



Marriage market and intra-household allocation: essays in economics of family and education

Marion Goussé

► To cite this version:

Marion Goussé. Marriage market and intra-household allocation: essays in economics of family and education. Economics and Finance. Institut d'études politiques de paris - Sciences Po, 2014. English. NNT : 2014IEPP0018 . tel-03516515

HAL Id: tel-03516515

<https://theses.hal.science/tel-03516515>

Submitted on 7 Jan 2022

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

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

Institut d'Etudes Politiques de Paris
ECOLE DOCTORALE DE SCIENCES PO

Programme doctoral en économie

Doctorat en Sciences économiques

Marriage Market and Intra-Household Allocation

- *Essays in Economics of Family and Education* -

Marion Goussé

Thèse dirigée par

Jean-Marc Robin, Professeur des Universités, Sciences Po Paris

Soutenue à Paris le 17 Juin 2014

Jury :

M. Pierre-André Chiappori - Professor of Economics, Columbia University
M. Nicolas Jacquemet - Professeur des Universités, Université de Lorraine - *Rapporteur*
M. Thierry Mayer - Professeur des Universités, Sciences Po Paris
M. Jean-Marc Robin - Professeur des Universités, Sciences Po Paris
M. Bernard Salanié - Professor of Economics, Columbia University - *Rapporteur*

Remerciements

Il y a maintenant presque 4 ans, je cherchais un professeur en économie du travail qui voulait bien m'encadrer pour une thèse sur des expérimentations aléatoires sur les programmes de remplacement en emploi adressés aux chômeurs... Après quelques échecs, j'ai frappé à tout hasard à la porte du nouveau professeur d'économie de Sciences Po, Jean-Marc Robin. Peu emballé par mon sujet, il m'a tout de même proposé de m'encadrer pour une thèse sur des modèles structurels de formation des couples... Je n'ai pas compris tout de suite... Et puis je me suis lancée dans ce projet !

C'est donc un peu par hasard que je soutiens aujourd'hui une thèse en micro-économie structurelle sur le marché du mariage qui devrait m'emmener dans quelques mois au Canada. Un grand merci donc à mon directeur de thèse Jean-Marc Robin pour m'avoir offert cette opportunité et pour m'avoir fait travailler sur un sujet finalement très porteur et riche en enseignements. Au cours, de ma thèse, j'ai en plus eu la chance de partir travailler à l'université de Columbia à New-York pendant 3 mois où j'ai été fort bien accueillie par Pierre-André Chiappori et Bernard Salanié qui sont aujourd'hui membres de mon jury de soutenance. Je les en remercie beaucoup. Je remercie également Bernard Salanié pour avoir accepté de me soutenir pour le Job Market et pour être rapporteur de cette thèse. Je remercie aussi Nicolas Jacquemet d'avoir accepté d'être membre de mon jury et rapporteur pour ma thèse ainsi que Thierry Mayer pour avoir accepté d'être membre de ce jury mais également pour avoir pris le temps d'être placement officer pour le Job Market cette année. Je remercie aussi Robert Gary-Bobo qui est co-auteur du 4^{ème} chapitre de cette thèse et qui m'a accueillie plusieurs fois au CREST pour nos séances de travail. Merci également à Alfred Galichon pour son intérêt, ses conseils et son soutien pendant mon Job Market. Je remercie aussi Noémie Le Donné qui a co-écrit avec moi le 3^{ème} chapitre de cette thèse et qui en plus d'être une très bonne amie est une excellente collègue de travail. Enfin, je remercie aussi Guy Laroque pour avoir créé et pour faire vivre toutes les semaines le reading group qui connaît un grand succès auprès des étudiants.

J'aimerais aussi remercier ceux sans qui ma thèse serait probablement allée beaucoup plus vite... Tout d'abord mes chers compagnons du Job Market à savoir François, le bubbly guy, bientôt surfeur à L.A, Gabriel, qui appréciera bientôt les calories vides des empanadas de Madrid et Maxime, bientôt économètre flegmatique à Londres.

Au fil des années et des bureaux (de la salle J401 jusqu'au grenier en passant par la bibliothèque H405 mais aussi par des demi-salles obscures), j'ai cotoyé tout un tas de graines de chercheurs qui ont rendu ces quatre années de thèse à Sciences Po finalement très agréables. Je pense d'abord aux anciens, aux petites blagues de Thomas, aux soirées arrosées de Camille, à

l'inspiration d'Etienne et à l'incroyable énergie de Dilan mais aussi à la compagnie d'Elisabeth, d'Elizabeta, de Gong, de Jérôme, de Lisa, de Marlène et d'Urszula. Je pense aussi aux équipes successives des ATER qui viennent d'Aix-Marseille ou de PSE pour nous aider à donner des cours aux étudiants. Merci à Clément, Morgane et Sarra pour leur non-accent de Marseille, et merci au boulevard Jourdan de nous avoir envoyé Marc, Xavier, Arthur, l'ours tyran et Adrien, que je remercie pour une excellente coopération pour le TD d'économétrie. Merci aussi à Matthieu, LE post-doc du département et à Cécile pour organiser régulièrement des piques-niques de retrouvailles sur les quais de Seine ou aux Tuileries. J'en profite aussi pour souhaiter bon courage à tous les suivants, à Jamil, Corentin, Meradj, à Alexis qui assume enfin son orientation vers la sociologie et les sciences politiques, à Jules, le chouchou d'Alain Besoin, à Camilo, la Flecha de Colombia et à Guillaume le philosophe écrivain.

Enfin, il y a les petits que j'ai découverts à mon retour des Etats-Unis et qui ont véritablement dynamisé l'ambiance du bureau des doctorants. Je pense d'abord à Florian, notre regretté délégué qui embaumait le bureau d'huiles essentielles biologiques, à Guillaume, le dérouleur des Ardennes, à Valeria, l'amoureuse des archives qui pourtant (allez, on peut le dire maintenant) affiche en cachette des images de propagande en Economie, et bien sur à Elisa, Emanuele, Eve, Ivan et à Clément. Enfin, il y a aussi les vraiment tout petits qui ayant pour la plupart été mes anciens élèves de TD me font comprendre qu'il est grand temps que je leur laisse la place... Je pense ainsi à Assia, Clara, Eiji, Joli P, Quentin et Simon mais aussi à nos compagnons du LIEPP dont Jean-Louis, Léa, Guillaume et Odran qui participent à la bonne ambiance du grenier même s'ils maintiennent toujours secret leur précieux séminaire.

Si j'ai fait des études si longues c'est aussi probablement parce que c'est là que je trouve les amis que je garde au cours du temps. Merci à tous ceux-là de m'encourager sans me poser trop de questions. Je remercie ainsi la bonne équipe d'amis de lycée de Lorient, avec lesquels je fête les 10 ans du bac cette année, mes amis de prépa, les toorch d'LLG, mes amis de master et mes amis d'école, les statisticiens économistes, qui à Londres ou à Paris, font marcher les assurances et les salles de marchés.

Bien sur je remercie aussi la család Lepage pour leur accueil chaleureux et je remercie aussi énormément toute ma famille pour un soutien et un amour inconditionnel.

Enfin, un grand merci à Guillaume avec qui je négocie joyeusement depuis plus de sept ans, qui a un indice de masse corporelle parfait et un niveau d'étude tout à fait idéal et qui est prêt à tenter avec moi l'aventure au Canada.

Abstract

This dissertation deals with family formation, family organization and education systems. The first part focuses on the household formation and within-household resource sharing whereas the second part focuses on education of junior high schools students.

The first two chapters study how people choose their partners and how they share their income. First, I focus on household formation and I model the matching process as follows: people meet randomly and decide to match or not depending on the marriage gain they would benefit from living together. People can choose their partner according to their education level, their wage and their physical attractiveness. I use American data, and I observe who matches with whom and who stays single and for how long. I then recover the preferences of individuals in terms of mating and the efficiency of the marriage market. The second chapter attempts to understand how the efficiency and the sorting of the marriage market could impact economic outcomes such as wealth and income inequalities or men and women labor supplies. In this chapter, individuals work and enjoy leisure and consumption. When they marry, they share their income and decide how much each of them will work on the market and at home to raise children or do the housework. I estimate my model on British data. I recover the amount of monetary transfers which exist between household members and show that these transfers make married women work less on the market and married men work more.

