people.

Couples A man i with income x gets matched with a woman j whose income is y and they consume respectively in private a and b and have a public consumption G. As we assume the transferability of utility, they maximize the joint utility :

$$\tilde{U}(a,G) + \tilde{V}(b,G) + \sigma\varepsilon_i(y) + \tau\eta_j(x)$$

under the constraint:

$$a + b + G \leq x + y$$

As the stochastic term $\sigma\varepsilon_i(y) + \tau\eta_j x$ does not depend on consumptions, the choice of consumption maximizes under the constraint budget the term:

$$\tilde{U}(a, G) + \tilde{V}(b, G)$$

i.e

$$\max_{a,b,G}(\xi + \zeta + a + b)(G - \rho)^{\delta} \text{ under the constraint } a + b + G \leq x + y$$

We obtain the following solution. The public consumption is given by:

$$G(x, y) = \frac{\delta}{1 + \delta}(x + y) + \frac{\rho + \delta(\xi + \zeta)}{1 + \delta}$$

and the sum of the private consumptions is:

$$a(x, y) + b(x, y) = x + y - G(x, y) = \frac{1}{1 + \delta}(x + y) - \frac{\rho + \delta(\xi + \zeta)}{1 + \delta}$$

We then deduce the indirect joint utility of the couple Φ_{ij}.

$$\Phi_{ij}(x, y) = \frac{1}{\delta}\left(\frac{\delta}{1 + \delta}\right)^{\delta+1}(x + y + \xi + \zeta - \rho)^{\delta+1} + \sigma\varepsilon_i(y) + \tau\eta_j(x)$$

This is the gross utility from sympathy and consumption the couple has to share between the two partners. The indirect joint surplus is:

$$S(x, y) = \frac{1}{\delta}\left(\frac{\delta}{1 + \delta}\right)^{\delta+1}(x + y + \xi + \zeta - \rho)^{\delta+1}$$

The net joint surplus is then:

$$\bar{S}(x, y) = \frac{1}{\delta}\left(\frac{\delta}{1 + \delta}\right)^{\delta+1}\left[(x + y + \xi + \zeta - \rho)^{\delta+1} - (x + \xi - \rho)^{\delta+1} - (y + \zeta - \rho)^{\delta+1}\right]$$

This is the joint net systematic gain from marriage of a couple of characteristic (x, y). The systematic surplus of a man of attributes x from a union with a woman of attributes y, that is the indirect utility he gets from consumption in this union is:

$$U(x, y) = \frac{\sigma}{\sigma + \tau}\left(\bar{S}(x, y) - \tau\ln\left(\frac{1 - \alpha}{\alpha}f_0(x)\right) + \tau\ln\left(\frac{1 - \beta}{\beta}g_0(y)\right)\right)$$

$$+\frac{1}{\delta}\left(\frac{\delta}{1+\delta}\right)^{\delta+1}(x+\xi-\rho)^{\delta+1}$$

and the systematic surplus of a woman of attributes y with a man of attributes x is:

$$V(x,y)=\frac{\tau}{\sigma+\tau}\left(\bar{S}(x,y)+\sigma\ \ln\left(\frac{1-\alpha}{\alpha}f_0(x)\right)-\sigma\ \ln\left(\frac{1-\beta}{\beta}g_0(y)\right)\right)$$

$$+\frac{1}{\delta}\left(\frac{\delta}{1+\delta}\right)^{\delta+1}(y+\zeta-\rho)^{\delta+1}$$

These are the individual consumption based utilities of each of the two partners. They are the individual net systematic gains from marriage of the partners. Note that, up to this step, the model has provided only the public consumption of the couple and the sum of the two private consumptions. We can now derive the private consumptions.

Private Consumptions within households As individuals do not change their preference, the individual consumption based utilities are exactly the value of their consumption preference function with respect to their consumptions within the households. Considering a man with characteristic x matched to a woman with characteristic y, with a public consumption G, and respective private consumptions a and b, the respective consumption based utilities of each of these two partners at the equilibrium are $\tilde{U}(a,G)$ and $\tilde{V}(b,G)$. The public good consumption is given by:

$$G(x,y)=\frac{\delta(x+y)}{1+\delta}+\frac{\rho+\delta(\xi+\zeta)}{1+\delta}$$

The private consumptions are given by:

$$a(x,y)=\left(\frac{1+\delta}{\delta}\right)^{\delta}\frac{U(x,y)}{(x+y+\xi+\zeta-\rho)^{\delta}}-\xi$$

and

$$b(x,y)=\left(\frac{1+\delta}{\delta}\right)^{\delta}\frac{V(x,y)}{(x+y+\xi+\zeta-\rho)^{\delta}}-\zeta$$

We can remark from these expressions that the if the partners consume the same level of private good, then the happier within the union is the one who has the greatest taste for public good.

Discussion on the main motivation of marriage: Consumption or Love ? In this development, we will show that this framework allows us to identity somehow the main nature of the union. We know that there are two reasons for getting matched on this market that are the public consumption and the sympathy shocks that we will assimilate simply to love. In the case the interest of an individual for marriage is more based on consumption than on love, he will be seen as materialist and in the other case, that is when his motivation for marriage is highly more based on love, he will be seen as romantic. In other words, we will compare for each individual in couple, his/her utility from consumption to his/her sympathy term. To be able to compare with a norm the materialism of the individuals, we propose what follows. Let us consider a man with income x matched to a woman with income y and whose utility is u_i from this and whose utility if he remained single is u_i^0. We also consider a woman with income y matched to a man with income x and whose utility from union is v_j and whose utility if she remained single is v_j^0. The indexes of materialism of these individuals are respectively:

$$\frac{U(x,y)-U^0(x)}{u_i-u_i^0} \quad \text{and} \quad \frac{V(x,y)-V^0(y)}{v_j-v_j^0}$$

