Marriage gains: who should you marry?*

Marion Gousse†

December 7, 2011

PRELIMINARY AND INCOMPLETE

Abstract

To evaluate the impact of taxation reforms and family policy programs on labor supply, we need to understand the intrahousehold allocation of time and consumption. Collective models initiated by P.A.Chiappori developed an estimation strategy to recover the sharing rule of resources within the household from the sole observation of labor supplies. However, these models analyse the sharing rule within existing marriages, which can give an incomplete picture of the determinants of the well-being of men and women. Indeed, the marriage market, as Becker ([2], 1973) has emphasized, is an important determinant of the distribution between men and women: the knowledge of who marries whom allows a better understanding of the distribution of revenues and labor supplies. Recently, search models of marriage have been exploited (Shimer and Smith ([18], 2000) or Jacquemet and Robin ([14], 2011). Following the approach of J-M.Robin and N.Jacquemet, this article proposes a multidimensional search and matching model of marriage. Whereas existing search marriage models allow people to match only on their wages, this model allows people to choose their partner according to their wages and their education levels. This model tries to account for the homogamy phenomenon that is observed in the data: people are often married with people who have the same education level. This paper also exploits the physical attractiveness as a potential factor of marriage and as a potential bargaining power in intrahousehold allocation.

Keywords: Marriage search model, Collective labor supply, Structural estimation

JEL classification: C78, D83, J12, J22

*I thank Jean-Marc Robin for his supervision. All errors remain mine.
†Sciences Po, Paris, Departement d’Economie, E-mail: marion.gousse@sciences-po.org
1 Introduction

Motivation

Evaluation of family policy programs is usually made with unitary models which consider the household as a unit. The two individuals of the household are supposed to pool their income and maximize a neoclassical household utility function subject to the household’s total income constraint and the time constraint. Labor supplies are then derived using traditional microeconomic theory instruments. However, households are composed of several individuals with different resources and preferences. As these unitary models do not take into account the decision process within the household, they must fail in evaluating the demand function of the household and labor supply functions of its members. Indeed, many empirical studies show that classical results of the standard theory were not confirmed by the data. Therefore, to evaluate the impact of taxation reforms and family policy programs on labor supply, we need to understand the intra-household allocation of time and consumption.

Collective models of the household

In 1998, P.A.Chiappori proposed a first answer to modelise the repartition of resources: the collective models ([8](1988), [9](1992)). Assuming that the household decision process always leads to a Pareto-optimal agreement, collective models are able to estimate the sharing rule of resources within the household. However, these models analyse the sharing rule within existing marriages, which can give an incomplete picture of the determinants of the well-being of men and women. Indeed, the marriage market, as Becker ([2],1973) has emphasized, is an important determinant of the distribution between men and women: the knowledge of who marries whom allows a better understanding of the distribution of revenues and labor supplies.

The work on bargaining models with threat points initiated by Manser and Brown ([13],1989) is a first step in this direction. Bargaining models allow the formation and the dissolution of marriage to depend on the repartition of the resources of the household between the two members. These models allow people to marry if and only if marriage brings a surplus to each member and allows the separation of married people if the match does not brings the surplus anymore. Thus, bargaining models also provide an opportunity for integrating the analysis of distribution within marriage with a matching or search model of the marriage market. Indeed, recent dynamic models of marriage which include search costs have been exploited using bargaining. Some structural models with several periods as in M. Mazzocchi ([17],2007) or in Cauwet, Guner and Knowles ([6],2002) and some search models as in Shimer and Smith ([18],2000), Wong ([20],2003) or Robin and Jacquemet ([14],2011). These search models assume that single people regularly meet another single people following a Poisson process. For each new meeting, singles have to choose between marrying this new individual or waiting and looking for a better option. Then
the objective is to find the pareto-equilibrium, that is the situation when no agent can improve his situation by marrying, quitting, or changing partner without degrading the situation of another agent. Then the characterization of the equilibrium defines a group of singles and a group of married people and let us now who marries whom at the equilibrium.

Who marries whom? Homogamy and physical attractiveness

French statistics (Insee, 2006,[19]) on the repartition of couples among social groups (blue collar, employee, white collars and executives,...) show that there is a strong social homogamy in France, that is both members of a couple belong much more to the same social group than if they were matched randomly.

Consequently, social homogamy and particularly educational homogamy is still a reality and is even increasing in developed countries. Indeed, since 1950, women are getting more and more educated and are marrying men who have the same educational level. The american sociologist Robert D.Mare [16] document a strong 'U-turn' in educational homogamy in the U.S over the 20th Century. Spousal resemblance on educational attainment was very high in the early 20th Century, trended down to an all time low for young couples in the early 1950s, and have been increasing steadily since then. Many explanations can be found. The sociologists have studied that question and found that there is an unconscious attraction to socially similar people. Indeed, Michel Bozon and Francoise Héran in "La formation des couples"(2006) [4] point out two important factors who are mutually reinforcing and which lead naturally to homogamy. First, there is the influence of preferences: individuals prefer partners who share social traits as religious beliefs and practice, race-ethnicity and leisure activities. For other traits as economic success or physical attractiveness, they prefer partner who are more successful or attractive relative to their sex than themselves. However, the competition between individuals on the marriage market make them date with partners who are approximately as successful or attractive as they are. Then competition in the marriage market implies that best educated men marry best educated women. This is the common result of Positive Assortative Matching found by Gary Becker in 1973 ((2]) and more recently by Shimer and Smith (2000) ((18]). This explanation holds if men and women value the level education of their partner which seems to be true for men and for women. Indeed, if educated women used to be more often single that less educated women, it is not the case anymore. Indeed, Chiapppori et al.[10] (2011) found out that educated women have become more and more attractive on the marriage market.

The second reason to explain homogamy is the influence of the environment: individuals face different opportunities for meeting and mating with other individuals. Indeed, people have a differential exposure to potential marriage partners that arises from socioeconomic segregation of schools, neighborhood, places of worship, leisure which tend to foster marriages that are much more homogamous than would be expected on the basis of chance. If meeting place have been diversified with urbanisation and development or leisure, sociability and leisure areas are still
socially segmented which increases chances to meet people of the same social background.

Then homogamy could be explained by homophiliala, by a bigger probability to meet someone from one's social group, but also by competition on the marriage market for the best educated people.

Another important factor of marriage is obviously the physical attraction between two people. However, beauty is difficult to measure and even if beauty indicators exist, I haven’t found an important database collecting this beauty indicator and other important variables as socio-economic characteristics, familial status, etc.

Consequently, some studies use the Body Mass Index (BMI)\(^1\) as a proxy for beauty. Indeed, in their study Chiappori et al. (2009) ([7]) try to evaluate agent’s preferences for some partner’s qualities and to estimate the substitutability between them. He separates qualities of physical attraction and qualities of productivity (socio-economical status) and evaluates agents’ preferences assuming that they are homogenous among men and homogenous among women (but different between sexes). Results of his studies show that men are more interested in physical characteristics than women and that substitutability between productivity and physical characteristics is much more important among women (roughly speaking, women value equally a rich and a not so handsome man and a poor but handsome man whereas men only prefer beautiful women.)

In this paper, I follow the approach of N. Jacquemet and J-M. Robin and I integrate a collective model of the household in a search and matching model of marriage where people match on two characteristics. I first allow people to match on their wages and their education level, then I allow them to match on their wages and a indicator of their physical attractivity. I estimate the model on two different databases : the French database Enquete Emploi 2009 and the American database PSID 2009. This paper is inspired from the estimation strategy of Choo and Siow (11122006) who use the density of couples of different types in the population as a measure of the public good which is produced in the household. More couples of particular type exist more this particular marriage brings a surplus to the couple.

This paper then tries to estimate the surplus of a particular match and how the share of this surplus is done between the household members. The identification of sharing is made using the change in worked hours between singles and married people. Some fundamentals elements are missing as the presence of children an household production. If these elements have already been introduced in collective models : the presence of children in Blundell, Chiappori et al. (3, 2005), the household production in Couprie (12, 2007), the introduction of taxation in Donni (13, 2003).

I present the theoretical framework in section 2. Section 3 presents the data and section

\(^1\) \(BMI = \frac{\text{weight}(kg)}{\text{size}(m)^2}\), the Body Mass Index is chosen to represent physical attractivity. Some other indicators exist as the ratio of height on waist or the ratio of height on chest,...
the estimation strategy. In section 5, I present the results obtained by the bidimensional matching on wages and education level whereas in section 6, I present the results obtained with a bidimensional matching on wages and the Body Mass Index (BMI). Section 7 concludes.

2 Model

This section presents the framework of the model and the main assumptions. I will consider two different populations of agents that are likely to match. A match is considered in this paper as a two people household, that is a group of two people who love each other and who decide to live together and make common future projects. For convenience, I use indifferently the terms match, marriage or couple. I also consider that the two populations who are likely to match are males and females (population $m$ and $f$). Each agent is defined by the population he belongs to but also by his individual type.

In this paper, the agent type is defined by his wage and another characteristic which can be his education level or his physical attractivity. The assumption made here is that these characteristics are fundamental in the marriage matching process. They are the intrinsic characteristics of the individual and they do not evolute in time.