The last two chapters of this dissertation deal with education of young students. They focus on the French education system and on the impact of grade retention policies. PISA data show that French students score have decreased between 2000 and 2009. In the third chapter which is a joint work with Noémie Le Donne, we use distributional decomposition methods to assess to which extent this decrease can be attributed to the changes in students characteristics or to the changes in school returns. We find that grade retention has a more negative effect in 2009 than in 2000. However, PISA data are not longitudinal and we can't disentangle what comes from a change in the retention return from what comes from a selection effect. In the fourth chapter of this thesis, which is a joint work with Robert Gary-Bobo and Jean-Marc Robin, we find an estimation strategy to get rid of this selection effect and we use a panel data on French high-school students to evaluate the impact of grade retention on their scores.

Contents

	Page
I Introduction	1
1 Economic Imperialism: Marriage and Education as Economic Issues . .	3
2 Family Formation and Intra-Household Allocation of Resources	5
3 Evaluation of Education Systems	8
II Marriage Market and Intra-Household Allocation	13
1 Marriage gains : who should you marry ?	15
1 Introduction	16
2 Model	18
3 Data	21
4 Estimation strategy	23
5 Estimation results	25
6 Simulation of equilibrium	31
7 Conclusion	35
Appendices	38
2 Marriage Market and Intra-Household Allocation	43
1 Introduction	44
2 Model	46
3 Data	49
4 Specification and estimation	55
5 Estimation results	59
6 Simulations	69
7 Extensions	75
8 Conclusion	77
Appendices	80
III Grade retention in French Junior High School	91
3 Why do Inequalities in 15-year-old Cognitive Skills Increase in France between 2000 and 2009?	93

1	Introduction	94
2	Main changes in French educational policies and their expected consequences on 15-year-old performances	95
3	French PISA 2000 and 2009 data	99
4	Decomposition techniques	103
5	Results	106
6	Discussion	115
	Appendices	121
4	Grade Retention and Unobserved Heterogeneity	129
1	Introduction	130
2	Data	132
3	Preliminary Analysis : IV Estimates	134
4	A Model of Knowledge-Capital Accumulation	138
5	Estimation Method	143
6	Estimation Results	144
7	The Treatment Effects of Grade Retention	149
8	Conclusion	156
	Appendices	160
	IV Conclusion	163

Part I

Introduction

"I was interested in social problems but felt that economics had the tools by which to handle these long-term interests and social questions"

Gary Becker, Nobel Lecture, 1992

1 Economic Imperialism: Marriage and Education as Economic Issues

Marriage, family and education issues do not belong to the traditional core of economics. Yet their impact on economic development, productivity growth, innovation, poverty and inequalities is huge. Economists then have to understand individual social behaviors to derive their aggregate effect on economic outcomes (Becker, 1993). Accumulation and transmission of human capital across generations is the main vector that economists focus on through which families and education impact economic development. For economists, human capital¹ is the stock of skills, knowledge, habits, social and personal attributes, including creativity and cognitive abilities which help to produce economic value. Social capital accumulates as any other type of capital: it can improve with education, experience or training².

Human capital is essential both at the macro-economic and at the micro-economic level. At the aggregate level, human capital is an important factor for economic development, productivity growth, and innovation (Aghion and Howitt, 1992; Romer, 1990). At the individual level, human capital (e.g education and non-cognitive skills) is the first determinant of labor market earnings (Heckman et al., 2006). In all societies, families play an essential role in the accumulation and the transmission of human capital. Family formation and intra-household organization may have consequently a significant impact on the formation of inequalities of education and income and on their transmission across generations. We can indeed identify three sources of inequalities naturally conveyed by family structures: homogamy, inter-generational transmission of human capital and role specialization within the household. First, preferences of individuals to mate with their likes mechanically increase inequalities across household (Greenwood et al., 2014). Second, these income inequalities are transmitted across generations as a large literature highlights the crucial role of the family environment in children human capital accumulation for both cognitive and non-cognitive skills (Carneiro and Heckman, 2003; Piketty, 2003; Cunha et al., 2010; Heckman et al., 2006). Besides, family income and market credit constraints also influence education investment decisions (Acemoglu and Pischke, 2001). Finally, within-household arrangements may also create inequalities through the gender division of roles within the household and through the decision made by the household members on the education choice for the children and on the part of the income which is devoted to them. Several studies show that when transfer payments are given to women rather than to their husbands,

¹Sociologists work with a different concept of human capital defined by Pierre Bourdieu which includes cultural capital, social capital, economic capital, and symbolic capital.

²Human capital has been a very controversial term as it may give the impression that it treats people as machines. Approaching schooling as an investment rather than a cultural experience was considered unfeeling and extremely narrow. It has been elected the Un-word of the year in 2004.

expenditures on children increase (Thomas, 1993; Lundberg et al., 1997). Empowerment of women would then have a strong effect on development and growth (Doepke et al., 2012).³

The significant effect of family background on children's future earnings linked with the increase in assortative mating is fostering inequalities. Inequalities in earnings are worrying if they limit earning mobility across generations which appears to be the case in high-income countries (Autor, 2014; Corak, 2013). To counteract the trend in increasing inequalities, we need to promote children's human capital in a way that offers greater benefits to the more disadvantaged. Studying the determinants of schooling choices and the process of accumulation of knowledge by students is of great importance to build efficient education systems, to improve growth and innovation and to give equal opportunities of learning to all children. For a good understanding of the mechanisms at play, we need to answer several questions: what are the mechanisms of couples formation ? Who marries whom ? What is the within-household decision process ? To which extent does it depend on wages inequalities between men and women or on their preferences for leisure and consumption ? Do existing education systems provide the same learning opportunities to all children independently of their family background ? Are the school reforms which target disadvantaged children efficient ?

Using the economic approach, this dissertation provides some elements of answers to these questions. To understand the aggregate impact of social behavior on economic outcomes, we need to focus on the individual. The economic approach of individual social behaviors is the rational choice theory which assumes that individuals weigh benefits and costs before taking any decision. The rationality assumption was first restricted to model financial decisions such as investment, saving and consumption. Since Gary Becker, economists have extended it to any individual decisions such as marriage, fertility, education or even crime as soon as individuals make decisions through weighing the advantages and disadvantages of alternative actions. Actions are constrained by income, time, imperfect memory and calculating capacities, and other limited resources, and also by the available opportunities in the economy and elsewhere.

In the first part of this dissertation, I use a micro-economic approach to model marriage decision and within-household allocation of resources. First I propose new methodologies to estimate preferences of individuals in mating and I implement them on American data (PSID) and on English data (BHPS). I then develop a model of the within-household decision process and I show how the decisions made are linked to the marriage market. Labor supplies of men and women are then estimated taking into account the matching process.

In the second part, I focus on the French high-school system and I look at its efficiency to improve performances of students independently of their family backgrounds. Using PISA database in 2000 and 2009, and the panel DEPP of 1995, I look at performance inequalities and at the impact of the grade retention practice and the program of priority education area on school results. Before going through the four chapters of my dissertation, I present the main concepts in family economics and education systems.

³In 2006, the World Bank launched its Gender Action Plan, which was explicitly justified with the effects of female empowerment on economic development.

2 Family Formation and Intra-Household Allocation of Resources

2.1 Family Gain

Despite an increase in single person households over the past few years, the couple is still a dominant organization for individuals: more than 75% of the 25-65 year old individuals live in couples in the OECD countries. The large majority of individuals look for a partner with whom they will live and may have children (OECD, 2011).