Note that the differences $u-u^0$ and $v-v^0$ are non negative. These ratios are in $]-\infty,+\infty[$ and the higher they are, the more materialist the partner is, and the less they are, the more in love the partner is. We do not observe consumptions but we can retrieve them. As we have:

$$U(x,y)-U^0(x)=\tilde{U}(a,G)-\tilde{U}(a^0,G^0) \quad \text{and} \quad V(x,y)-V^0(y)=\tilde{V}(b,G)-\tilde{U}(b^0,G^0)$$

and the functions $\tilde{U}$ and of $\tilde{V}$ are increasing in consumption, we can deduce that if the difference $U-U^0$ is positive then, the consumption vector (a,G) is more preferable than the consumption vector (a^0,G^0). So if the index is positive, we can say that the individual has gained in term of consumption and if the index is negative he/she has lost in term of consumption. But note that a loss in consumption implies automatically a gain in term of sympathy as the net utility is necessarily positive and the individual is then romantic. If the difference $U-U^0$ is low and the difference $u-u^0$ is high then the ratio is low and that shows that the main part of the utility from marriage of this individual is based on the sympathy term. To understand this approach of quantifying the main motivation of marriage, let us consider for instance a first couple in which we have: $U > U^0$ and $V < V^0$. In this union, the utility from consumption of the man is greater than the utility from consumption he would have if he were single. As the utility from consumption is increasing in function of the public good and in function of the sum of the private consumption, this implies that the public consumption of the man in this union is greater than the public consumption he would have if he remained single. This union is a real good deal for him in term of consumption. For the woman, things are quite different. Her consumption based utility is less than the consumption based utility she can get if she were single. As her public consumption has increased, we can deduce that her private consumption has decreased by getting married. So, she has a private consumption less than her partner. The choice of the man is highly more based on consumption than on sympathy for his partner, and the choice of the woman is contrarily more based on sympathy because the consumption based utility she can have for being single is higher than her current consumption based utility, so her total utility in this union derived much from her love to her partner. The index of materialism is negative for the woman and positive for the man. So we can conclude that the man is more materialist than his wife.

In the following subsection, we will focus on the computation and the estimation method.

5.4 Empirical Application

5.4.1 Computation and Parametric Inference

Computation The individual surpluses $U(x, y)$ and $V(x, y)$ depend on the parameters of the consumption preference $\theta = (\delta, \rho, \xi, \zeta)$. We denote $\Theta \subset]0, 1[\times] - \infty, 0] \times [0, +\infty[\times [0, +\infty[$ the set of θ. We will show a way to compute the functions $U^\theta(x, y)$ and $V^\theta(x, y)$ and we propose a parametric inference of the model. For identification, we consider the case $\sigma = \tau = 1$. In the light of Chapter II, we propose the following approach to compute the functions $U^\theta(x, y)$ and $V^\theta(x, y)$. We recall the following operators defined in Chapter II:

$$\begin{aligned}\mathbf{A}(\bar{U}^\theta)(x,y) &= \ln\left(1-\frac{\nu}{1-\nu}\int_0^{+\infty}\frac{f(x)}{g(y)}\frac{\exp\left(\bar{U}^\theta(x,y)\right)}{1+\int_0^{+\infty}\exp\left(\bar{U}^\theta(x,y)\right)dy}dx\right)\\ &\quad+\ln\left(1+\int_0^{+\infty}\exp\left(\bar{U}^\theta(x,y)\right)dy\right)\end{aligned}$$

and

$$\begin{aligned}\mathbf{B}(\bar{V}^\theta)(x,y) &= \ln\left(1-\frac{1-\nu}{\nu}\int_0^{+\infty}\frac{g(y)}{f(x)}\frac{\exp\left(\bar{V}^\theta(x,y)\right)}{1+\int_0^{+\infty}\exp\left(\bar{V}^\theta(x,y)\right)dx}dy\right)\\ &\quad+\ln\left(1+\int_0^{+\infty}\exp\left(\bar{V}^\theta(x,y)\right)dx\right)\end{aligned}$$

We also recall:

$$\bar{U}_0^\theta(x,y) = \frac{1}{2}\bar{S}^\theta(x,y)+\frac{1}{2}\ln\left(\frac{(1-\nu)g(y)}{\nu f(x)}\right)$$

and

$$\bar{V}_0^\theta(x,y) = \frac{1}{2}\bar{S}^\theta(x,y)+\frac{1}{2}\ln\left(\frac{\nu f(x)}{(1-\nu)g(y)}\right)$$

As showed in Chapter II, the sequences $(\bar{U}_k)_{k\in\mathbb{N}}$ and $(\bar{V}_k)_{k\in\mathbb{N}}$ defined as follows converges to $\bar{U}^\theta$ and $\bar{V}^\theta$:

$$\bar{U}_{k+1}^\theta(x,y) = \frac{1}{2}\bar{S}^\theta(x,y)+\frac{1}{2}\ln\left(\frac{(1-\nu)g(y)}{\nu f(x)}\right)+\frac{1}{2}\mathbf{A}(\bar{U}_k^\theta)(x,y)$$

and

$$\bar{V}_{k+1}^\theta(x,y) = \frac{1}{2}\bar{S}^\theta(x,y)+\frac{1}{2}\ln\left(\frac{\nu f(x)}{(1-\nu)g(y)}\right)+\frac{1}{2}\mathbf{B}(\bar{V}_k^\theta)(x,y)$$

The convergence of this sequence as ensured by the fact that the exact functions $\bar{U}$ and $\bar{V}$ verify the equations:

$$\forall\, x,\, \forall\, y,\ \bar{U}^\theta(x,y) = \frac{1}{2}\bar{S}^\theta(x,y)+\frac{1}{2}\ln\left(\frac{(1-\nu)g(y)}{\nu f(x)}\right)+\frac{1}{2}\mathbf{A}(\bar{U}^\theta)(x,y)$$

and

$$\forall\, x,\, \forall\, y,\ \bar{V}^\theta(x,y) = \frac{1}{2}\bar{S}^\theta(x,y)+\frac{1}{2}\ln\left(\frac{(1-\nu)g(y)}{\nu f(x)}\right)+\frac{1}{2}\mathbf{B}(\bar{V}^\theta)(x,y)$$

The uniqueness of $\bar{U}^\theta$ and of $\bar{V}^\theta$ guaranteed by their identification ensures the convergence of the sequences respectively to $\bar{U}^\theta$ and to $\bar{V}^\theta$.