I suppose that agents are egoistic, that is they derive their utility only from their own consumption and leisure. If this assumption is reasonable for a single individual, it is however a strong assumption for an individual in couple. Indeed, we can suppose, that a married man care about the utility of his wife or even care about what she consumes (he can dislike when she smokes or like when she buys new clothes...), and he can also care about her leisure time and enjoys when she has more time to spend with him. However for tractable reasons, I will keep the model of egoistic agent.

I also assume that when two people decide to live together, it is not only because they will make economies of scale but it is also because they like each other and they like spending time together. Then marriage or cohabitation brings them some additional utility which will be called the match quality.

2.1 Indirect utilities

When two individuals are married (or cohabiting), they produce together a public good or make economies of scales which slackens their budget constraint. We will note $Q_{ij}$ this economy of scales.

In this paper, this public good is only a monetary good which is created without any inputs\(^2\). It is just a saving of money made where the members of the household live together. It can come from the sharing of a rent or the sharing of meals or from a tax reduction due to the pooling of

\(^2\)This creation of monetary resources only due to coexistence of two loving people also appeared in the initial model of Manser and Brown (1980) ([15])

5
declared fiscal resources ...

This economy of scales has to be reparted between the two members. Following the literature on collective models, I suppose that the household take Pareto-efficient outcomes. Moreover, I add an additional assumption as in the Jacquemot-Robin model ([14]) that the members of the household make a Nash bargaining to share their resources. This bargaining obtain Pareto-efficient outcomes. Then the program of the household can be decentralised and follows a two steps processus. First, the members share their non-labor income resources plus the additional resource created by the life together. Then each members maximises his own utility subject to his new budget constraint.

I assume that non-labor income resources are null. Then let $t_{mij}$ be the part of the public good obtained by the member $m$ of type $i$ in a match with a member $j$, then we will have $Q_{ij} = t_{mij} + t_{fij}$. The program of the married agent is:

$$\max_{c_{i,h}} \ u_{mij}$$
$$s.e \ c_{mij} \leq w_i(T - t_i) + t_{mij}$$

$p$ is the price of the private consumption good (a hicksian aggregate good) which can be normalized to one.

I will use a demand system linear in expenditure, that is the demand function for a good $c^h_i$ by a man $h$ who has a total income of $R^h$ has the following linear form:

$$c^m_i = a^m_i(p) + b^m_i(p)R^h$$

(1)

with $p$ the price vector of goods.

That is, the consumption of the good $c^m_i$ depends on individual characteristic $a^m_i(p)$ and on a part of the total income which is constant across men $b^m_i(p)$. Then, this part is a function of price which does not depend on the marital status of the individual man. In a same way, I define $a^h_i(p)$ and $b^h_i(p)$ to get the demand for a good $i$ by a woman $h$.

To obtain a demand system of this form compatible with the maximisation of a consumer’s utility, you need to use an indirect utility of the following form:

$$u_{mij} = \frac{K_{mij}^d (w_iT + t_{mij} - A_{mi}(w_i))}{B_m(w_i)}$$

and $u_{mi} = \frac{(w_iT - A_{mi}(w_i))}{B_m(w_i)}$ for a single man.

with $B_m(w_i)$ is an aggregate price index for a man of type $i$ and $A_{mi}(w_i)$ can be interpreted as the minimal consumption needed to reach a positive level of utility.

$K_{mij}$ corresponds to the match quality for the man of type $i$ when he marries a woman of type
\[ j. \text{ It can be different between men and women. } d_m \text{ is the weight of the match quality in the utility function.} \]

The optimal labor supply conditional on the level of public good is (by Roy’s identity or directly with the result of the program of maximisation).

\[ T - t^*_m = h^*_m = T - A'_m(w_i) - \frac{B'_m(w_i)}{B_m(w_i)}(w_i T + t_{mij} - A_m(w_i)) \] (2)

2.2 The search model

This paragraph presents the meeting process between two singles. I assume that only singles search for a partner. There is no “on the marriage search”.

I will use the the following notations : \( U \) is the number of singles, \( N \) is the number of matches. Let \( U_f \) and \( U_m \) be the respective number of single women and single men. Let \( u_f(i) \) be the density of single women of type \( i \). I note \( \lambda_m \) the instantaneous probability of an agent among the population \( m \) of meeting a new person of the population \( f \) and \( M(U_f, U_m) \) the number of meetings per period. Then \( \lambda_m = \frac{M(U_f, U_m)}{U_m} \). We then have the instantaneous probability of a meeting be \( \lambda = \frac{\lambda_m}{U_m} = \frac{M(U_f, U_m)}{U_f U_m} \). Let \( \delta \) be the separation rate and \( r \) the interest rate.

The following Bellman equation shows the present value of a single man of type \( i \):

\[ r W_{mi} = v_{mi} + \lambda(\mathbb{E}_{fj} (\max(W_{mij} - W_{mi}, 0)1(W_{fij} > W_{fj}))) \] (3)

with \( W_{mij} \) the present value of a man of type \( i \) married with a woman of type \( j \), \( W_{mi} \) the present value of single man of type \( i \), \( W_{fij} \), the present value of a woman of type \( j \) married with a man of type \( i \), \( W_{fj} \) the present value of single woman of type \( j \). \( \mathbb{E}_{fj} \) is the esperance of a man evaluated on all possible types of women he can meet.

The present value of a man of type \( i \) married with a woman of type \( j \) is

\[ r W_{mij} = v_{mij} + \delta (W_{mi} - W_{mij}) \] (4)

\[ W_{mij} = \frac{v_{mij} + \delta W_{mi}}{r + \delta} \]

When a match is formed, the two members start a Nash bargaining and choose the repartition of the public good \( Q_{ij} \) that has been created by the fact that they live together. The Nash bargaining power of the man is \( \beta \) and the Nash bargaining power of the woman is \( (1-\beta) \). I assume as in the Jacquemet-Robin modelisation, that the public good is not entirely determined by the characteristics of the household but has also an undetermined compoant. Then \( Q_{ij} = C(i, j) + z \) and \( z \) is a random variable. Then the Nash bargaining of the household is the maximisation of the following program:
\[
\max_{t_m, t_f} (W_{mij} - W_{mi})^\beta (W_{fij} - W_{fj})^{1-\beta} \\
\text{s.c } t_m + t_f \leq Q_{ij} = C(i, j) + z
\]

Then the surplus of a marriage for an \( i \)-type agent with a \( j \)-type agent is

\[
S_{mij} = W_{mij} - W_{mi} = \frac{v_{mij} - rW_{mi}}{r + \delta}
\]

\[
S_{fij} = W_{fij} - W_{fj} = \frac{v_{fij} - rW_{fj}}{r + \delta}
\]

Then the Nash-bargaining is easier to derive and we get the following formula for the sharing of the public good\(^3\):

\[
t_m(i, j, z) + w_i T - A_{mi}(w_i) - \frac{s_m(i) + w_i T - A_{mi}(w_i)}{K^{d_{mij}}_{mi}} = \beta (S(i, j) + z)
\]

\[
t_f(i, j, z) + w_j T - A_{fj}(w_j) - \frac{s_f(j) + w_j T - A_{fj}(w_j)}{K^{d_{fij}}_{fj}} = (1 - \beta) (S(i, j) + z)
\]

with

\[
s_m(i) = B_{mi}(w_i) r W_{mi} - w_i T + A_{mi}(w_i)
\]

\[
s_f(j) = B_{fj}(w_j) r W_{fj} - w_j T + A_{fj}(w_j)
\]

\[
S(i, j) = C(i, j) + w_i T - A_{mi}(w_i) + w_j T - A_{fj}(w_j) - \frac{s_f(j) + w_j T - A_{fj}(w_j)}{K^{d_{fij}}_{fj}} - \frac{s_m(i) + w_i T - A_{mi}(w_i)}{K^{d_{mij}}_{mi}}
\]

\(s_m(i)\) and \(s_f(j)\) can be interpreted as the individual surplus for a single person. It corresponds to a pondered difference between a single present value and his instantaneous utility.