There are important gains from living in couple, both economic and emotional. Browning et al. (2014) identify several broad sources of potential material gain from living in couple⁴. First, the sharing of public (non rival) goods such as rent or electricity bills etc. Estimates of this gain are large and amount to 40 % of savings (Browning et al., 2013). Second, life in couple allows division of labor to exploit comparative advantage and increasing returns to scale (Becker, 1973, 1981). One partner could work at home while the other works in the formal labor market. Household work is the source of important production activities that should not be disregarded. Individuals (particularly women) spend a significant fraction of their available time on what economists call “household production” and which comprises immediate tasks such as cleaning, cooking, etc., but also long term investments in health, education and others. Third, marriage allows for extending credit and coordination of investment activities. For example, one partner may work when the other is in school to increase his/her human capital for better future labor earnings. Commitments are crucial for the implementation of such a program (Dufwenberg, 2002). A woman will be hesitant to support her husband through long superior education if she expects him to break the marriage when he graduates. Finally, another advantage of marriage is risk pooling: one partner may work while the other is sick or unemployed. Intuitively one partner can transfer resources to his/her partner in good times and taking resource from him/her in bad times. Then, both partners can be made strictly better off, provided that their incomes are not perfectly correlated (or that risk aversions differ). Hess (2004) finds that couples with a higher correlation in incomes are more likely to divorce, suggesting that effects of mutual insurance on the gains from marriage are higher when the partners’ incomes are less correlated. Shore (2010) finds that the correlation in spouses’ earnings respond to the business cycle: it is higher for couples whose marriage spans longer periods of high economic activity.

Browning et al. (2014) remark that from an economic point of view, the couple is a partnership for the purpose of joint production and joint consumption where consumption and production are broadly defined to include goods and services such as companionship and children. Yet, the gain to live in couples may vary with the characteristics of the partners and it is interesting to look at how couples are formed.

2.2 Homogamy (Assortative Mating)

Are people looking for the most educated or rich individual? Are they looking for people the most similar to themselves? A remarkable fact in couples’ characteristics is the importance of

⁴Most of example of this section are taken from Browning et al. (2014).

social homogamy (OECD, 2011; Vanderschelden, 2006): men and women often partner with people with similar social backgrounds as theirs. Despite the ongoing individualization and liberalization of lifestyle (cohabitation, multiplication of partners,...), we observe a persistence of social attraction and repulsion in partner's choice. A famous survey conducted by the sociologists Bozon and Héran (1991) analyzed mating behavior of French couples. It shows that individuals are not aware of their attraction to socially similar people. Most of the interviewed individuals highlighted that they met their partner by chance. Moreover, agents do not value consciously the social proximity. This criterion comes far away behind health, moral qualities and physical appearance. These authors show that men particularly value physically attractive women. They are also interested in their relational and appearance qualities may be because of the traditional role of women in social representation and social mediation. To a lesser extent, women also value physical appearance and some particular qualities of security and protection such as high height or strength⁵. However women are particularly interested in a high social and professional status. The desired qualities also depend on social background: wealthy women value education and high social status whereas more modest women are more interested in social and professional stability and physical strength. It particularly concerns young women who marry early as marrying an older man allows them to get an earlier access to financial independence.

Bozon and Héran point out two important factors which are mutually reinforcing and which lead naturally to homogamy: preferences and environment. First, individuals prefer partners with whom they share social traits such as religious beliefs, ethnicity and leisure activities. For other traits such as economic success or physical attractiveness, they prefer partners who are more successful or attractive relative to their gender 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. Second, individuals face different opportunities for meeting and mating with other individuals. People have 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 places have been diversified with urbanization and development or leisure, sociability and leisure areas are still socially segmented which increases chances to meet people of the same social background.

Homogamy is a persistent phenomenon which has even increased over the last fifty years. Mare (2008) documents this increase in the US. He first shows that homogamy is a self-reinforcing phenomenon as it is passed down from parents to children through two channels: first, homogamous parents seem to transmit homogamous values to their children with education, second they raise their children in a more homogenous background increasing their chances to meet homogenous partners. Thus, inter-generational transmission may be a cause of the increase in educational resemblance of spouses. Second, Mare highlights the effect of the timing of graduation and of marriage on assortative mating and remarks that both the

⁵A large part of evolutionary literature, inspired by the Darwinist theory predicts that men look for women able to reproduce whereas women look for men able to reproduce and to protect them (Buss et al., 2000).

median age at first marriage and the median age at leaving school have increased over these last fifty years along with the educational homogamy trend. Indeed, there has been a strong increase in the economic return to schooling in the last century ([Autor, 2014](#)) which has raised the graduation age. When the schooling return is important, individuals have greater incentives to pursue their education and to place more weight on the educational level of their prospective marriage partner. This effect has been particularly strong for women who have been getting more and more educated since 1950. [Caucutt et al. \(2002\)](#) show that delay in marriage and fertility is due to increases in the number of years of education of women but it is also due to the marriage matching market and to the return from experience on labor market. First, educated women delay their marriage until they find the best prospect (which leads to a delay in their fertility). Second, women delay fertility because the increase in return from experience gives them incentives to work immediately and increase their future gains. Again, we would expect that low educated level women marry earlier as they have more incentives to marry and get a social status that they could not obtain from their parents or that they could not get by themselves.

Individuals have several incentives to invest in human capital during their youth: they anticipate better labor market earnings but also better marriage prospects due to increase in attractiveness on the marriage market. [Chiappori et al. \(2011\)](#) show that women have a higher education premium than men. Education raises returns from marriage by increasing four factors: marriage probability, partner's education level, surplus of the match, and household resources' share in the match. Indeed, the household is viewed as a two-person organization where each partner has a say.

2.3 Family decision making: a collective choice

What is the decision process of a household which is not a single individual but a group of several members with different preferences and different personal resources? How do they make decisions on production, consumption, labor supply and education of children?

Different assumptions can be made on the level of cooperation among household members ([Browning et al., 2014](#)). Couples can be represented as cooperative (or collective), that is individuals tell each other the truth about their preferences and don't hide any information. In this case, an optimal decision for the couple is made jointly. Conversely, couples can be represented as non-cooperative (or strategic), in which case each member in the household maximizes his/her own utility taking his partner's resources and actions as given. Non-cooperative couples do not necessarily make optimal decisions. One strong support for cooperation within couples is that each partner knows the preferences of the other and can observe their consumption behavior and the two interact often and regularly so that we would expect that they would find ways to exploit any possibilities to find the optimal solution. Nonetheless, the existence of some phenomena such as domestic violence and child abuse, as well as the demand for marriage counseling and family therapy, suggests that family behavior is sometimes inefficient. The support for non-cooperative theory is that partners can't make binding, costlessly enforceable agreements since legal institutions do not provide for external enforcement of contracts regarding

consumption, labor supply and allocation of resources within marriage. Besides, efficiency is difficult to maintain overtime. If household members cannot commit through a permanent contract, changes in opportunities offered to each member will affect their bargaining power and will lead to inefficiency (Mazzocco and Yamaguchi, 2007). Experimental economics show that non-cooperative behaviors of couples can indeed appear. For instance Cochard et al. (2009) observe that women in couple are more opportunistic when they are married and have children. Del Boca and Flinn (2012) estimate a model that allows for cooperative and non-cooperative decision making in the household. Based on PSID data, they found that about one-fourth of American couples behave in a non-cooperative way.

Another element to consider in the within-household decision process is the bargaining power of each individual. It has been repeatedly proved that the redistribution of resources within the household depends on the relative bargaining strength of each spouse (Chiappori, 1988, 1992). Bargaining powers may depend on the relative income of each partner but may also depend on external factors. Following the intuition of Becker, these factors could reflect the situation in the marriage market which could affect opportunities of spouses outside marriage and so the intra-household balance. Chiappori et al. (2002) test this assumption and show that the sex ratio (the relative supplies of males and females in the marriage market) and divorce legislation influence the bargaining power of individuals. When there is a relative scarcity of women, the distribution of gains from marriage is shifted in their favor. The same effect would be obtained if divorce legislation was made more favorable to women. I show in the first two chapters of this dissertation that physical attractiveness and family values of individuals can also influence the distribution of resources.