We deduce a second order of approximation of $\bar{U}^\theta$ and $\bar{V}^\theta$ as:

$$\bar{U}_1^\theta(x,y) = \frac{1}{2}\bar{S}^\theta(x,y) + \frac{1}{2}\ln\left(\frac{(1-\nu)g(y)}{\nu f(x)}\right) + \frac{1}{2}\mathbf{A}(\bar{U}_0^\theta)(x,y)$$

and

$$\bar{V}_1^\theta(x,y) = \frac{1}{2}\bar{S}^\theta(x,y) + \frac{1}{2}\ln\left(\frac{\nu f(x)}{(1-\nu)g(y)}\right) + \frac{1}{2}\mathbf{B}(\bar{V}_0^\theta)(x,y)$$

Parametric Inference by Maximum Likelihood We denote respectively by $\mathcal{C}$, $\mathcal{S}_m$ and $\mathcal{S}_f$ the set of couples, the set of single men and the set of single women. The probability of a union (m, w) where m is a man with income $x_m = x$ and w a woman with income $y_w = y$ is the product of the conditional probability of the attraction of the man m by the woman w with respect to $\{x_m = x\}$ with the conditional probability of the attraction of the woman w by the man m with respect to $\{y_w = y\}$. The probability of the attraction of a man with income x by a woman with income y is:

$$\frac{\alpha\pi^\theta(x,y)}{f(x)} = \frac{e^{\bar{U}^\theta(x,y)}}{1 + \int_0^{+\infty} e^{\bar{U}^\theta(x,y)}dy}$$

The probability of the attraction of the woman with income y by a man with income x is:

$$\frac{\beta\pi^\theta(x,y)}{g(y)} = \frac{e^{\bar{V}^\theta(x,y)}}{1 + \int_0^{+\infty} e^{\bar{V}^\theta(x,y)}dx}$$

The probability a man with income x remains single is:

$$\frac{(1-\alpha)f_0^\theta(x)}{f(x)} = \frac{1}{1 + \int_0^{+\infty} e^{\bar{U}^\theta(x,y)}dy}$$

And the probability a woman with income x remains single is:

$$\frac{(1-\beta)g_0^\theta(y)}{g(y)} = \frac{1}{1 + \int_0^{+\infty} e^{\bar{V}^\theta(x,y)}dx}$$

The likelihood is given :

$$L(\theta)=\prod_{(i,j)\in\mathcal{C}}\left(\frac{\alpha\beta(\pi^{\theta}(x_i,y_j))^2}{f(x_i)g(y_j)}\right)\prod_{i\in\mathcal{S}_m}\left(\frac{(1-\alpha)f_0^{\theta}(x_i)}{f(x_i)}\right)\prod_{j\in\mathcal{S}_f}\left(\frac{(1-\beta)g_0^{\theta}(y_j)}{g(y_j)}\right)$$

We then write this likelihood as:

$$L(\theta)=\prod_{(i,j)\in\mathcal{C}}\left[\frac{e^{\bar{S}^{\theta}(x_i,y_j)}}{\left(1+\int_0^{+\infty}e^{\bar{U}^{\theta}(x_i,y)}dy\right)\left(1+\int_0^{+\infty}e^{\bar{V}^{\theta}(x,y_j)}dx\right)}\right]$$

$$\times\prod_{i\in\mathcal{S}_m}\left[\frac{1}{1+\int_0^{+\infty}e^{\bar{U}^{\theta}(x_i,y)}dy}\right]\times\prod_{j\in\mathcal{S}_f}\left[\frac{1}{1+\int_0^{+\infty}e^{\bar{V}^{\theta}(x,y_j)}dx}\right]$$

We maximize the log-likelihood with respect to $\theta=(\delta,\rho,\xi,\zeta)$.

5.4.2 Empirical Results

We use the data of PSID on the period 1968-2001. For description of the data, refer to Chapter II. The variables X and Y represent respectively here the hourly income of men and the hourly income of women. We must recall that we do not observe consumptions; although, we estimate the parameters of the consumption preference specified in the model, and we derive then the consumptions of the individuals.

We estimate the gains from marriage and the individual consumptions within households and their evolution over the period of study. We also estimate the inequality indexes of consumption and of surplus. The comparison between married and singles as we have done in Chapter II, brings here another approach based on the consumption to reveal the impact of marriage on inequality.

Estimation Results The figure 33 describes the evolution of the estimated parameters δ. We remark that the parameter δ has increased on the period 1968-1970, and decreased on the period 1970-1983. It increased on the period 1983-1991 and then slightly decreased on the period 1991-1999.