The two individuals \( i \) and \( j \) decide to marry if and only if

\[
S(i, j) + z > 0
\]

\(^3\) We remark that if \( d_m = d_f = 0 \), that is if there is no impact of match quality in the utility function and the match only happens for monetary reasons, then we have

\[
t_m(i, j, z) - s_m(i) = \beta (S(i, j) + z)
\]

\[
t_f(i, j, z) - s_f(j) = (1 - \beta) (S(i, j) + z)
\]

with

\[
s_m(i) = B_{mi}(w_i) r W_{mi} - w_i T + A_{mi}(w_i)
\]

\[
s_f(j) = B_{fj}(w_j) r W_{fj} - w_j T + A_{fj}(w_j)
\]

\[
S(i, j) = C(i, j) - s_f(j) - s_m(i)
\]
The matching probability can then be calculated as

\[ a(i, j) = P\{S(i, j) + z > 0|i, j\} \]

\[ = 1 - P\{z \leq -S(i, j)|i, j\} \]  

(8)

2.3 Equilibrium

The characterisation of the equilibrium allows us to close the model. To solve for a market equilibrium, we have to describe how new singles enter the market over time. Burdett and Coles ([5], 1999) review the different cases that have been considered in the literature. Here I suppose there is no entry of new singles, however partnerships are destroyed at some exogenous rate \( \delta > 0 \) whereupon both return to the single market. At the equilibrium, there is equality between inflows and outflows for each type of marriage. We will note \( n(i, j) \) is the density of couples of type \( i \) for the \( m \) member and of type \( j \) for the \( f \) member. Then we have for all couple of type \((i, j)\), the equality between the number of outflows and the number of inflows:

\[ \delta n(i, j)N = \lambda_u m(i)u_f(j)U_mU_f \mathbb{P}(1(W_{mi} > W_{mj})1(W_{fj} > W_{fj})) \]

(9)

\[ \delta n(i, j)N = \lambda_u m(i)u_f(j)U_mU_f a(i, j) \]

\[ a(i, j) = \frac{\lambda_u m(i)u_f(j)U_mU_f}{\delta n(i, j)N} \]  

(10)

We see here that the data will allow us to compute the match probability \( a(i, j) \). Then using the equation (8) and making some additional assumptions on the law of \( z \), we will be able to derive the match surplus and the present value of individuals. The details will be exposed in section 4.

3 Data

I use two databases to estimate the model. I use the French Database : Enquête Emploi 2009 to estimate the matching on wages and education level whereas I use the American data base PSID 2009 to estimate the matching on wages and BMI.

In both cases, I limit my sample to households composed of one single member (with children or not) and household composed of a couple (with or without children). I eliminate households composed with several adults who are not in couple. I only keep households where all the adults are between the age of 18 and 65, where all people work and declare their wages and the number of hours worked. Then I trim the 1% top and bottom wages and the 1% top and bottom of numbers of hours worked. I don’t use any information of the non labor resources of the households.
Enquete Emploi 2009

I dispose of 30 047 individuals with 10 701 couples, 5 067 single women and 3 578 single men. The ratio $U_m/U_f$ is 0.71 which correspond to a strong deficit of men in comparison to female and the proportion of married people is $2N/(2N + U_m + U_f) = 71\%$. I also use information on their education level. There are 6 levels of diplomas which are described in table 3.

<table>
<thead>
<tr>
<th>Education level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>No diploma, primary school, intermediate school</td>
</tr>
<tr>
<td>3</td>
<td>Professional formation</td>
</tr>
<tr>
<td>4</td>
<td>High school graduate (professional and general)</td>
</tr>
<tr>
<td>5</td>
<td>College Undergraduate</td>
</tr>
<tr>
<td>6</td>
<td>College Graduate</td>
</tr>
<tr>
<td>7</td>
<td>PhD, Grandes Ecoles</td>
</tr>
</tbody>
</table>

The following graphs present the distribution of hourly wages (in €) and education level in the population.

![Graph](https://via.placeholder.com/150)

Figure 1: Wage density by sex and marital status. Enquete Emploi 2009

Men have higher wages than women, particularly married men. Women hourly wages are centered around 8.7 € whereas single men hourly wages are centered around 9 € and married men hourly wages around 10 €
PSID 2009

I dispose of 7,477 individuals with 2,451 couples, 1,651 single women and 924 single men. The ratio $U_m/U_f$ is 0.56 which also correspond to a strong deficit of men in comparison to female and the proportion of married people is $2N/(2N + U_m + U_f) = 66\%$. I also use information on their body mass index.

The following graphs presents the distribution of hourly wages (in $\) and Body Mass Index in the population.

Figure 2: Histogram of education level by sex and marital status. Enquete Emploi 2009

Figure 3: Log wage density by sex and marital status. PSID 2009

Wages are much more dispersed in the U.S. Married men and women have higher wages than singles. And single men have higher wages than single women.
Men have higher BMI than women and married males have higher BMI than single men but married women have lower BMI than single women. It can be due to an effect of selection on the marriage market: only the slimest are attractive and find a partner. Weight and marital status are linked for several reasons which are exposed in the article of Averet et al. ([1], 2008)

4 Estimation

4.1 Specification

Agent type

Each individual is defined by two characteristics: his wage \((w_i)\) and another characteristic \((k_i)\). So each individual is indexed by \((w_i, k_i)\).

We note \(m(w_i, w_i)\) the joint density of the index \(k_i\) and the wage \(w_i\) among \(m\) population. We also note \(n(w_i, k_i, w_j, k_j)\), and \(a(w_i, k_i, w_j, k_j)\) the respective density of couples of characteristics \((w_i, k_i)\) and \((w_j, k_j)\) and the probability of the match of a man with characteristics \((w_i, k_i)\) and a woman of index \((w_j, k_j)\).

Then integrating equation (9) on these two characteristics among the female population, we have:

\[
\int_{w_f, k_f} \delta n(w_i, k_i, w_j, k_j) N = \lambda u_m(w_i, k_i) U_m U_f \int_{w_f, k_f} u_f(w_j, k_j) a(w_i, k_i, w_j, k_j)
\]

\[
\delta (m(w_i, k_i) - u_m(w_i, k_i)) = \lambda u_m(w_i, k_i) \int_{w_f, k_f} u_f(w_j, k_j) a(w_i, k_i, w_j, k_j)
\]
which gives:

$$u_m(w_i, k_i) = \frac{m(w_i, k_i)}{1 + \frac{1}{\delta} \sum_{w_j, k_j} u_f(w_j, k_j) a(w_i, k_i, w_j, k_j)}$$

$$u_f(w_j, k_j) = \frac{f(w_j, k_j)}{1 + \frac{1}{\delta} \sum_{w_i, k_i} u_m(w_i, k_i) a(w_i, k_i, w_j, k_j)}$$

(11)

Possible specifications of the match quality

Following the explanations of homophany which were presented in the introduction, the match quality of marriage can be modelled in different ways. If we follow the sociological explanation of homophilia, the match quality depends on the gap between the education levels of the members of a couple, then more close they are in education level and more the match quality is high. A function of this gap could be the following one.

$$K_{mij} = K_{fij} = \exp \left( - \frac{(dip_i - dip_j)^2}{2\sqrt{dip_i}} \right)$$

Otherwise, following the ranking explanation, the match quality is just increasing is the level of diploma of the other member. Then the quality of the match for an agent $i$ only depends on the education level of his partner $j$.

$$K_{mij} = dip_{fj}$$

$$K_{fij} = dip_{mi}$$

If we use the data on the Body Mass Index, the match quality of an agent $i$ can rise when the BMI of his partner is “attractive”. Men prefer slim women but women prefers strong men. Using the distribution of the BMI in the US population (cf. figure (4)), I define the following index of match quality:

$$K_{mij} = \exp \left( - \frac{(BMI_j - 20)^2}{2\sqrt{BMI}} \right)$$

$$K_{fij} = \exp \left( - \frac{(BMI_i - 25)^2}{2\sqrt{BMI}} \right)$$

4.2 Estimation

Using the data, we can compute the matching probability:

$$a(w_i, k_i, w_j, k_j) = \frac{\delta}{\lambda} \frac{n(w_i, k_i, w_j, k_j) * N}{u_m(w_i, k_i) * u_f(w_j, k_j) * U_m * U_f}$$

We estimate the density functions with Kernel estimation. Details are presented in appendix.\footnote{The computation of this model needs a mathematical software and a sufficient memory storage to stock 4-D matrices. Indeed, each couple has 4 characteristics and using numerical integration, we have to compute the joint densities of couples on a grid of 30 points. We need about two hours to generate the matrix of the joint density of couple $n(w_i, k_i, w_j, k_j)_{ij}$ or the matrix of conditional mean of hours for married people $h_m(w_i, k_i, w_j, k_j)_{ij}$.}
Assuming that the distribution function of $z$ is $G(z) = \Phi(\frac{z}{\sigma})$ with $\Phi$ a function of variance 1 and mean 0 (we'll use a standard gaussian law), we can also compute the total surplus of the match using the formula (8)

$$S(w_i, k_i, w_j, k_j) = -G^{-1}(1 - a(w_i, k_i, w_j, k_j))$$

$$= -\sigma_z \Phi^{-1}(1 - a(w_i, k_i, w_j, k_j))$$

Then, the computation of the present value of single is now possible. Indeed, we have

$$1(W_{mijz} > U_{mi})1(W_{fjiz} > U_{fj}) > 0 \iff S(k_i, w_i, w_j, k_j) + z \geq 0$$

then after rearranging the equation of transfers 6 and the individual surplus equation 5, we obtain the following Bellman equation for a single individual:

$$rW_{mi}(w_i, k_i) = v_{mi} + \lambda(\bar{E}_{fjz}(\max(S_{mij}(t_{mijz}), 0)))$$

$$= v_{mi} + \lambda \bar{E}_{fjz}\left(\max\left(\frac{\beta}{(r + \delta)}B_{mi}(w_i)(S(k_i, w_i, k_j, w_j + z), 0]\right)\right)$$

$$= v_{mi} + \frac{\beta \lambda}{(r + \delta)B_{mi}(w_i)} \int \int_{k_f, w_{fj}} \max(S(k_i, w_i, k_j, w_j + z), 0)u_f(k_j, w_j)dG(z)$$

we then simplify the last expression using the term of equation (7), $s_m(i)$ to write the following formulas:

$$s_m(w_i, k_i) = \frac{\beta \lambda}{(r + \delta)} \int \int_{k_f, w_{fj}} \max(S(w_i, k_i, w_j, k_j) + z, 0)u_f(w_j, k_j)dG(z)$$

$$s_m(w_i, k_i) = \frac{\beta \lambda}{(r + \delta)} \int \int_{k_f, w_{fj}} \left(\int \max(S(w_i, k_i, w_j, k_j) + z, 0)dG(z)\right)du_f(w_j, k_j)$$