To understand how resources are shared within the household, we need to uncover the decision process between the two members. This is an important issue as the redistribution of powers toward women could increase expenditure on children, and consequently increase their educational achievement and ultimately the global growth. However, in addition to the family background, the other main factor of children human capital accumulation is the existing education system.

3 Evaluation of Education Systems

It is very important that governments invest in education and build efficient education systems for several reasons. First, a general increase in human capital leads to productivity growth and innovation. Second, subsidized education is an important means along with taxation, towards redistributing resources within the society and towards correcting for inequalities created through families transmissions. All OECD countries provide a general comprehensive school system, and education of children is free and compulsory in most of them for the 5-15 years old (OECD, 2011). Since 2000, the PISA evaluations allow researchers to compare the skills of 15 year-old students of different OECD countries in different subjects such as reading, math and science. These evaluations make possible the comparison of different education systems to assess their efficiency. One remarkable result is that over all OECD countries, educational attainment

is strongly linked to family background: the "social gradient", that is the difference between the expected performance of two individuals separated by a given amount of socio-economic status, is significant for all OECD countries. It is below the average of OECD countries for Canada, Sweden and Finland, slightly above the average in France and substantially above the average in the UK ([OECD, 2004](#)). Education systems seem to fail in providing the same learning opportunities to all children.

Some public policies have been implemented to help disadvantaged students such as vouchers to help poor students to attend better schools located in a wealthier neighborhood. Other programs give additional resources to schools located in particularly poor areas to incitate them to form smaller class sizes and provide more teaching hours. The use of the grade retention practice (which allows the student to repeat a year of schooling if he had not acquired enough knowledge during the current year) tries also to give more chances to students with difficulties. Other public policies may have helped to reduce the social gradient such as pre-school provision for 3 year-old children or a late sorting of students between general and vocational tracks.

All these reforms are costly and we need to know if they are efficient. However, evaluation of education practice is difficult as the individual latent ability is unobserved. The identification of the effect of an education reform requires either a natural experiment or good data and advanced econometric tools. Public policy evaluation is a growing body of Economics and econometricians have been developing powerful identification strategy to evaluate treatment effects ([Heckman and Vytlacil, 2005](#)). I present in the last chapter of this dissertation a structural evaluation of the grade retention practice.

Bibliography

- D. Acemoglu and J-S. Pischke. Changes in the wage structure, family income, and children's education. *European Economic Review*, 45(4):890–904, 2001.
- P. Aghion and P. Howitt. A model of growth through creative destruction. *Econometrica*, 60(2): 323–51, 1992.
- D.H. Autor. Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science*, 344(6186):843–851, 2014.
- G.S. Becker. A theory of marriage: Part I. *Journal of Political economy*, 81(4), 1973.
- G.S. Becker. A treatise on the family, 1991. *Enlarged Edition (Cambridge, Mass.: Harvard University Press)*, 1981.
- G.S. Becker. Nobel lecture: The economic way of looking at behavior. *Journal of political economy*, pages 385–409, 1993.
- M. Bozon and F. Héran. La formation du couple. *Population*, 1:152–159, 1991.
- M. Browning, P-A. Chiappori, and A. Lewbel. Estimating consumption economies of scale, adult equivalence scales, and household bargaining power. *The Review of Economic Studies*, 80(4): 1267–1303, 2013.
- M. Browning, P-A. Chiappori, and Y. Weiss. *Economics of the Family*. Cambridge University Press, 2014.
- D.M. Buss, T.K. Shackelford, and G.J. LeBlanc. Number of children desired and preferred spousal age difference: context-specific mate preference patterns across 37 cultures. *Evolution and Human Behavior*, 21(5):323–331, 2000.
- P. Carneiro and J.J. Heckman. Human capital policy. 2003.
- E.M. Caucutt, N. Guner, and J. Knowles. Why Do Women Wait? Matching, Wage Inequality, and the Incentives for Fertility Delay. *Review of Economic Dynamics*, 5(4):815–855, 2002. ISSN 1094-2025.
- P-A. Chiappori. Rational household labor supply. *Econometrica*, 56(1):63–90, 1988.
- P-A. Chiappori. Collective labor supply and welfare. *Journal of Political Economy*, 100(3): 437–467, 1992.
- P-A. Chiappori, B. Fortin, and G. Lacroix. Marriage market, divorce legislation, and household labor supply. *Journal of Political Economy*, 110(1):37–72, 2002.
- P-A. Chiappori, B. Salanié, and Y. Weiss. Partner choice and the marital college premium. *Discussion Papers*, 2011.
- F. Cochard, H. Couprie, and A. Hopfensitz. Do spouses cooperate? and if not: Why? *TSE Working Paper*, 9, 2009.

- M. Corak. Income inequality, equality of opportunity, and intergenerational mobility. *The Journal of Economic Perspectives*, pages 79–102, 2013.
- F. Cunha, J.J. Heckman, and S.M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.
- D. Del Boca and C. Flinn. Endogenous household interaction. *Journal of Econometrics*, 166(1):49–65, 2012.
- M. Doepke, M. Tertilt, and A. Voena. The economics and politics of women’s rights. *Annual Review of Economics*, 4:339–372, 2012.
- M. Dufwenberg. Marital investments, time consistency and emotions. *Journal of Economic Behavior & Organization*, 48(1):57–69, 2002.
- J. Greenwood, N. Guner, G. Kocharkov, and C. Santos. Marry your like: Assortative mating and income inequality. Working Paper 19829, NBER, January 2014. URL <http://www.nber.org/papers/w19829>.
- J.J. Heckman and E. Vytlacil. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–738, 2005.
- J.J. Heckman, J. Stixrud, and S. Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482, 2006.
- G.D. Hess. Marriage and consumption insurance: what’s love got to do with it? *Journal of Political Economy*, 112(2):290–318, 2004.
- S.J. Lundberg, R.A. Pollak, and T.J. Wales. Do husbands and wives pool their resources? evidence from the united kingdom child benefit. *Journal of Human Resources*, 32(3), 1997.
- R.D. Mare. Educational assortative mating in two generations. *Department of Sociology, University Of California Los Angeles*, 2008.
- M. Mazzocco and S. Yamaguchi. Labor Supply, Wealth Dynamics, and Marriage Decisions. *University of Wisconsin Economics Department Working paper*, 2007.
- OECD. Pisa 2003 key findings. what makes a school system perform? seeing school system through the prism of pisa. 2004.
- OECD. Doing better for families. 2011.
- T. Piketty. The impact of divorce on school performance: Evidence from france, 1968-2002. 2003.
- P.M. Romer. Endogenous technological change. *The Journal of Political Economy*, 98(5), 1990.
- S.H. Shore. For better, for worse: Intrahousehold risk-sharing over the business cycle. *The Review of Economics and Statistics*, 92(3):536–548, 2010.
- D. Thomas. The distribution of income and expenditure within the household. *Annales d’Economie et de Statistique*, pages 109–135, 1993.

M. Vanderschelden. «Position sociale et choix du conjoint: des différences marquées entre hommes et femmes». *Données Sociales, Insee*, 2006.

Part II

Marriage Market and Intra-Household Allocation

Chapter 1

Marriage gains : who should you marry ?

Abstract

This paper proposes a general equilibrium approach to examine the marriage market and within household transfers. Using observations on who matches with whom, I derive information on how efficiently the market functions and on agent's preferences in terms of mate choice. I use a multidimensional search and matching model which allows people to choose their partner according to their education level, their wage and their physical attractiveness. I first estimate a gender dependent unidimensional marriage index then I implement a multidimensional matching model to allow for complementarities in characteristics of partners. In both specifications, I find that wages are more important in men's attractiveness than in women's attractiveness whereas physical characteristics and education are more important for women. Some disparities exist across age groups. The weight of wage increases over the lifetime whereas the weights of education and physical attractiveness decrease. Besides, I find that the marriage market is more efficient for young individuals who are more selective. Finally, I compute the marriage gain and the within-household transfers and show that men get in average 56% of the surplus. I also look at the effects of taxation on matching patterns. I compare individual and joint taxation of the household gain and show that joint taxation slightly improves women's share of the surplus through less sorting on characteristics on the marriage market.