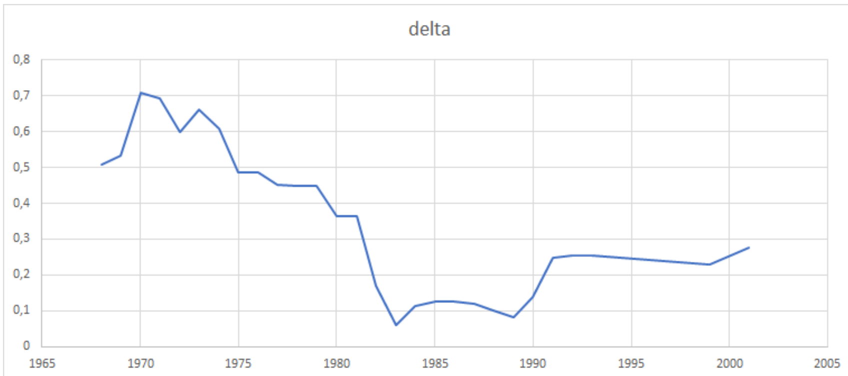


Figure 33: Evolution of δ

The figure 34 shows the evolution of the parameter ρ, of the taste for public good ξ of men and the evolution of the taste for public good ζ of women over the period of the study. We can remark quasi stability on the period $1968-1983$ around the value -2. It decreased in 1984 to around -2.5 stable on the rest of the period 1984-2001. The taste for public of men ξ, is lower then the taste for public good ζ of women. The parameter ξ has been stable relatively stable globally on the period of the study but seems to have decreased from 1999 to 2001. The parameter ζ has been stable on the period 1968-1983, and has then increased in 1984 to a stable value on the period 1984-1999 and increased then from 1999 to 2001.

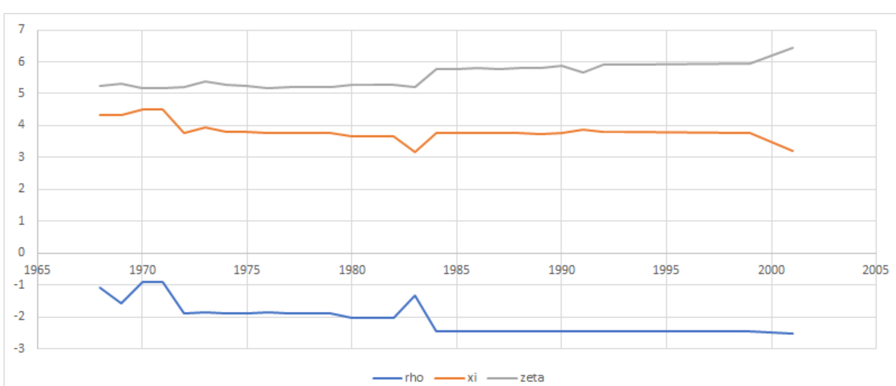


Figure 34: Evolution of ρ, ξ and ζ

The estimation shows that the taste for public good ζ of women is higher than the taste for public good ξ of men. This significant difference in the parameters ξ and ζ is actually linked to the income inequality between men and women. As the taste for public good is higher for women than for men, then considering a single man and single woman with same income, the consumption based utility of the single woman is higher than the consumption based utility of the single man. Actually, we have:

$$U^0(x) = \frac{1}{\delta}\left(\frac{\delta}{1+\delta}\right)^{\delta+1}(x+\xi-\rho)^{\delta+1}$$

and

$$V^0(y) = \frac{1}{\delta}\left(\frac{\delta}{1+\delta}\right)^{\delta+1}(y+\zeta-\rho)^{\delta+1}$$

Within the households, the interpretation does not imply this systematic conclusion. Considering a couple (m, w) where the two partners have the same income, in the case they have the same level of private consumption, then the man has a higher individual consumption based utility. And then for an equal sharing of the joint surplus, the man has to reduce his private consumption. The reduction of the private consumption of the man implies the increase of the public consumption and then the women, with her high taste for public good, she can increase her consumption based utility. We can formalize that as follows:

$$U(x,y) \searrow \Longleftarrow a(x,y) \searrow \Longrightarrow G(x,y) \nearrow \Longrightarrow b(x,y) \nearrow \Longrightarrow V(x,y) \nearrow \Longrightarrow U(x,y) - V(x,y) \searrow$$

We can evaluate the transfer of the man his wife by $-\Delta a(x,y) = \Delta b(x,y)$ in term of consumption or by $\Delta V(x,y)$ in term of utility. We recall that the private consumptions are:

$$a(x,y) = \left(\frac{1+\delta}{\delta}\right)^{\delta}\frac{U(x,y)}{(x+y+\xi+\zeta-\rho)^{\delta}} - \xi$$

and

$$b(x,y) = \left(\frac{1+\delta}{\delta}\right)^{\delta}\frac{V(x,y)}{(x+y+\xi+\zeta-\rho)^{\delta}} - \zeta$$

In other term, men detain the key to reduce inequality within households. They just need to reduce their private consumption and invest more in public expenses to increase mechanically the consumption based utility of their wife and this will lead to a reduction of the surplus gap within the household. In the following paragraph, we will actually confirm the inequality in consumption and utility within households.

Individual Consumptions The figure 35 represents the evolution of the nominal private consumptions. It shows an increase of the private consumptions of married men and of married women. The dynamics of the curves are very similar to the dynamics of the nominal incomes. We remark that the private consumption of married men is greater than the private consumption of married women. The figure 36 represents the evolution of the deflated private consumptions. We notice that the real private consumptions have increased on the period 1968-1989 and they have decreased then on the period 1989-2001. The part in real private consumption of married men is higher to the part of married women. The evolution of the real private consumptions shows that the gap between married men and married women have been reduced.