Then

$$\int \max(S(w_i, k_i, w_j, k_j) + z, 0)dG(z) = S(w_i, k_i, w_j, k_j)a(w_i, k_i, w_j, k_j) + \int^{+\infty}_{-S(w_i, k_i, w_j, k_j)} zdG(z)$$

$$= S(w_i, k_i, w_j, k_j)a(w_i, k_i, w_j, k_j) + \sigma \int^{+\infty}_{-S(w_i, k_i, w_j, k_j)} wd\Phi(z)$$

$$= S(w_i, k_i, w_j, k_j)a(w_i, k_i, w_j, k_j) + \sigma \phi\left(\frac{S(w_i, k_i, w_j, k_j)}{\sigma}\right)$$

$$= \mu(w_i, k_i, w_j, k_j)$$

then

$$s_m(w_i, k_i) = \frac{\beta \lambda}{(r + \delta)} \int_{k_f, w_{fj}} \mu(a(w_i, k_i, w_j, k_j))du_f(w_j, k_j)$$

and by symmetry

$$s_f(w_j, k_j) = \frac{(1 - \beta)\lambda}{(r + \delta)} \int_{k_{mi}, w_{mi}} \mu(a(w_i, k_i, w_j, k_j))du_m(w_i, k_i)$$
So we can compute the domestic good:

\[
C(w_i, k_i, w_j, k_j) = S(w_i, k_i, w_j, k_j) - w_i T - w_j T + A_{mi}(w_i) + A_{fj}(w_j) \\
+ \frac{s_f(w_j, k_j) + w_i T - A_{fj}(w_j)}{K_f(k_i, k_j)^{d_f}} + \frac{s_m(w_i, k_i) + w_i T - A_{mi}(w_i)}{K_m(k_i, k_j)^{d_m}}
\]

### 4.3 Inference from hours

To compute the transfers and the public good, I need the identification of \( A_f \) and \( A_m \). If \( B_f \) and \( B_m \) are known functions, \( A_f \) and \( A_m \) can be recovered from hours worked by single individuals. Indeed, by setting the transfer at zero in the equation (2), you can write the following linear differential equation:

\[
\frac{d(w_i T - A_{mi}(w_i))}{dw_i} - \frac{B'_{mi}(w_i)}{B_{mi}}(w_i T - A_{mi}(w_i)) = h^0_{mi}(w_i) \\
\frac{d(w_j T - A_{fj}(w_j))}{dw_j} - \frac{B'_{fj}(w_j)}{B_{fj}}(w_j T - A_{fj}(w_j)) = h^0_{fj}(w_j)
\]

whose solution is

\[
w_i T - A_{mi}(w_i) = B_{mi}(w_i) \int_0^{w_i} \frac{h^0_{mi}(w)}{B_{mi}(w)} \, dw \\
w_j T - A_{fj}(w_j) = B_{fj}(w_j) \int_0^{w_j} \frac{h^0_{fj}(w)}{B_{fj}(w)} \, dw
\]

As we know the hours worked by the members of couples, the equation (2) gives us a formula to estimate \( B_{mi} \). Indeed, we have

\[
h_m(i, j, z) = T - A'_{mi}(w_i) - \frac{B'_{mi}(w_i)}{B_{mi}(w_i)}(w_i T + t_m(i, j, z) - A_{mi}(w_i)) \\
h_m(i) = T - A'_{mi}(w_i) - \frac{B'_{mi}(w_i)}{B_{mi}(w_i)}(w_i T - A_{mi}(w_i))
\]

then

\[
h_m(i, j, z) - h_m(i) = -\frac{B'_{mi}(w_i)}{B_{mi}(w_i)} t_m(i, j, z) \tag{14}
\]

Then integrating the preceding equation, we get

\[
\int_{z>z_{xy}} h_m(w_i, k_i, w_j, k_j, z) - h_m(k_i, w_i) = -\frac{B'_{mi}(w_i)}{B_{mi}(w_i)} \int_{z>z_{xy}} t_m(w_i, k_i, w_j, k_j, z)
\]

\[
\overline{h_m(w_i, k_i, w_j, k_j)} - \overline{h_m(k_i, w_i)} = -\frac{B'_{mi}(w_i)}{B_{mi}(w_i)} \left( \overline{h_m(w_i, k_i, w_j, k_j)} - h_m(k_i, w_i) \right)
\]

We will consequently regress the ratio \( \overline{h_m(w_i, k_i, w_j, k_j)} - \overline{h_m(k_i, w_i)} \) on \( \overline{t_m(z>z_{xy})} \) and use the following formula

15
\[ \frac{B_m'(w_i)}{B_m(w_i)} = -\frac{\int \left( \overline{h_m}(w_i, k_i, w_j, k_j) - h_m(k_i, w_i) \right) \overline{T_m}(w_i, k_i, w_j, k_j)n(w_i, k_i, w_j, k_j) dj}{\int \overline{T_m}(w_i, k_i, w_j, k_j)^2 n(w_i, k_i, w_j, k_j) dj} \]

Remind the equation (6) to write the following formula:

\[ \overline{t}_{m,i} = -w_i T + A_{mi}(w_i) + \frac{s_m(w_i, k_i) + w_i T - A_{mi}(w_i)}{K_m(k_i, k_j)^{d_m}} + \beta \mathbb{E}(S(i, j) + z| i, j, z > -S(i, j)) \]

\[ = -w_i T + A_{mi}(w_i) + \frac{s_m(w_i, k_i) + w_i T - A_{mi}(w_i)}{K_m(k_i, k_j)^{d_m}} + \beta \frac{\mu(a(w_i, k_i, w_j, k_j))}{a(w_i, k_i, w_j, k_j)} \]

Then if \( d_m \neq 0 \), the transfers are also expression of \( A_{mi} \), so we have actually a complex expression for \( A_{mi} \) which has the following form:

\[ A_{mi}(w_i) = T(A_{mi}(w_i), d_m, d_f) \]

More precisely, \( F(w_i) = A_{mi} - w_i T \) is the solution of the following differential equation:

\[ F'(w) = -\frac{F(w)^2 \int \Delta h_m(1 - \frac{1}{K_m})n(x, y) dy + F(w) \int \Delta h_m(\frac{s_m}{K_m} + \beta S x y)n(x, y) dy}{F(w)^2 \int (1 - \frac{1}{K_m})^2 n(x, y) dy + 2F(w) \int (\frac{s_m}{K_m} + \beta S x y)(1 - \frac{1}{K_m}) n(x, y) dy + \int (\frac{s_m}{K_m} + \beta S x y)^2 n(x, y) dy} \]

Then solving the differential equations\(^5\) allows you to recover \( A_m \) and \( A_f \) and calculate the transfers and the worked hours. I repeat the entire process for each value of \( d_m \) and \( d_f \) and choose the couple \((d_m, d_f)\) which minimizes the differences between estimated hours and observed worked hours.

### 4.4 Characterisation of the equilibrium

An equilibrium is a fixed point of \((u_m, u_f, s_m, s_f)\) of the following system of equations where the first two equations determine equilibrium wage distributions for singles (derived from (11)) and the last two equations determine equilibrium present values of singleness.

\[ u_m(w_i, k_i) = \frac{1}{1 + \frac{\lambda}{r + \delta}} \int_{k_j, w_j} u_f(w_j, k_j)a(w_i, k_i, w_j, k_j) \]

\[ u_f(w_j, k_j) = \frac{1}{1 + \frac{\lambda}{r + \delta}} \int_{k_i, w_i} u_m(w_i, k_i)a(w_i, k_i, w_j, k_j) \]

\[ s_m(w_i, k_i) = \frac{\beta}{r + \delta} \int_{w_i, k_i} \left( \int_z \max(S(w_i, k_i, w_j, k_j) + z, 0) dG(z) \right) du_f(w_j, k_j) \]

\[ s_f(w_j, k_j) = \frac{(1 - \beta)}{r + \delta} \int_{w_i, k_i} \left( \int_z \max(S(w_i, k_i, w_j, k_j) + z, 0) dG(z) \right) du_m(w_i, k_i) \]

with

\[ a(w_i, k_i, w_j, k_j) = 1 - G \left( -C(i, j) - w_i T - w_j T + A_m(w_i) + A_f(w_j) + \frac{s_f(w_j, k_j) + w_i T - A_f(w_j)}{K_f(k_i, k_j)^{d_f}} + \frac{s_m(w_i, k_i) + w_i T - A_m(w_i)}{K_m(k_i, k_j)^{d_m}} \right) \]

\(^5\)I work with the matlab function ode45
and $\lambda = \frac{M(U_m, U_f)}{U_m U_f}$.