1 Introduction

How individuals sort themselves into marriage has important implications for the distributions of income, labor supply, and fertility (Becker, 1973). First, positive assortative mating in the marriage market contributes to income inequality across households (Greenwood et al., 2014). Second, within household inequalities also depend on the state of the marriage market. It is now well known that the household members don't share their income equally but bargain over it. Factors conditioning the process of match formation such as sex ratios or rules about divorces implicitly determine the outside option of each household member and influence the sharing rule (Chiappori et al., 2002).

The simple analysis of spousal correlation for different characteristics is insufficient to understand agent's preferences in mate choice. Observable association of spousal traits also results from the distribution of characteristics in the population, from the frequency of contacts among singles and from the frequency of couples separations. A structural investigation is needed to better understand who matches with whom.

This paper proposes a multidimensional search and matching model of marriage. It is a search model where agents are ex-ante heterogeneous, with each represented by a marriage index as in Wong (2003). When matched, the individuals produce together a good they can consume. This good may come from economies of scale (sharing the rent, the electricity, increasing returns to scale in cleaning, meal preparation)¹. I assume that the production of this good depends on the observable characteristics of the partners and on an unobservable match specific component. This marriage gain is not specified a priori but is estimated non-parametrically using the estimation strategy of Jacquemet and Robin (2013). Single individuals meet randomly and decide whether they marry. They evaluate the match anticipating what will be the surplus generated by the match and how it will be split. Individuals bargain *a la Nash*² to choose an optimal sharing rule. Either the surplus is high enough and both want to match, or it is not and both prefer to stay single. Identification of matching patterns is obtained with the steady-state assumption of search models as in Shimer and Smith (2000). Assuming that the market is at the equilibrium in the data, matching patterns are recovered from observed joint distributions of characteristics among couples and among singles³.

This paper adds two important contributions to the existing search models on the marriage market. The first one is the endogenisation of the separation. When people are matched, some shocks can hit the unobserved characteristics and lower the value of the match. In that case, the match breaks up. Only couples with high enough complementarities in observed characteristics will last. This new setting allows me to identify meetings and separations frequencies. Using the panel structure of the PSID, I observe partnership duration and singlehood duration and I

¹Many studies have attempted to estimate the additional revenue generated by living in couples (Browning et al., 2013; Couprie, 2007)

²Nash-Bargaining within household is a particular case of collective models: it gives Pareto efficient outcomes. It is a simple way to include the outside option in the bargaining process. See Manser and Brown (1980), McElroy and Horney (1981) and McElroy (1990) for deeper discussion on Nash-Bargaining.

³Similar strategies are used in models with perfect information as in Choo and Siow (2006), Chiappori et al. (2011b) and Galichon and Salanié (2011)

can identify meeting parameters and quality shock parameters.

The second contribution of this paper is to allow people to match on several continuous characteristics. Heterogeneity of individuals varies in an infinite number of dimensions which can be important in couples formation. One of the most important observable dimension of heterogeneity on the marriage market is education or the social group (Mare, 2008; Bozon and Hérin, 1991). Wages and physical attractiveness are other important dimensions.

First, I build a marriageability index composed of wages, education and physical characteristics. Wong (2003) and Chiappori et al. (2012a) also build such indexes and shed light on some complementarities of characteristics. Chiappori et al. (2012a) show that a substitution exists between physical characteristics and productivity characteristics for men whereas it doesn't exist for women. A poor handsome man can be as attractive as a rich non good looking man whereas a woman can't compensate a disadvantage look with higher education. Studies of Chiappori et al. (2011a) and Wong (2003) show there exist differences between races. Afro-Americans don't have the same preferences than white Americans. However, Chiappori et al. (2012a) do not consider a frictional general equilibrium setting and identify their index on existing couples whereas Wong (2003) identifies her index by specifying the match surplus as the product of partner's indexes. In this paper, the surplus is estimated non-parametrically and can be any functional of the indexes of the partners. I allow the index to be gender dependent and estimate it on different age groups. Second, I extend the setting to true multidimensional matching where an agent type is a vector of several characteristics and not a unidimensional index. Only a few papers consider multidimensional matching. Some papers consider matching on two characteristics, one of which is a binary characteristic (Chiappori et al., 2010), but only the recent work of Dupuy and Galichon uncovers matching preferences over many different continuous characteristics. They recovered traits or characteristics that are very attractive in marriage (Dupuy and Galichon). To my knowledge, this paper is the first which consider two-sided multidimensional matching in a search setting.

The model provides several results. First, I find that wages account for 30% of men's attractiveness and 20% of women's attractiveness whereas physical characteristics account for 30% for men and 35% for women⁴. Education accounts then for 40% for men and 45% for women. Some disparities exist across age groups. The weight of wage increases over the lifetime whereas the weights of education and physical attractiveness decrease. Besides, I recover the positive assortative mating result. Without imposing any a priori assumptions on the marriage gain, I show that it is supermodular in indexes. People want to marry with the highest index partner which leads at the equilibrium to a high correlation of indexes within couples. Second, the multidimensional matching model presents the form of the marriage gain with respect to characteristics. Men seem to prefer very thin women whereas women prefer heavy men. The more far away men and women are from 'ideal' physical norms, the less they will select their partner on that criterion. Considering wages, high-wage men and women want to match with high-wage partners whereas low-wage women and low-wage men seem to care much less. Third, I look at the efficiency of marriage markets. Selectivity in mating increases

⁴Chiappori et al. (2010) and Dupuy and Galichon also show that physical attractiveness was an important dimension for women attractiveness.

with the rate of contacts among singles and decreases with the rate of couple separations. Efficiency is then a function of the ratio of these two rates. Even if these both rates are higher for the young, I find that the marriage market of the young is more efficient than the one of the oldest individuals. Young individuals (below 35 years old) are more selective. Fourth, I identify the within-household transfers and show that men get around 56% of the marriage gain.

Finally, I implement a taxation setting of the household good. For redistributive motives (vertically across households or horizontally between singles and couples), a social planner may want to tax or subsidize the good produced by couples. However, taxation can impact matching pattern. As welfare reforms have an impact on the sharing rule ([Chiappori et al., 2002](#)), family policies may influence marital sorting through divorce and couples formation ([Francesconi et al., 2009](#); [Bitler et al., 2004](#)). I compare matching equilibria obtained with individual and joint taxation of the household and I show that joint taxation slightly improves women's share of the surplus through less sorting on characteristics on the marriage market.

The model is described in section 2 and the data in section 3. Section 4 presents the estimation strategy and section 5 the results. Equilibrium conditions and simulations are computed in section 8. Section 6 concludes.

2 Model

2.1 The marriage market

I consider a marriage market with L_m males and L_f females that are likely to match. In this paper, a match is a two-person household (married or cohabiting). The number of married couples is denoted by N and the respective numbers of single males and single females are $U_m = L_m - N$ and $U_f = L_f - N$. I assume that only singles search for a partner ruling out “on-the-marriage search”. I denote λ the instantaneous probability of a meeting between a random single woman and a random single man. Then $\lambda_m = \lambda U_f$ and $\lambda_f = \lambda U_m$ are the respective instantaneous probabilities of an agent among the population m and f of meeting a new person of the populations f and m .