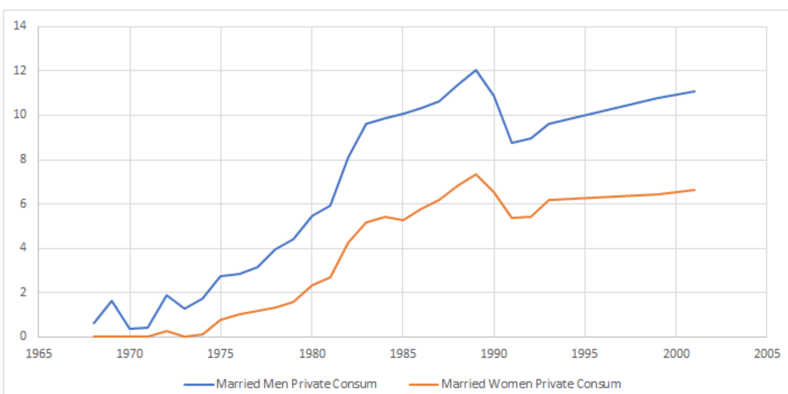


Figure 35: Private Consumption of Married

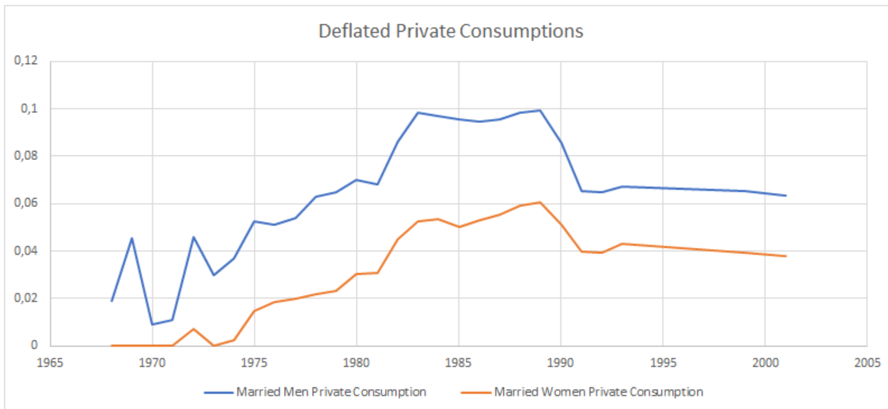


Figure 36: Private Consumption of Married

The figure 37 compares the evolution of nominal consumptions of married individuals and of single individuals. We notice that married men have the highest consumptions and single women have the lowest ones. This was expected since the increasing dynamics of the consumptions are very similar to the dynamics of the incomes. A very interesting remark is despite married women have an hourly income less than the single men, we can see that married women have a greater consumption than single men on the period 1968-1981 and from 1990 till 2001. This shows that consumption is a justification of marriage. We deflate the total consumptions in the figure 38 and we remark a decrease of the real consumption of married men; the real consumption of married women has decreased on the period 1970-1983 and has slightly increased then on the period 1983-2001; the real consumptions of single individuals seem to be stable over the period of the study.

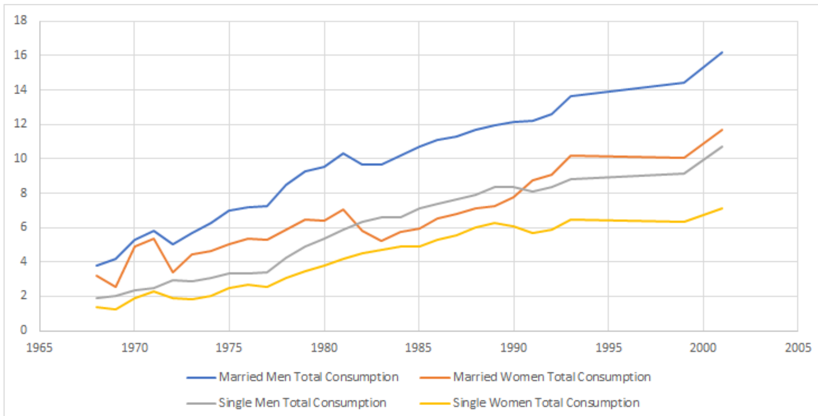


Figure 37: Total Individual Consumption

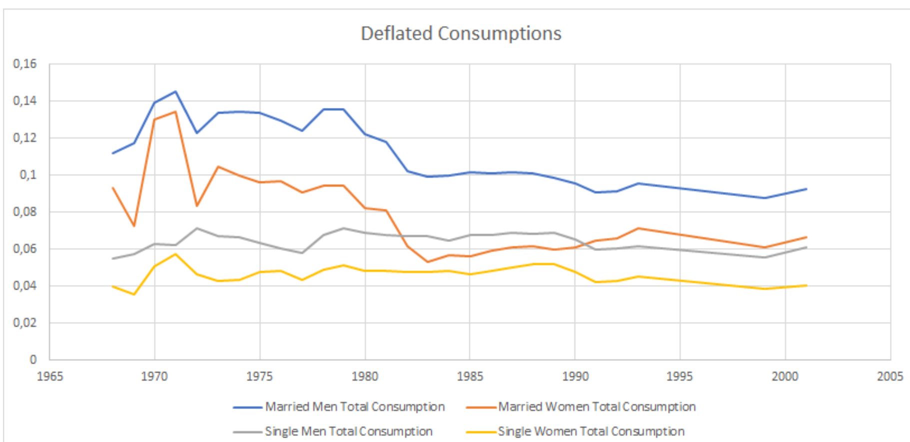


Figure 38: Total Individual Consumption

The figure 39 shows that the total consumption of married men represents about 60% of the total consumption of the household. This proportion is very close to the weight of men in the total household resources as we found in Chapter II.

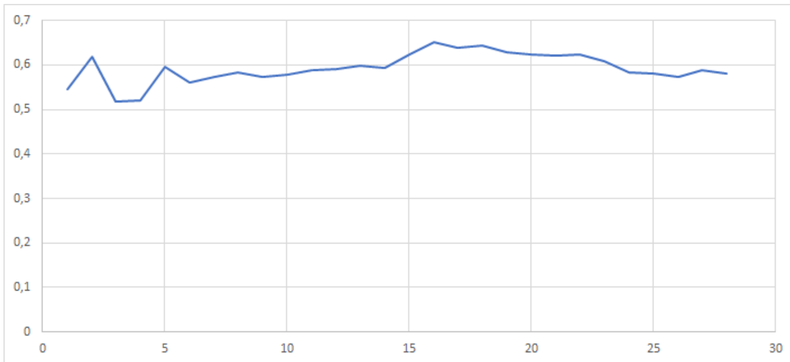


Figure 39: Consumption Part of Married Men within Households

Social Surplus and Individual Surpluses The figure 40 shows the evolution of the individual surpluses on the period of the study. It shows that the individual surplus of men from marriage and consumption is higher to the individual surplus of women. These surpluses have mainly increased on the period 1983-2001.