In this section we want to verify that we can go backward, that is, calculate the equilibrium wage distributions and labor supply functions from the previous nonparametric estimates of the structural parameters, namely, the marriage externality function $C(i, j)$ and the preference parameters. To calculate the equilibrium we postulated a Cobb-Douglas meeting function $M(U_m, U_f) = M_0 \sqrt{U_m U_f}$ and estimated $M_0$ as $M_0 = \lambda \sqrt{U_m U_f}$ for the calibrated value of $\lambda$. Despite the lack of a global contraction mapping property, I found that the standard fixed-point iteration algorithm, $x_{n+1} = T x_n$ worked well in practice, even starting far from the equilibrium (likewith $s_m(w_i, k_i) = 0$ and $u_m(w_i, k_i) = m(w_i, k_j)$).

We have to choose the value of the set of fixed parameters $(\delta, r, \beta, \sigma_z)$. Following empirical data, we choose the following sets. Then we estimate $\lambda$.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>$\delta$</th>
<th>$r$</th>
<th>$\beta$</th>
<th>$\sigma_z$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enquete Emploi 2009</td>
<td>$1.3$</td>
<td>$1$</td>
<td>0.5</td>
<td>1000</td>
<td>$1.9 \times 10^{-4}$</td>
</tr>
<tr>
<td>PSID 2009</td>
<td>$1.3$</td>
<td>$1$</td>
<td>0.5</td>
<td>1000</td>
<td>$1.9 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

For each dataset, the algorithm converges to the equilibrium observed in the data. Indeed, we obtain the fixed point $(u_m, u_f, s_m, s_f)$ which corresponds to the density and present value we observe in the data.

5 Results: matching on education level

The data confirm the homogamy phenomena. I found a correlation of 0.57 between the diploma of men and women married together (the correlation of their wages is about 0.40).

Figure 5 presents the joint density of hourly wages (in €) among couples and the joint distribution of diplomas among couples. The correlation of wages among couples is not evident. Many different configurations can be obtained.
Figure 5: Couple density according to log wages. The highest density is in red. Enquête Emploi 2009

Following the equation (12), we can compute the probability of a match according to their wages. We can see on the figure 6 that the match probability is increasing in the wages of both members. It is also increasing in the education level of women. Educated women seem to be attractive on the marriage market.

Figure 6: Projection of Match probability on household members’ wages. Enquête Emploi 2009

**Present value of singles**

Following the estimation method presented in section 4, I compute the present value of singles according to their education level and their wages. Integrating the present value $s_m(w_i, k_i)$ on all diplomas of single men and $s_m(w_i, k_i)$ on all wages of single men, I present on the figure 7, the average single present value of single men conditional on wages (upper graph) and conditional on diplomas (bottom graph). I do the same for single women present values.
Figure 7: Present value of single projected on wages and on education level. Enquête Emploi 2008

The figure 7 shows that single men present value rises sharply with wages between 10 € and 20 € then the slope is lower. The more single men are productive, the more they can make a good match from which they can obtain a large part of the surplus. The present value of women rises much more strongly with her diploma than it does for men. Generally men have higher single present value than women. Indeed, they are less numerous making them more attractive, besides they will capt a higher share of the match surplus when they marry.

Transfers and public good

We can compute the transfers and the public good following the estimation strategy of section 4. In a first step I will estimate the simple model where $d_m = d_f = 0$, that is there is no impact of match quality in the model. In that case, the expression of $t_m$ and $t_f$ are very simple. Indeed, we have

$$t_m(i, j, z) - s_m(i) = \beta (S(i, j) + z)$$
$$t_f(i, j, z) - s_f(j) = (1 - \beta) (S(i, j) + z)$$

and taking the average on formed couple, we have:
\[
\begin{align*}
\overline{t_{m,i,j}} &= s_m(w_i, k_i) + \beta \mathbb{E}(S(i, j) + z | i, j, z > -S(i, j)) \\
\overline{t_{f,i,j}} &= s_f(w_j, k_j) + \beta \mathbb{E}(S(i, j) + z | i, j, z > -S(i, j))
\end{align*}
\]

and

\[
C(i, j) = S(i, j) + s_f(j) + s_m(i)
\]

I construct the matrices \(C(i, j), t_{m,i,j},\) and \(t_{f,i,j},\) and I use polynomial method to compute \(C(i, j), t_{m,i,j},\) and \(t_{f,i,j}\) for each couple of my data base (details are in appendix). Then I can regress them on the variables of the model. The following table presents the results.
Table 1: Household production and transfers for $d_m = d_f = 0$

<table>
<thead>
<tr>
<th></th>
<th>Cxy (Std.Err.)</th>
<th>Tfxy (Std.Err.)</th>
<th>Tmxy (Std.Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2033.10</td>
<td>455.20</td>
<td>209.38</td>
</tr>
<tr>
<td>Wages</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$w_m$</td>
<td>11.91</td>
<td>2.33</td>
<td>8.60</td>
</tr>
<tr>
<td>$w_m^2$</td>
<td>2.39</td>
<td>1.26</td>
<td>-0.03</td>
</tr>
<tr>
<td>$w_m^3$</td>
<td>-0.06</td>
<td>-37.62</td>
<td>2.03</td>
</tr>
<tr>
<td>$w_f$</td>
<td>-128.72</td>
<td>2.93</td>
<td>0.36</td>
</tr>
<tr>
<td>$w_m^2$</td>
<td>8.21</td>
<td>0.30</td>
<td>(0.01)</td>
</tr>
<tr>
<td>$w_m^3$</td>
<td>-0.16</td>
<td>-0.06</td>
<td>(0.001)</td>
</tr>
<tr>
<td>$w_m * w_f$</td>
<td>0.43</td>
<td>-0.11</td>
<td>-0.16</td>
</tr>
<tr>
<td>$</td>
<td>w_m - w_f</td>
<td>$</td>
<td>-11.21</td>
</tr>
</tbody>
</table>

Male diploma

<table>
<thead>
<tr>
<th></th>
<th>Ref.</th>
<th>Ref.</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>diph = 2</td>
<td>-27.50</td>
<td>0.19</td>
<td>-15.58</td>
</tr>
<tr>
<td>diph = 3</td>
<td>-45.26</td>
<td>3.04</td>
<td>-40.00</td>
</tr>
<tr>
<td>diph = 4</td>
<td>-82.00</td>
<td>1.17</td>
<td>-57.90</td>
</tr>
<tr>
<td>diph = 5</td>
<td>-84.60</td>
<td>2.40</td>
<td>-71.37</td>
</tr>
<tr>
<td>diph = 6</td>
<td>-220.90</td>
<td>-2.50</td>
<td>-131.72</td>
</tr>
<tr>
<td>diph = 7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Female diploma

<table>
<thead>
<tr>
<th></th>
<th>Ref.</th>
<th>Ref.</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>dipf = 2</td>
<td>11.63</td>
<td>1.00</td>
<td>0.71</td>
</tr>
<tr>
<td>dipf = 3</td>
<td>33.19</td>
<td>2.15</td>
<td>6.31</td>
</tr>
<tr>
<td>dipf = 4</td>
<td>23.86</td>
<td>-0.44</td>
<td>4.44</td>
</tr>
<tr>
<td>dipf = 5</td>
<td>29.85</td>
<td>-0.27</td>
<td>9.18</td>
</tr>
<tr>
<td>dipf = 6</td>
<td>26.36</td>
<td>0.21</td>
<td>3.77</td>
</tr>
<tr>
<td>dipf = 7</td>
<td>18.62</td>
<td>2.00</td>
<td>3.25</td>
</tr>
</tbody>
</table>

$|diph - dipf|$

<table>
<thead>
<tr>
<th></th>
<th>Ref.</th>
<th>Ref.</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.63</td>
<td>0.20</td>
<td>0.72</td>
</tr>
</tbody>
</table>

The public good is increasing in male’s wage and decreasing in female’s wage. It is also decreasing in the male education level. The larger is the gap between the education levels of the household members, the larger is the public good. Here, we observe an aditional explanation for homogamy: closer the members are in education level and higher is their household production of economy of scales. The transfer to male is strongly increasing in male’s wage and decreasing in female’s wage. The more the male is educated, the smaller is his part of the public good. The transfer to the female is increasing in both members’ wages and is strongly increasing in her education level.
The figure 8 presents the average transfer to the male conditional on his hourly wage (on the left) and conditional on his education level (on the right). The figure 9 presents the average transfer to the female conditional on her hourly wage (on the left) and conditional on his education level (on the right).