2.2 Individual types

Individuals differ in a fixed set i of K exogenous characteristics that contains their gender and other continuous characteristics. In the case of couples, I will use $i \in \mathbb{R}^K$ for husband's type and $j \in \mathbb{R}^K$ for wife's type. Let $U(i)$ and $L(i)$ denote the distributions of i in the population of singles and in the whole population, respectively. Let also $U_m(i)$ and $U_f(j)$ denote the distributions of the types for male and female singles separately. I similarly define distribution functions $L_m(i)$ and $L_f(j)$ for the whole populations of men and women. Lastly, let $N(i, j)$ be the distribution of matches with a male of type i and a female of type j . I also use lower-case notations, such as $u(i)$ or $u_m(i)$ or $n(i, j)$ to denote corresponding densities with respect to some adequate measure that is not necessarily the Lebesgue measure, but that we still simply denote as di or dj . Hence,

$l_m(i) = \int n(i, j) dj + u_m(i)$, meaning that the number of males of type i in the economy is equal to the number of all married males and all single males of this type.

2.3 Marriage contracts

When they are single, individuals get zero instantaneous utility. However, when a match is formed, the two members get a positive utility from the consumption of a good Q that they produce together. I assume as in the Jacquemet-Robin model, that the good created depends on the observable characteristics of the partners and on a match specific component z : $Q_{ij}(z) = C(i, j) + z$. The match specific component z is drawn from some distribution G at the time of the first meeting and is infrequently updated with new draws z' from the same distribution G at random times following a Poisson process with parameter λ_z . The gain is then shared between the two members of the couple, the man gets $t_{i|j}(z)$ and the woman gets $t_{j|i}(z)$ such that $t_{i|j}(z) + t_{j|i}(z) = Q_{ij}(z)$. When a new z is drawn, the value of being in couple is modified, and the partners bargain again over the new surplus. The couple can separate if they do not agree on a new sharing rule. Then the present value of a man of type i married with a woman of type j and a match specific component z is :

$$(r + \lambda_z)W_{i|j}(z) = t_{i|j}(z) + \lambda_z \int_{z'} \max(W_{i|j}(z'), V_i^0) dG(z'), \quad (1)$$

where r is the discount rate and the second term of the right-hand side is the option value of divorce or match continuation after a shock to the match specific component. The present value of a single man of type i , V_i^0 is

$$rV_i^0 = \lambda \iint \max(W_{i|j}(z) - V_i^0, 0) U_f(j) dj dG(z). \quad (2)$$

that is then the sum of the instantaneous utility of being single (which is nil) plus his expected surplus from a match.

2.4 Surplus, Nash bargaining and transfers

When a match forms, the two members start to bargain on the repartition of the gain $Q_{ij}(z)$ that has been created by the match. I model the decision process with a Nash bargaining where the threat point is to stay single. The respective threat points are then the respective outside options of the man and the woman and so their present value as singles. I denote β and $(1 - \beta)$ the respective bargaining power of the man and the woman. Then the Nash bargaining of the household is the maximization of the following program :

$$\begin{aligned} \max_{t_i, t_j} & (W_{i|j}(z) - V_i^0)^\beta (W_{j|i}(z) - V_j^0)^{1-\beta} \\ \text{s.t.} & t_i + t_j = Q_{ij}(z) = C(i, j) + z. \end{aligned}$$

$W_{i|j}(z) - V_i^0$ is the surplus of the man and $W_{j|i}(z) - V_j^0$ is the surplus of the woman. This model has the important property of transferable utility models: both surplus are simultaneously

positive or negative. Either the match generates enough surplus to make both people want to marry or it doesn't and they both refuse to marry. The maximization of this program with respect to t_i leads to the following equality :

$$\frac{W_{i|j}(z) - V_i^0}{\beta} = \frac{W_{j|i}(z) - V_j^0}{1 - \beta}.$$

Denoting total surplus $S(i, j, z)$ such that $S(i, j, z) = (W_{i|j}(z) - V_i^0) + (W_{j|i}(z) - V_j^0)$, we have

$$\begin{aligned} W_{i|j}(z) - V_i^0 &= \beta S(i, j, z) \\ W_{j|i}(z) - V_j^0 &= (1 - \beta) S(i, j, z). \end{aligned} \tag{3}$$

The two individuals i and j decide to marry if and only if they both obtain a positive surplus from the match, that is if $S(i, j, z) > 0$. Then using equation 2, we can derive the following expressions for the total surplus

$$(r + \lambda_z)S(i, j, z) = C(i, j) + z - rV_i^0 - rV_j^0 + \lambda_z \int_{z'} \max(S(i, j, z'), 0) dG(z'),$$

and the transfers

$$\begin{aligned} t_{i|j}(z) &= rV_i^0 + \beta \left[(r + \lambda_z)S(i, j, z) - \lambda_z \int_{z'} \max(S(i, j, z'), 0) dG(z') \right] \\ t_{j|i}(z) &= rV_j^0 + (1 - \beta) \left[(r + \lambda_z)S(i, j, z) - \lambda_z \int_{z'} \max(S(i, j, z'), 0) dG(z') \right]. \end{aligned} \tag{4}$$

The total surplus is the difference between the total expected gain of the match and the individual single present value. The remarkable fact of this expression is the linearity of the surplus in z . We will use the notation $s(i, j)$ such that

$$(r + \lambda_z)S(i, j, z) = s(i, j) + z.$$

Linearity in z allows us to link the total surplus of the couple to the match probability. When two individuals of each population meet, they decide to match if and only if the surplus is positive. Then the matching probability between a man of type i and a woman of type j when they meet can be computed with

$$\begin{aligned} a(i, j) &= \mathbb{P} \{s(i, j) + z > 0 | i, j\} \\ &= 1 - \mathbb{P} \{z \leq -s(i, j) | i, j\} \end{aligned} \tag{5}$$

2.5 Equilibrium

The characterization of the equilibrium allows us to close the model. To solve for a market equilibrium, I suppose there is no entry of new singles⁵, but the partnerships of type (i, j) are destroyed at rate $\lambda_z(1 - a(i, j))$ whereupon both return to the single market. At the equilibrium, there is equality between inflows and outflows for each type of marriage. Then we have for all couples of type (i, j) , the equality between the number of outflows and the number of inflows:

$$\lambda_z(1 - a(i, j))N(i, j) = \lambda U_m(i)U_f(j)a(i, j), \quad (6)$$

The match probability $a(i, j)$ can be then computed using data after having estimated the parameters λ and λ_z . Using the equation (5) and choosing a specification for $G(z)$ the distribution function of z , I can derive the marriage surplus through the following simple relation

$$s(i, j) = -G^{-1}(1 - a(i, j)).$$

The whole model is then identified and can be estimated on data.

3 Data

3.1 The sample

I estimate the model using the PSID survey where I follow individual marriage history from 1977 to 2011 using the family and individual samples. I merge the individual file with the marriage history file. I only keep households composed of heterosexual couples and single member households who are between 22 and 64 and present in 1999⁶. I drop all observations with missing information on their wage, their height, their weight and their education level. When married, both spouses have to work and declare their wage. I trim the 1% top and bottom of all these variables. The wage is the hourly wage rate in dollars, the education level is the number of year of education, and the BMI is the weight (in kilograms) divided by the square of the height (in meters). The maximal length of observation is 34 years from 1977 to 2011. The final sample is composed of more than 8000 individuals that I follow at least 2 years between 1999 and 2011. I observe about 1500 couple separations and 700 couple formations⁷.

3.2 Data description

Table 1 presents some descriptive statistics and shows that men and women in couple are in average more educated and have higher wages than singles. However, whereas married men are fatter than single men, married women are in average thinner than single women.

⁵Burdett and Coles (1999) review the different cases that have been considered in the literature considering how new singles enter the market overtime.

⁶The anthropometric information is not available before 1999.

⁷In appendix, table 8 presents sample size, singles' rate and sex ratio.