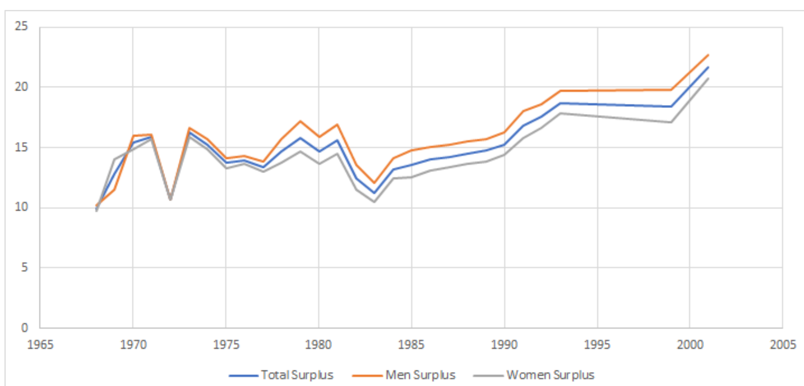


Figure 40: Evolution of the social surplus and averaged individual surpluses

The figure 41 brings more detail about the individual surplus. Married men seem to be the ones who the greatest surplus from marriage and consumption. They are followed by married women. Single men and single women have very close surpluses and they are the lowest on the market.

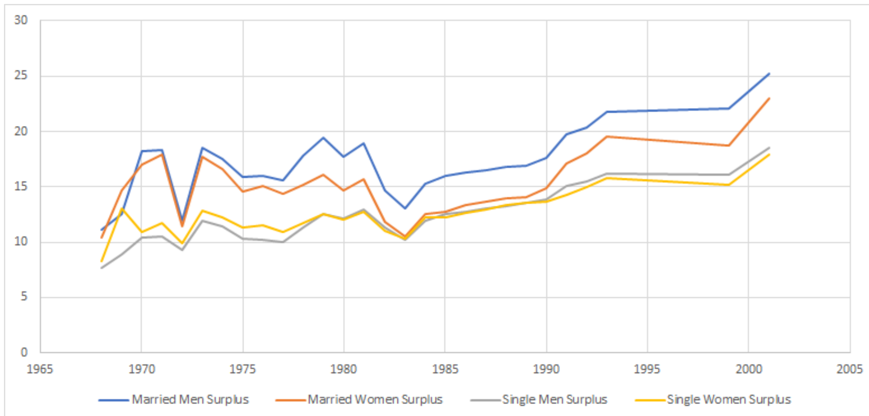


Figure 41: Married Individuals' surplus and Singles' surplus

We can understand clearer the mechanism of transfer within households. The fact that men have a greater part of surplus is explained by their weight of their income in the total resources. This gives them a greater private consumption within the household. To ensure a more egalitarian sharing of the joint surplus, they can transfer a part of their utility to the household through a more important participation to the public good. And then, as wives have a greater taste for public good i.e $\zeta > \xi$, they can benefit from a higher public consumption and this will lead to an increase of their individual utility.

Conversion rate between marriage surplus and income: The value of marriage In Chapter II, we have defined a parameter λ that was the conversion rate between income and the net marriage utility. Its inverse $1/\lambda$ was interpreted as the value of marriage. We have just noticed that the trend of that value of marriage estimated in the chapter was similar to the trend of the parameter δ. Here, we define with a new and simple approach the conversion rate between marriage surplus and income by:

$$\frac{\mathcal{W}(\pi,\kappa)}{\nu\mathbb{E}(X)+(1-\nu)\mathbb{E}(Y)}$$

where X and Y represent respectively the hourly incomes of men and of women. The estimation of this ratio gives the following graph.

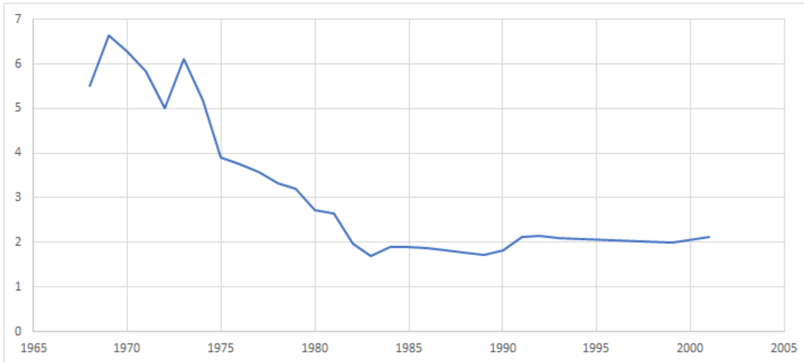


Figure 42: Evolution of the value of marriage

We can remark that, likely to the trend of $1/\lambda$, this ratio has mainly decreased on the period 1970-1983, and then been stable on the period 1983-2001. This result is very analogous to the what we found with $1/\lambda$. We can interpret it as the crash of the value of marriage with respect to the value of dollar. To be precise, considering an individual A living in the 1970s and an individual B living in the 1990s such that they have both the same income, the utility from marriage of B will be expected to be lower than the utility from marriage of man A. The real value of the social welfare of the marriage market has decreased over the time on the period 1970-1983 and has been stable then.

Consumption Inequality and Surplus Inequality We finally analyze the consumption inequality ad the surplus inequality. This analysis is important as a complement to the study in the Chapter II to measure the impact of marriage on inequalities.