![Figure 8: Projection of $t_m$ on wages (on the left) and on diploma (on the right)](image1)

![Figure 9: Projection of $t_f$ on wages (on the left) and on diploma (on the right)](image2)

Figure 8 shows that men capt an increasing part of the public good when their hourly wage rise and that men’s part doesn’t vary a lot with their education level nor with their wife’s wage. The bargaining power of men is then increasing in their wage. Figure 9 shows that female’s part of the public good in only increasing in her wage when her wage is between 5€ and 15€.
Caracterisation of the sharing rule

Figure 10: Projection of the sharing rule on wages (on the left) and on diploma (on the right)

The sharing rule is the ratio of the transfer to the man over the sum of transfers that is the public good. This ratio is between 0.4 and 0.7. Men cap a larger part of the surplus. This is strongly increasing in his wage between 5 € and 20 € and the stabilises.

Is leisure a normal good?

The estimation strategy of section 4.4 shows how to estimate $B_m(w_i, k_i)$, $b_m(w_i, k_i)$, $B_f(w_j, k_j)$, $b_f(w_j, k_j)$, $A_f(w_j, k_j)$, and $A_m(w_i, k_i)$, from the observation of hours worked. I estimate them on data and I integrate the six previous functions on all diplomas. I present on the figure 11 the average of these functions conditional on wages. The results are close to those estimated by Jacquemet-Robin in their unidimensional model of search. (I present in appendix the result of their model estimated on French Data. Their estimated $B_m$, $b_m$, $A_m$, $B_f$, $b_f$, $A_f$ are presented on Figure 29).

\[ b_m(w_i, k_i) = \frac{b_m'(w_i, k_i)}{B_m'(w_i, k_i)} \text{ and } b_f(w_j, k_j) = \frac{b_f'(w_j, k_j)}{B_f'(w_j, k_j)} \]
The results show that leisure is an inferior good for men and women. Indeed, when transfers to male increase, male increase their number of worked hours according to the equation 14. Jacquemet and Robin also found this result. However, leisure seems to be a normal good for women who decrease their labor supply when their resources rise. Minimal consumption naturally rises with wage.

Hours

The equation 14 gives us a way to estimate the hours worked for married individuals. Figure 12 presents the difference between the average of worked hours for married people conditional on wages that I estimated on data with Kernel estimation and the estimated average I made using the model. The error term doesn’t exceed 20% on the entire grid of wages for men but is very important for women on the edges of the grid (for women with very high wages married to men with very low wages). When I compute the average of hours worked conditional on only one’s wage, the accuracy is better. Indeed, I present on the Figure 13 the difference between the observed average of hours of married men conditional on their wages and the estimated average of hours by the model. The error doesn’t exceed 6% for men but can reach 8% for women.
Figure 12: $100 \times (\text{hmxhat}-\text{hmx}) / \text{hmx}$
6 Results: matching on physical attractiveness

Figure 14 presents the joint density of hourly wages (in $) among couples and the joint distribution of BMI among couples. Hourly wages are not very dispersed. The correlation between wages is 0.33 whereas the correlation between BMI is 0.27.

Figure 14: Couple density according to wages (graph on the left) and according to their BMI (graph on the right). The highest density is in red. PSID 2009

The matching probability is presented on Figure 15. It is increasing in wages of both members of the household but much more rapidly in the male wage. The match probability is also higher for couple with similar BMI. Indeed, we observe a high diagonal on the graph. Competition
for the slimmest people lead to positive assortative mating: the slim match the slim and the fat match the fat.

Figure 15: Match probability according to wages (graph on the left) and according to their BMI (graph on the right). The highest density is in red. PSID 2009

Present value of singles

I present on the figure 16, the average single present value of single men conditional on wages (upper graph) and conditional on BMI (bottom graph). I do the same for single women present values.
As in the preceding estimation, single men present value rise with their wage, the relation is also true for women. The more single are productive and the more they can make a good match from which they can obtain a large part of the surplus.

The relation between single present value and their BMI is interesting. The slimmer the women are and the higher their present value is: slim women have high probability to make a good match and keep a large part of the surplus. This is not the same for men. Their present value rises with their BMI until it reaches 24 then stabilizes and decreases after 33. Men seem to be more attractive with a BMI between 24 and 33.

**Characterisation of transfers and household production**

We can compute the transfers and the public good following the estimation strategy of section 4. In a first step I will estimate the simple model where $d_m = d_f = 0$ as in the preceding section.
Table 2: Household production and transfers for matching on BMI with $d_m = 0$, $d_f = 0$

<table>
<thead>
<tr>
<th>Variable</th>
<th>$C_{xy}$</th>
<th>Std.error</th>
<th>$tm_{xy}$</th>
<th>Std.error</th>
<th>$tf_{xy}$</th>
<th>Std.error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2127.90</td>
<td>(99,15)</td>
<td>-1235.90</td>
<td>(28,43)</td>
<td>868.17</td>
<td>(20,08)</td>
</tr>
<tr>
<td>Wage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$w_h$</td>
<td>26.49</td>
<td>(1,44)</td>
<td>17.86</td>
<td>(0,41)</td>
<td>-1.11</td>
<td>(0,36)</td>
</tr>
<tr>
<td>$w_h^2$</td>
<td>0.12</td>
<td>(0,04)</td>
<td>-0.12</td>
<td>(0,01)</td>
<td>0.06</td>
<td>(0,01)</td>
</tr>
<tr>
<td>$w_h^3$</td>
<td>-0.002</td>
<td>(0,0003)</td>
<td>-0.0003</td>
<td>(0,0001)</td>
<td>-0.0004</td>
<td>(0,0001)</td>
</tr>
<tr>
<td>$w_f$</td>
<td>11.69</td>
<td>(0,65)</td>
<td>9.74</td>
<td>(0,57)</td>
<td>-0.13</td>
<td>(0,02)</td>
</tr>
<tr>
<td>$w_f^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$w_f^3$</td>
<td>0.0006</td>
<td>(0,0003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$w_h * w_f$</td>
<td>0.03</td>
<td>(0,02)</td>
<td>0.02</td>
<td>(0,002)</td>
<td>0.02</td>
<td>(0,01)</td>
</tr>
<tr>
<td>$</td>
<td>w_m - w_f</td>
<td>$</td>
<td>-14.68</td>
<td>(0,48)</td>
<td>-0.83</td>
<td>(0,11)</td>
</tr>
<tr>
<td>Body Mass Index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$bmi_h$</td>
<td>123.48</td>
<td>(5,46)</td>
<td>97.34</td>
<td>(1,56)</td>
<td>-3.65</td>
<td>(0,67)</td>
</tr>
<tr>
<td>$bmi_h^2$</td>
<td>-2.89</td>
<td>(0,09)</td>
<td>-1.56</td>
<td>(0,03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$bmi_f$</td>
<td>-147.96</td>
<td>(3,76)</td>
<td>-6.82</td>
<td>(1,08)</td>
<td>-34.65</td>
<td>(0,93)</td>
</tr>
<tr>
<td>$bmi_f^2$</td>
<td>0.92</td>
<td>(0,06)</td>
<td>0.04</td>
<td>(0,02)</td>
<td>0.36</td>
<td>(0,01)</td>
</tr>
<tr>
<td>$bmi_h * bmi_f$</td>
<td>2.48</td>
<td>(0,10)</td>
<td>0.12</td>
<td>(0,03)</td>
<td>0.15</td>
<td>(0,02)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.92</td>
<td>0.94</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The household public good increases with both household member’s wages. It also increases with the Body Mass index of the man but strongly decreases with the BMI of the woman. The transfer to the man is increasing with man wage. A rich man has a larger bargaining power and it decreases with the woman wage. Indeed, she gains some bargaining power when she gets a higher wage. Transfers to women increase with their wage and decrease with their spouse’s wage. Transfers to women decrease strongly with their BMI. Indeed, if she is less attractive, she loses some bargaining power. For the same reason, the transfer to the woman is strongly decreasing in her BMI.

Figure 17 and 18 show the average of transfers condition on their wages and their BMI.
Figure 17: Projection of $t_m$ on wages (on the left) and on BMI (on the right)

Figure 18: Projection of $t_f$ on wages (on the left) and on BMI (on the right)
Sharing rule

Figure 19: Projection of the sharing rule on wages (on the left) and on BMI (on the right)

The ratio of the man transfer on the total public good varies between 0.35 and 0.7 but is generally higher than 0.5. It increases sharply in his wage and decreases in her wage. The average of this ratio conditional on BMI shows that man’s part of the surplus rises with the BMI of the woman and is at the top when is BMI is 30.
Is leisure a normal good?

Figure 20: Aggregated prices

Hours fit

Figure 21: 100*(hmxyhat-hmxy)/hmxy
Figure 22: 100\%(\frac{\hat{h}_m(w_i)-h_m(w_i)}{h_m(w_i)}, \frac{\hat{h}_f(w_j)-h_f(w_j)}{h_f(w_j)})

7 Conclusion

Results

- A multidimensional search and matching model of marriage which fits the data.
- Marriage according to diploma and according to BMI
- Men capt the largest part of the public good. Transfers to males rise with their wages and decrease with their diplomas. Transfers to female increase with their education level.
- Bargaining power of men rise with their wage, bargaining power of women rise with their education level
- Bargaining power of women strongly decrease with their physical attractiveness (BMI), which is not the case for men.
- Household public good is increasing in male wages and female education levels.