Table 1: Summary statistics

	Education	Wage rate	BMI	Age
Married men	13.26 (2.54)	19.59 (13.00)	27.34 (4.00)	41.81 (9.88)
Married women	13.27 (2.42)	13.97 (8.85)	25.26 (5.03)	39.71 (9.50)
Single men	13.03 (2.50)	14.67 (9.60)	25.06 (5.13)	38.35 (10.59)
Single women	12.71 (2.30)	12.20 (8.25)	27.31 (5.90)	39.22 (10.92)

Standard deviations are in brackets

Table 2 presents correlations of certain characteristics within couples. Age correlation is very high, around 0.9 for the whole sample. Spouse education correlation is also very high (around 0.6) whereas wage correlation and BMI correlation are also important but much lower. Education correlation is constant across age groups whereas wage correlation decreases with age (this may be due to a specialization effect within couples. Women spend more time at home raising children at the expense of their career and their wage.) and BMI correlation increases with age (this may be due to the fact that they live together and have the same habits (Averett et al., 2008)).

Table 2: Spousal Correlations

Age range	Education	Wage	Bmi	Age	Number of couples
22-35	0.59	0.35	0.13	0.75	947
All	0.56	0.28	0.20	0.90	3480

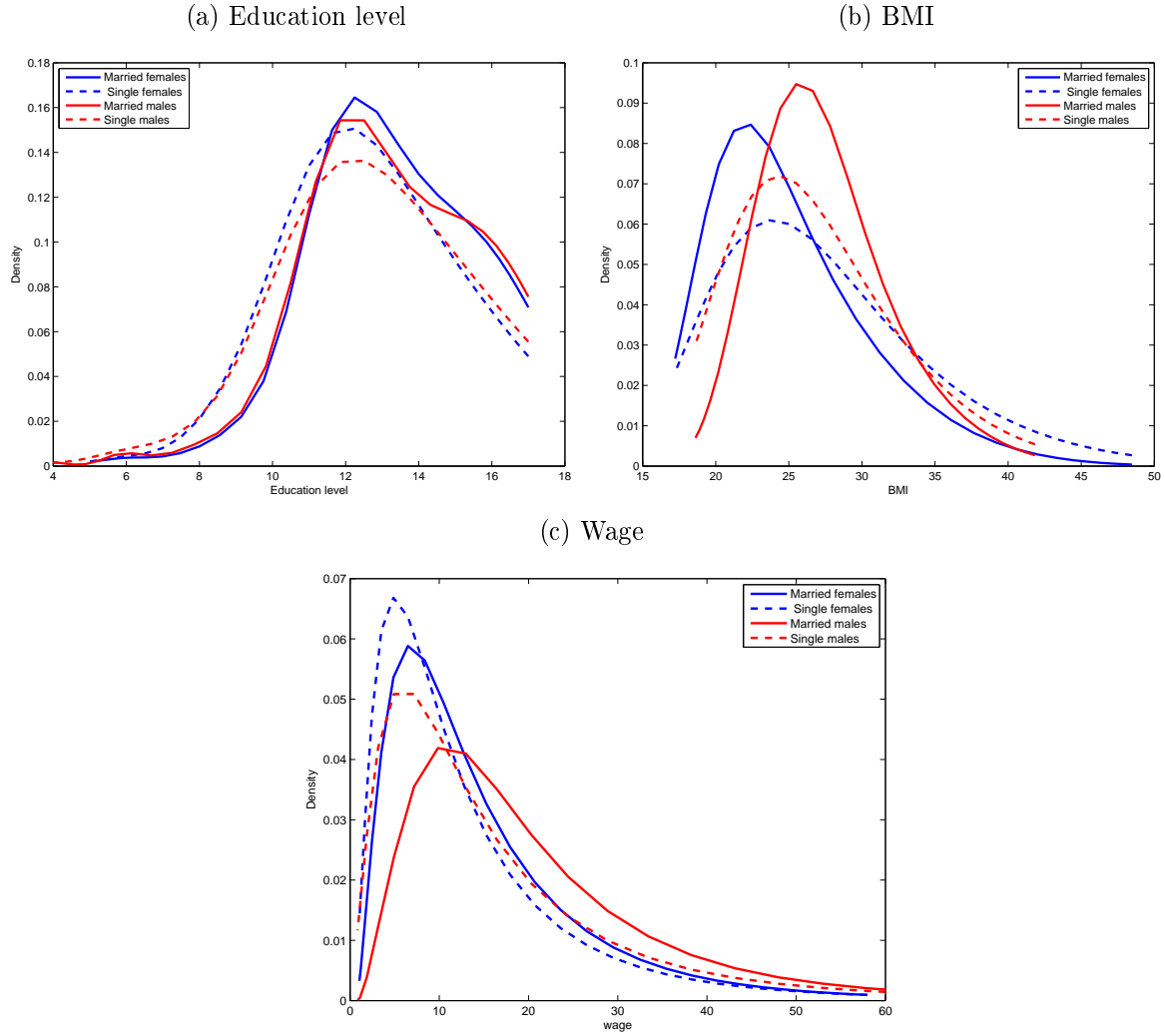
Table 3 presents the complete duration spells of singlehood and marriage we observe in data. We observe 45% of complete durations of men singlehood and 33% of women singlehood. We observe only 16% of complete marriage durations which last a little bit longer (17 years in average whereas men singlehood last around 13 years and women singlehood last 14 years).

Table 3: Complete Duration Spells

	Number	Share	Average	Std. Dev.
Men singlehood	296	45.2 %	12.89	7.82
Single women	367	32.8 %	14.13	7.79
Marriage duration	571	16.4 %	17.30	8.26

Figure 2 represents the distribution of wages, education and BMI conditional on marital status. We notice that what was observed for the means on table 1 holds for the whole distributions. Married people are more educated and have higher wages. Married women are the thinnest whereas married men are the fattest.

Figure 1: Densities by sex and marital status. PSID 1999



4 Estimation strategy

4.1 Agent's type

An index of several characteristics

I construct a marriageability index as a combination of different characteristics as in [Wong \(2003\)](#). I consider three different dimensions. Two account for the socioeconomic status of the individual and are the education level (denoted e) and the wage rate (denoted w) and one accounts for the physical attractiveness and is the opposite of the Body Mass Index (BMI) (denoted $b = -BMI$)⁸. For each gender, I standardize these three variables to adjust their mean to 15 and their variance to 3. These three standardized variables (still noted w , e , and b for convenience) will then be combined in a marriageability index. I allow this index to be gender dependent with different vector of weights for each variable : α_m is the vector of weights for men and α_f is the vector of weights for women. The indexes have the following expression :

⁸[Chiappori et al. \(2012a\)](#) also use the BMI as a proxy for physical attractiveness and found that physical attractiveness decreases unambiguously with the BMI.

$$I_m = \alpha_{m_1}w + \alpha_{m_2}e + (1 - \alpha_{m_1} - \alpha_{m_2})b$$

$$I_f = \alpha_{f_1}w + \alpha_{f_2}e + (1 - \alpha_{f_1} - \alpha_{f_2})b.$$

A truly multidimensional type

If we want to relax the assumption that agents are ranked the same way by all men and by all women and find some complementarities in characteristics of the partners, we have to consider truly multidimensional types. The type is then a vector of several characteristics. To keep the model tractable and easily estimable, I will only consider a bi-dimensional type using the wage and the BMI. Then the type i of an agent would be $i = (w, b)$.

4.2 Estimation by an exponential duration model

As in Wong (2003), I use maximum likelihood estimation to identify the matching parameters and the weights of the indexes. Identification of all parameters requires knowledge of couples' characteristics and individuals' marriage history. Agents can be single or married at the first interview in 1999. Information on the duration of singlehood or marriage is obtained by following single agents after the first interview and using information on previous marital history available in the survey. When singles match, I observe the characteristic of their new partner and I follow the subsequent duration of the match.

I denote T_{0b} (T_{0f}) the elapsed (residual) duration of singlehood for the single people at the time of the interview. Therefore, the duration of singlehood is $T_0 = T_{0b} + T_{0f}$. I denote C_{0b} and C_{0f} the corresponding censoring indicators.