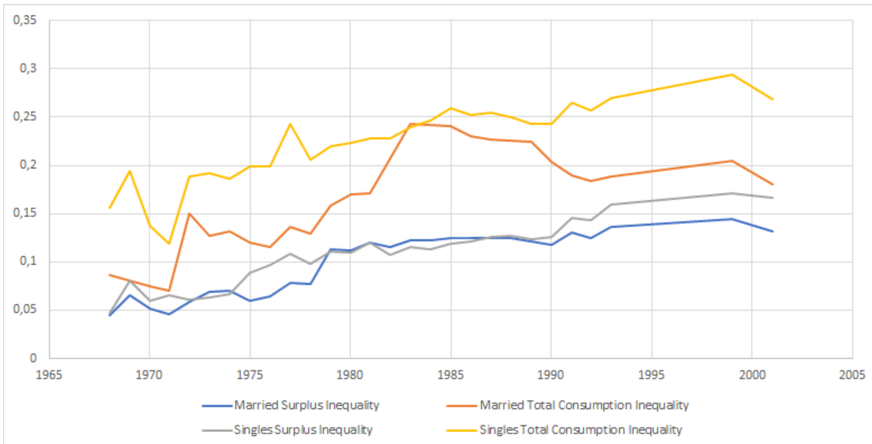


Figure 43: Evolution of the Gini Index of individual consumption and of the individual surplus

The figure 43 shows that the surplus inequality is lower than the consumption inequality. We observe a global increase of the inequality of consumption for singles and a global increase of surplus inequality in the population. The consumption inequality of married individuals has increased on the period 1968-1983 and decreased then on the period 1983-2001. The surplus inequality of married individuals is slightly lower than the surplus inequality of single individuals. But the difference is clearer in term of consumption inequality. In fact, we remark that the consumption inequality is higher for singles than for married. This confirms the result we found in the chapter 2 i.e marriage reduces inequalities at least in term of consumption.

5.5 Conclusion

In this paper, we have studied the decision of marriage and of consumption. The approach of investigation is based on the model proposed in Chapter II which is a continuous extension of the discrete model of Choo and Siow (2006) and an alternative approach of the continuous model of Dupuy and Galichon (2014). By assuming the transferability of utility, and by considering a stability of the marriage market in the sense of Gale and Shapley (1962), we have identified the consumption based utility of individuals within their household at the equilibrium of the market. This utility from consumption is in fact the systematic part of the utility. The second part of utility is stochastic and quantifies the sympathy shock between the partners or the auto-sympathy term when the individual is unmatched. From the identification of the individual utility based on consumption, we show that the intra-household consumption is the solution of an system of equations with three equations and three unknown that are the two private consumptions of the partners and their common public consumption. Especially, by considering a particular specification of Dupuy and Galichon to model the stochastic part of the utility, and a Generalized Quazi-linear form for consumption preferences, we derived the private consumptions and the public consumption within the household, depending on their incomes and the tastes for public good, in function of the their individual systematic surpluses and these individual systematic surpluses can be computed by different methods: in fact for instance, Galichon and Salanié (2015) proposed for that an algorithm and in Chapter II, we suggested a way of recursive analytical approximation from a precise starting point that corresponds to the analytical solutions of the individual surpluses when the matching rate tends to zero. The estimation approach is based on the maximum likelihood. We estimate the model on data from PSID on the period 1968-2001. The estimation of the model shows a difference in taste for public good for women and men. Actually, it revealed that women have a higher taste for public good than man. One implication of that, it is that for the same level of income, single women have a higher consumption based surplus than men. And within households, for a same level of private consumption, men have a higher individual surplus than their female partner and in that case for an equal sharing of the joint surplus men have to diminish their private consumption and invest more in public good or in direct transfer towards their wife to increase her individual surplus. This leads us to suggest that the happiness of men on this market comes more from sympathy than from consumption. In a sense we can then say that men marry for love i.e the stochastic part of the utility.. This makes sense as they have the highest incomes in the population. The marriage decision of women seems to be more explained by the deterministic part of the utility. We also observed a quasi stability of the taste for public good for men on the period of the study. The taste for public of women has been stable on the period 1968-1983 and has increased in 1984 to a higher value stable on the period 1984-1999 and increased from 1999 to 2001. The private consumption of married men is greater than the private consumption of married women. We noticed that the real private consumptions have increased on the period 1968-1989 and they have decreased then on the period 1989-2001. The part in real private consumption of married men is higher to the part of married women. The evolution of the real private consumptions shows that the gap between married men and married women have been reduced. We also remarked a decrease of the real consumption of married men; the real consumption of married women has decreased on the period 1970-1983 and has slightly increased then on the period 1983-2001; the real consumptions of single individuals seem to be stable over the period of the study. Furthermore, the estimation shows that married have a greater consumption than singles and have then a greater surplus. This is explained by the fact that married have higher incomes than singles. We noticed that consumptions and individual surpluses have dynamics similar to the dynamics of the incomes. The happiest on the market are clearly married men. In fact they benefit from the highest consumptions and highest surpluses. The comparison between married and singles shows that marriage increases consumption and this gives a justification for unions. We also noticed that there exist a weaker inequality of consumption and of surplus among married than among singles. This confirms the observed result in the chapter 2 according to which marriage reduces income inequality. Actually, this income inequality reduction is achieved through a more egalitarian distribution of the consumption by marriage. Moreover we observed an increase of the inequality of consumption and of surplus. Finally, we proposed another approach to compute to conversion rate between marriage surplus and income alternatively to the approach suggested in the chapter 2. This conversion rate can be interpreted as a monetary value of marriage on the market. We observed that this value has increased on the period 1968-1970, has then decreased drastically on the period 1970-1983 and has been stable then between 1983 and 2001. This result is very conform to what we observed Chapter II.