Caveats and extensions

- Estimation only possible on working people
- No inclusion of children
- Where does Household public good come from? It would be interesting to invest the problem of home production and time use.
References


8 Appendix

8.1 Descriptive statistics: enquête emploi 2008

Figure 23: Conditional mean of hours according to wages by marital status. Enquête Emploi 2009

8.2 Descriptive statistics: PSID 2009

Figure 24: Conditional mean of hours according to wages by marital status. PSID 2009
8.3 Results of the unidimensional level

Here, I present the result obtained by the model of Robin-Jacquetem (2011) implemented on the Enquête Emploi 2009 (They use the PSID). That is the characterisation of the public good and the transfers when you suppose that people match only on their wages. Following the equation (12), we can compute the probability of a match according to their wages. We can see on the two following graph that the match probability is increasing in the wages of both members of the couple.

![Match probability graph](image)

**Figure 25:** Match probability projected on wages. Enquête Emploi 2009

Then, the present value of a single individual is represented as a function of wages on the following graph. It is clear that present value of a single man rises with his wage. However, the relation is less clear for women...

![Individual value graph](image)

**Figure 26:** Present value of individuals

The public good created is all the more important since both wages of the members of the household are important.
Figure 27: Household public good created

Figure 28: Transfers

Characterisation of the public goods and the transfers:
<table>
<thead>
<tr>
<th></th>
<th>Cxy</th>
<th>Std.err.</th>
<th>Tmxy</th>
<th>Std.err.</th>
<th>Tfy</th>
<th>Std.err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2814,8</td>
<td>(7,08)</td>
<td>-117,41</td>
<td>(1,14)</td>
<td>308,71</td>
<td>(1,67)</td>
</tr>
<tr>
<td>$w_m$</td>
<td>152,82</td>
<td>(1,08)</td>
<td>88,67</td>
<td>(0,17)</td>
<td>3,13</td>
<td>(0,25)</td>
</tr>
<tr>
<td>$w_m^2$</td>
<td>-6,94</td>
<td>(0,07)</td>
<td>-4,09</td>
<td>(0,01)</td>
<td>-0,12</td>
<td>(0,02)</td>
</tr>
<tr>
<td>$w_m^3$</td>
<td>0,11</td>
<td>(0)</td>
<td>0,06</td>
<td>(0)</td>
<td>0,002</td>
<td>(0,0003)</td>
</tr>
<tr>
<td>$w_f$</td>
<td>-70,1</td>
<td>(1,07)</td>
<td>-2,93</td>
<td>(0,17)</td>
<td>2,09</td>
<td>(0,25)</td>
</tr>
<tr>
<td>$w_f^2$</td>
<td>2,97</td>
<td>(0,08)</td>
<td>0,1</td>
<td>(0,01)</td>
<td>-0,11</td>
<td>(0,02)</td>
</tr>
<tr>
<td>$w_f^3$</td>
<td>-0,05</td>
<td>(0)</td>
<td>-0,001</td>
<td>(0,0003)</td>
<td>0,001</td>
<td>(0,0004)</td>
</tr>
<tr>
<td>$w_m \times w_f$</td>
<td>1,38</td>
<td>(0,03)</td>
<td>0,08</td>
<td>(0,01)</td>
<td>0,03</td>
<td>(0,01)</td>
</tr>
<tr>
<td>$</td>
<td>w_m - w_f</td>
<td>$</td>
<td>-32,07</td>
<td>(0,25)</td>
<td>-1,29</td>
<td>(0,04)</td>
</tr>
</tbody>
</table>

| $R^2$ | 0,96 | 0,99 | 0,41 |

Table 3: Characterisation of transfers and household creation

Then, using data on hours worked, you can compute the aggregated price $B_m, B_f$ and the minimal consumption $A_m, A_f$. You obtain the following results.

![Graphs of Bmx, Bfy, bmfx, bfyx, Amx/T, Aly/T](image)

Figure 29: Aggregated prices and Minimal expenditure
Figure 30: Share of total earning that is created. (%)
Figure 32: 100% \left( \frac{h_m(w_i) - h_m(w_i)}{n_m(w_i)}, \frac{h_f(w_j) - h_f(w_j)}{n_f(w_j)} \right)

8.4 The Clenshaw-Curtis quadrature

Approximation of the numerical value of an integral

The Clenshaw-Curtis method consists in transforming the function (the integral of which we want to estimate) in a periodic function with the variable change \( x = \cos \theta \). Then we can use the Cosinus Discrete Transform of \( f(x) = f(\cos \theta) \) and evaluate this transform in few points (nodes).

We consider a function \( f(x) \) defined on the interval \([-1, 1]\) (With a variable change, we can reach all other intervals). Then we try to estimate the value of

\[
\int_{-1}^{1} f(x)dx = \int_{0}^{\pi} f(\cos \theta)\sin \theta d\theta
\]

\( f(\cos \theta) \) is a \( 2\pi \)-periodic function which is even and real. This function is then equal to the following Fourier series:

\[
f(\cos \theta) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(k\theta)
\]

After a double integration by parts, we show that

\[
\int_{0}^{\pi} f(\cos \theta)\sin \theta d\theta = a_0 + \sum_{k=1}^{\infty} \frac{2a_{2k}}{1 - (2k)^2}
\]

However, it is difficult to calculate the Fourier coefficients \( a_k \). Indeed, their exact formula is:

\[
a_k = 2 < f(\cos \theta)\cos(k\theta) > = \frac{2}{2\pi} \int_{-\pi}^{\pi} f(\cos \theta)\cos(k\theta)d\theta = \frac{2}{\pi} \int_{0}^{\pi} f(\cos \theta)\cos(k\theta)d\theta
\]

With \( < | > \), the notation for scalar product.

Then, we will try to evaluate the function \( f(\cos \theta) \) in \( N+1 \) points (nodes) equidistant on the segment \([0, \pi]\) and use the Cosinus Discrete Transform of type I. The \( N+1 \) nodes used are the \( \theta_n = \frac{n\pi}{N} \) with \( n = 0, 1, \ldots, N \).
Cosinus Discrete Transform and Discrete Fourier Transform

The Cosinus Discrete Transform of type I is then given by:

\[ f(\cos \theta_n) = \frac{1}{2} (f(\cos(0)) \cos(0) + f(\cos(\pi)) \cos(n\pi)) + \sum_{k=1}^{N-1} f(\cos(\frac{k\pi}{N})) \cos(\frac{k\pi}{N}n) \]

\[ = \frac{f(1)}{2} + \frac{f(-1)}{2}(-1)^n + \sum_{k=1}^{N-1} f(\cos(\frac{k\pi}{N})) \cos(\frac{k\pi}{N}n) \]

for \( n = 0, 1, ..., N \).

We then multiply the first and the last term by \( \frac{1}{2} \) (Euler-Maclaurin Formula).

Using the equation (17) and the Inverse function of the TCD of type I (which is the same transformation multiplied by the coefficient \( \frac{2}{N} \)), we can obtain an estimation of coefficients \( a_n \) given by:

\[ a_n \approx \frac{2}{N} \left[ \frac{f(1) + f(-1)}{2} + \sum_{k=1}^{N-1} f(\cos(\frac{k\pi}{N})) \cos(\frac{k\pi}{N}n) \right] \]

To calculate the integral of equation (18), it is enough to know the even coefficients pairs of the TCD. Their formula is then

\[ a_{2n} \approx \frac{2}{N} \left[ \frac{f(1) + f(-1)}{2} + f(0)(-1)^n + \sum_{k=1}^{N-1} \left( f(\cos(\frac{k\pi}{N})) + f\left(-\cos\left(\frac{k\pi}{N}\right)\right) \right) \cos\left(\frac{k\pi}{N}n\right) \right] \]

(19)

Theses are computable and we can have an estimation of the value of the wanted integral with the equation (18).

Weight determination

The Clenshaw-Curtis method allows to calculate quadrature weights \( w_k \) such that

\[ \int_{-1}^{1} f(x) dx = \sum_{k=0}^{N} w_k f(\cos(\theta_k)) + R_n \]

with \( R_n \), an approximation error.