References

[1] Bargain, Olivier & Olivier Donni. 2012, Expenditure on Children: A Rothbarth-Type Method Consistent with Scale Economies and Parents' Bargaining, European Economic Review, vol. 56, pp. 792-813.

[2] Bergstrom. 1989, A Fresh Look at the Rotten Kid Theorem–and Other Household Mysteries, Journal of Political Economy, vol. 97, issue 5, 1138-59

[3] Bergstrom, Theodore C & Cornes, Richard C. 1983, "Independence of Allocative Efficiency from Distribution in the Theory of Public Goods," Econometrica, Econometric Society, vol. 51(6), pages 1753-1765

[4] Blundell, Richard, Pierre-André Chiappori & Costas Meghir. 2005 Collective Labor Supply with Children, Journal of Political Economy, vol. 113, no., pp. 1277-1306.

[5] Bourguignon, Frans & Pierre-André Chiappori. 1992, Collective Models of Household Behavior: An Introduction. European Economic Review, vol. 36, pp. 355-364

[6] Bourguignon, Frans, Martin Browning, & Pierre-André Chiappori. 2009, Efficient Intra Household Allocations and Distribution Factors: Implications and Identification, Review of Economic Studies 76, pp.503-26

[7] Browning M., F. Bourguignon, P.A. Chiappori & V. Lechene. 1994, Incomes and Outcome : A Structural Model of Intrahousehold Allocation, Journal of Political Economy, pp. 1067-97.

[8] Browning M. & P.A. Chiappori. 1998, Efficient Intra-Household Allocations: A General Characterization and Empirical Tests, Econometrica, Vol. 66, No. 6, pp. 1241-1278

[9] Browning, Martin, Pierre-André Chiappori & Arthur Lewbel. 2013, Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining Power, Review of Economic Studies, 80, pp. 12671303

[10] Browning, M., Chiappori, P. & Weiss, Y. 2014, Economics of the Family (Cambridge Surveys of Economic Literature), Cambridge University Press. doi:10.1017/CBO9781139015882

[11] Cherchye, Laurens, Thomas Demuynck, Bram De Rock & Frederic Vermeulen. 2017, Household Consumption When the Marriage Is Stable, American Economic Review, VOL. 107, NO. 6, pp. 1507-34

[12] Cherchye, Laurens, Bram De Rock, Arthur Lewbel & Frederic Vermeulen. 2015, Sharing Rule Identification for General Collective Consumption Models, Econometrica, Volume83, Issue 5, Pages 2001-2041

[13] Cherchye, Laurens, Bram De Rock & Frederic Vermeulen. 2011, The Revealed Preference Approach to Collective Consumption Behavior: Testing and Sharing Rule Recovery, Review of Economic Studies, Vol. 78, No. 1, pp. 176-198, Published By: Oxford University Press

[14] Cherchye, Laurens, Bram De Rock & Frederic Vermeulen. 2012, Married with Children: A Collective Labor Supply Model with Detailed Time Use and Intrahousehold Expenditure Information, American Economic Review, VOL. 102, NO. 7, pp. 3377-3405

[15] Chiappori, Pierre André. 1988, Rational household labor supply, Econometrica, Vol. 56, No. 1, pp. 63-90

[16] Chiappori. 1992, Collective Labor Supply and Welfare, Journal of Political Economy, Vol. 100, pp. 437-67

[17] Chiappori P.A. & I. Ekeland. 2007, The Micro Economics of Group Behavior 1 : General Characterization, Journal of Economic Theory, Vol. 130, pp. 1-26.

[18] Chiappori P.A. & I. Ekeland. 2009, The Micro Economics of Group Behavior 2 : Identification, Econometrica, Vol 77, issue 3, pp. 763-99.

[19] Chiappori P.A., B. Fortin & G. Lacroix. 2002, Marriage Market, Divorce Legislation and Household Labour Supply, Journal of Political Economy, Vol. 110, No. 1, pp. 37-72, Published By: The University of Chicago Press

[20] Chiappori & Weiss. 2017, Partner Choice and the Marital College Premium, American Economic Review, vol. 107, no. 8, pp. 2109-67

[21] Choo, Eugene & Aloysius Siow. 2006, Who marries whom and why, Journal of Political Economy, Vol. 114, No. 1, pp. 175-201

[22] Couprie, H., Peluso, E., & Trannoy, A. 2010, Is power more evenly balanced in poor households? Journal of Public Economics, 94(7), 493-507.

[23] Donni Olivier & Eleonora Matteazzi. 2012, On the importance of domestic production in collective models : Evidence from US data, Annals of Economics and Statistics, 105-106, p. 99-126

[24] Dunbar, Geoffrey R., Arthur Lewbel & Krishna Pendakur. 2013, Childrens Resources in Collective Households: Identification, Estimation, and an Application to Child Poverty in Malawi, American Economic Review, VOL. 103, NO. 1, FEBRUARY 2013, pp. 438-71

[25] Dupuy, Arnaud & Alfred Galichon. 2014, "Personality Traits and the Marriage Market," Journal of Political Economy, University of Chicago Press, vol. 122(6), p 1271-1319

[26] Gale & Shapley. 1962, College Admissions and the Stability of Marriage, Mathematical Association of America, Vol. 69, No. 1, pp. 9-15

[27] Galichon, Alfred & Bernard Salanié. 2012, Cupids invisible hand: social surplus and identification in matching models, Sciences Po publications

[28] Lewbel, Arthur & Krishna Pendakur. 2008, Estimation of Collective Household Models With Engel Curves, Journal of Econometrics, special issue on Estimating Demand Systems and Consumer Preferences, vol. 147, pp. 350-358.

[29] Lise, Jeremy & Shannon Seitz. 2011, Consumption Inequality and Intra-household Allocations, Review of Economic Studies,vol. 78, pp. 328-355.