Considering that \( N \) is even, we can calculate the \( \frac{N}{2} + 1 \) quadrature weights \( w_k \) verifying:

\[ \int_{-1}^{1} f(x) dx \approx \sum_{k=0}^{N/2} w_k \left( f(\cos(\theta_k)) + f(-\cos(\theta_k)) \right) \]

Besides, remind that

\[ \int_{-1}^{1} f(x) dx = a_0 + \sum_{k=1}^{\infty} \frac{2a_{2k}}{1 - (2k)^2} \]

Then we can calculate the weights \( w_k \) with the formula of \( a_{2k} \). We rewrite the equation (19) with the following matricial form:
\[
\begin{pmatrix}
a_0 \\
a_2 \\
\vdots \\
a_N \\
\end{pmatrix} = Dy = D
\begin{pmatrix}
y_0 \\
y_1 \\
\vdots \\
y_{N/2} \\
\end{pmatrix}
\]

with \( y_n = f(\cos(\theta_n)) + f(-\cos(\theta_n)) \) and \( D \), the TCD matrix such that:

\[
D_{i,j} = \begin{cases} 
\frac{1}{N} & \text{si } j = 1 \\
\frac{2}{N} \left( \cos \left( \frac{(i-1)(j-1)\pi}{N/2} \right) \right) & \text{si } j \notin \{1; N/2 + 1\} \\
\frac{1}{N}(-1)^{i-1} & \text{si } j = \frac{N}{2} + 1
\end{cases}
\]

Finally, the equation (18) let us write:

\[
\int_{-1}^{1} f(x)dx = d \begin{pmatrix}
a_0 \\
a_2 \\
\vdots \\
a_N \\
\end{pmatrix} = \begin{pmatrix}
1 & \frac{2}{1-1} & \frac{2}{1-16} & \ldots & \frac{2}{1-N^2} \\
\end{pmatrix} \begin{pmatrix}
a_0 \\
a_2 \\
\vdots \\
a_N \\
\end{pmatrix}
\]

So we have

\[
\int_{-1}^{1} f(x)dx = dDy = (dD)y = wy
\]

with \( w \) the vecteur of quadrature weights.

The quadrature weights are then:

\[
w_0 = \frac{1}{N} \left( 1 + \sum_{j=1}^{N/2} \frac{2}{1 - (2j)^2} \right)
\]

\[
w_{N/2} = \frac{1}{N} \left( 1 + \sum_{j=1}^{N/2} \frac{2(-1)^j}{1 - (2j)^2} \right)
\]

\[
w_k = \frac{2}{N} \left( 1 + \frac{(-1)^k}{1 - N^2} + \sum_{j=1}^{N/2-1} \frac{2}{1 - (2j)^2} \cos \left( \frac{2jk\pi}{N} \right) \right) \quad \forall k = 1, ..., N/2 - 1
\]

**Algorithm of Jorg Waldvogel**

J.Waldvogel using matrices, Féjer'quadrature and Discrete Fourier Transform derives a simple algorithm to obtain the weights of the Clenshaw-Curtis quadrature. He shows that the weights \( w = (w_0, w_1, ..., w_{N-1}) \) of the Clenshaw-Curtis quadrature rule are given by the inverse discrete
Fourier transform of the vector $v + g$, where $g$ and $v$ are defined below, and with $w_0 = w_N$.

$$v_k = \frac{2}{1 - (2k)^2}, k = 0, 1, \ldots, \left[\frac{N}{2}\right] - 1,$$

$$v_{\frac{N}{2}} = \frac{N - 3}{2 \left[\frac{N}{2}\right] - 1} - 1$$

$$v_{n-k} = v_k, k = 0, 1, \ldots, \left[\frac{N - 1}{2}\right]$$

$$g_k = -w_0, k = 0, 1, \ldots, \left[\frac{N}{2}\right] - 1,$$

$$g_{\frac{N}{2}} = w_0 [(2 - \text{mod}(N,2))N - 1]$$

$$g_{n-k} = g_k, k = 0, 1, \ldots, \left[\frac{N - 1}{2}\right]$$

The matlab code is then given by

```matlab
function [x,wcc] = ccquad(n)

% Clenshaw-Curtis quadratures by DFTs n>1
% Nodes: $x_k = \cos(\frac{k\pi}{n}), k = 0, \ldots, n$
% wcc = weights
% Compute $\int_{-1}^{1} f(x) dx = f * wcc$ for $f = [f(x_0)\ldots f(x_n)]$

K = [0 : n]'; x = cos(K*pi/n);
N = [1 : 2 : n - 1]; l =length(N); m = n - l;
v0 = [2./N/(N - 2); 1./N(end); zeros(m, 1)];
v2 = -v0(1 : end - 1) - v0(end : -1 : 2);

%Clenshaw-Curtis nodes: k = 0, 1, \ldots, n; weights: wcc, wccn = wcc0
g0 = -ones(n, 1); g0(1 + l) = g0(1 + l) + n; g0(1 + m) = g0(1 + m) + n;
g = g0/(n^2 - 1 + mod(n, 2)); wcc =real(ifft(v2 + g));
wcc = [wcc; wcc(1)];

8.5 Kernel density estimation

The kernel density estimation is a non-parametric way of estimating the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample.

Let $(x_1, x_2, \ldots, x_n)$ be an i.i.d sample drawn from some distribution with an unknown density $f$. We are interested in estimating the shape of this function $f$. Its kernel density estimator is:
\[ \hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) \]

where \( K() \) is the kernel. It is a symmetric but not necessarily positive function that integrates to one. The normal kernel is often used \( K(x) = \phi(x) \), where \( \phi \) is the standard normal density function.

The bandwidth \( h > 0 \) is a smoothing parameter. Intuitively one wants to choose \( h \) as small as the data allows, however there is always a trade-off between the bias of the estimator and its variance. The optimal bandwidth is of order of \( n^{-1/5} \), it minimises the asymptotic mean integrated squared error which is then of order of \( n^{-4/5} \).

It can be shown that, under weak assumptions, there cannot exist a non-parametric estimator that converges at a faster rate than the kernel estimator. Note that the \( n^{-4/5} \) rate is slower than the typical \( n^{-1} \) convergence rate of parametric methods.

### 8.5.1 Non-parametric estimation of joint density

We observe the value taken by three variables \( x, y \) on a population of size \( n \). Then, the estimation of the joint density of \( x \) and \( y \), \( f(x, y) \) on the \( n \) points \( (x_i, y_i) \) could be made with the Parzen Rosenblatt estimator on \( \mathbb{R}^2 \).

\[ \hat{f}_n(x, y) = \frac{1}{nh(n)^2} \sum_{i=1}^{n} K \left( \frac{1}{h(n)} \left( \frac{x_i - x}{h}, \frac{y_i - y}{h} \right) \right) \]

With \( K \) should be a Parzen-Rosenblatt kernel of \( \mathbb{R}^2 \) which means a function defined on \( \mathbb{R}^2 \) integrable and integrates to 1. \( K \) is so bounded and we have \( \lim_{||x|| \to +\infty} ||x||^2 K(x) = 0 \).

\( K \) could also be decomposed as : \( K(x, y) = K_1(x)K_1(y) \), with \( K_1: \mathbb{R} \to \mathbb{R} \). Then \( K_1 \) could be the probability distribution function of a standard normal distribution \( (x \to \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}) \).

Then we have

\[ \hat{f}_n(x, y) = \frac{1}{nh(n)^2} \sum_{i=1}^{n} K_1 \left( \frac{x_i - x}{h(n)} \right) K_1 \left( \frac{y_i - y}{h(n)} \right) \]

**Implementation**

Then, it is easy to implement it. Creation of the vector \( k_x = \frac{1}{h(n)\sqrt{2\pi}} \left( e^{-\left( \frac{x_1 - x}{h(n)} \right)^2}, ..., e^{-\left( \frac{x_n - x}{h(n)} \right)^2} \right) \)

and \( k_y = \frac{1}{h(n)\sqrt{2\pi}} \left( e^{-\left( \frac{y_1 - y}{h(n)} \right)^2}, ..., e^{-\left( \frac{y_n - y}{h(n)} \right)^2} \right) \) and compute \( \hat{f}_n(x, y) = \frac{1}{n} k_x k_y' \). This function is then evaluated for \( p \) values of \( x \)and \( y \) (the nodes of the Clenshaw-Curtis quadrature).
8.6 Non-parametric estimation of: $\mathbb{E}(h|x, y)$

We observe the value taken by three variables $x$, $y$ and $h$ on a population of size $n$. Then, the estimation of $\mathbb{E}(h|x, y)$ could be made with the Nadaraya-Watson estimator:

$$\hat{m}_n(x, y) = \frac{\sum h_i K \left( \frac{1}{n} \left( \frac{x_i - x}{h(n)} \right) \right)}{\sum K \left( \frac{1}{n} \left( \frac{y_i - y}{h(n)} \right) \right)}$$

With $K$, a bidimensional Parzen Rosenblatt kernel which can be decomposed as: $K(x, y) = K_1(x)K_1(y)$. And $K_1$ could be the probability distribution function of a standard normal distribution ($x \rightarrow \frac{1}{\sqrt{2\pi}}e^{-x^2}$).

$$\hat{m}_n(x, y) = \frac{\sum_{i=1}^{n} h_i K_1 \left( \frac{x_i - x}{h(n)} \right) K_1 \left( \frac{y_i - y}{h(n)} \right)}{\sum_{i=1}^{n} K_1 \left( \frac{x_i - x}{h(n)} \right) K_1 \left( \frac{y_i - y}{h(n)} \right)}$$

**Implementation**

Then the implementation is

$$\hat{m}_n(x, y) = \frac{1}{n} k_x \left( \begin{array}{cccc} h_1 & 0 & 0 & 0 \\ 0 & h_2 & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & h_n \end{array} \right) k_y'$$

46