Let's consider a man of type i who is single at first interview. T_{0b} and T_{0f} are supposed to be i.i.d. and have an exponential distribution with parameter $\lambda \int_j a(i, j)U_f(j)dj$, that is the probability to find a woman with whom the match will be formed. Then for an agent of type i , the individual contribution of singlehood duration until and including the time of exit into marriage or censoring is

$$L_i^{0a} = \left(\lambda \int_j a(i, j)U_f(j)dj \right)^{1-C_{0b}+1-C_{0f}} e^{-\left(\lambda \int_j a(i, j)U_f(j)dj \right)(T_{0b}+T_{0f})},$$

where $T_{0b} > 0$ and $T_{0f} > 0$. Events occurring after exit from being singlehood are independent of the events up to exit. Therefore, their probability is independent of the likelihood of being singlehood. The two events immediately following type i 's singlehood duration are first the realization of whom to match with and second the subsequent duration of the match. The first is given by the density of accepted type and the second by the density of the duration with or without censoring. I denote $T_{01} > 0$ the residual duration of marriage for the single man who meets and matches with a woman. If this duration is right censored, I let $C_{01f} = 1$. The subsequent contribution to the loglikelihood of the single man of type i who matches with a woman of type j for a certain period is then

$$L_{ij}^{0b} = u_f(j)(\lambda_z(1 - a(i, j)))^{1-C_{01f}} e^{-\lambda_z(1-a(i,j))T_{01}},$$

Then the contribution to the likelihood of this single man of type i is the product of the two previous likelihood :

$$L_{ij}^0 = L_i^{0a}(L_{ij}^{0b})^{1-C_{0f}}.$$

I build a similar loglikelihood for each woman who single at the time of the first interview.

Let $(T_{1b}, T_{1f}, C_{1f}, C_{1f})$ be similar notations for marriage durations. The contribution to the loglikelihood of a man of type i married with a woman of type j who separate after a certain period is then

$$L_{ij}^1 = (\lambda_z(1 - a(i, j)))^{1-C_{1b}+1-C_{1f}} e^{-\lambda_z(1-a(i,j))(T_{1b}+T_{1f})}.$$

The loglikelihood of the model is then equal to the product of the contribution of each couple and each single individual.

To build the likelihood, two operations must be performed first. I need to build the indexes i and j using vectors of weights α_m and α_f then I compute the match probability function using the parameters λ_z , λ and the steady-state flow restriction 6.

$$a(i, j) = \frac{N(i, j)}{\frac{\lambda}{\lambda_z} U_m(i) U_f(j) + N(i, j)}.$$

The structural parameters to be estimated are $(\alpha_m, \alpha_f, \lambda, \lambda_z)$.

5 Estimation results

5.1 Parameter estimation

Table 4 presents the estimates of the parameters for the unidimensional index model. Wages account for 30% of men's attractiveness and 20% of women's whereas education accounts for 40% for men and 45% for women. Then physical attractiveness accounts for 30% for men and 35% for women. I then estimate the model on a subgroup composed of the youngest individuals. As men are generally married with women younger by one year or two, I select 22-35 year old women and 23-36 year old men to maximize the number of couples falling in the age group. Some disparities exist across age groups. The weight of wages increases over the lifetime whereas the weights of education and physical attractiveness decrease. For each age group, the weight of wage is more important in men attractiveness than in women attractiveness. I also estimate the models with constraints. I estimate two two-component models: one using wage and education and the other using wage and the BMI. I also estimate three one-component models on wage, education and BMI. Results are presented in appendix on tables 9, 10, 11, 12 and 13. Loglikelihood tests reject these constrained models. Each variable is important in one's attractiveness.

Table 4 and table 5 present estimates of the meeting and separation parameters λ and λ_z for the index model and for the multidimensional model. λ and λ_z are structural parameters which reflect the marriage market efficiency. The marriage market is more efficient with a high λ (high rate of contact) and a low λ_z (higher length of marriage). A measure of the market efficiency could be the ratio $\frac{\lambda}{r+\lambda_z}$ which measures the relative speed of offers and can be a measure of the inverse of search frictions. The higher this ratio, the better the marriage market works. A high λ means that individuals will be selective as they will find quickly their best match. A low λ_z also incites individuals to be selective as the match won't break up quickly. For both the multidimensional model and the three-components index model, we observe a decrease in λ_z over age which reflects the fact that couples last longer among older individuals. We also observe for both models a decrease of λ over age which shows that the rate of contacts decreases with age. The market is then more flexible for young individuals. The resulting ratio $\frac{\lambda}{r+\lambda_z}$ is also decreasing with age (with r around 5%) meaning the marriage market is more efficient for young individuals. They can be more selective than older individuals. The parameters of quality shocks and meeting give an average duration of singlehood of 7.6 years for single men and 14.5 years for single women and a average duration of couples of 38.5 years.

Table 4: Estimates from the three-component index

Age range*	α_{m1}	α_{m2}	α_{f1}	α_{f2}	λ	λ_z	$\log L$
22-35	0.10 (0.01)	0.50 (0.01)	0.00 (0.94)	0.66 (0.03)	0.0078 (0.0010)	0.0580 (0.0001)	-9758
All	0.30 (0.16)	0.38 (0.16)	0.21 (0.09)	0.45 (0.33)	0.0035 (0.0077)	0.0270 (0.0040)	-25963

* Age range of women (age range of men is shifted by 1 additional year and is 23-36).

Table 5: Estimates from the bi-dimensional model (no index)

Age range*	λ	λ_z
22-35	0.0043 (0.0014)	0.0541 (0.0017)
All	0.00291 (0.0005)	0.0251 (0.0005)

* Age range of women (age range of men is shifted by 1 additional year).

5.2 Match probability, match gain and transfers

Three-component index model

Using the estimated weights, I compute the marriageability index for each individual and I represent on figure 2 its distribution in the population by sex and marital status (left) and its joint distribution among couples (right). We notice that men have higher index than women and that married men and women have higher indexes than singles. Finally, we can remark that the joint density of indexes is higher on the diagonal. Index correlation within couples is high.

Figure 2: Index density distributions

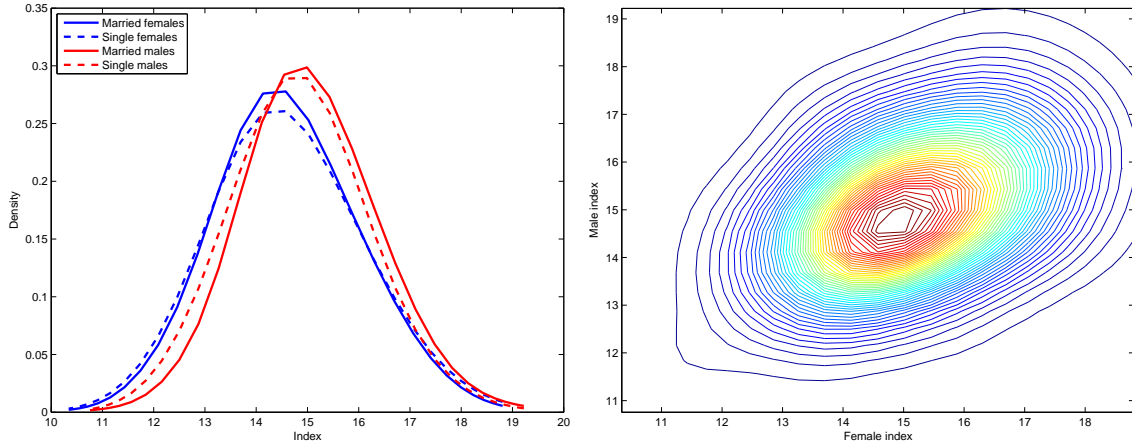


Figure 3 represents the marriage probability between individuals when they meet. The left panel is a 3D plot whereas the right panel represents the level curves. The woman index is on the x-axis and the man index is on the y-axis. We observe a strong positive assortative matching in the index for high indexes (>15). The higher the index⁹ the higher the probability to match with a high index partner. It is less clear for low indexes. The probability that a man with an index of 18 matches with a woman of index 18 when he meets her is 0.16 whereas the probability that he marries a woman of index 12 when he meets her is less than 0.02.

Figure 3: Marriage probability

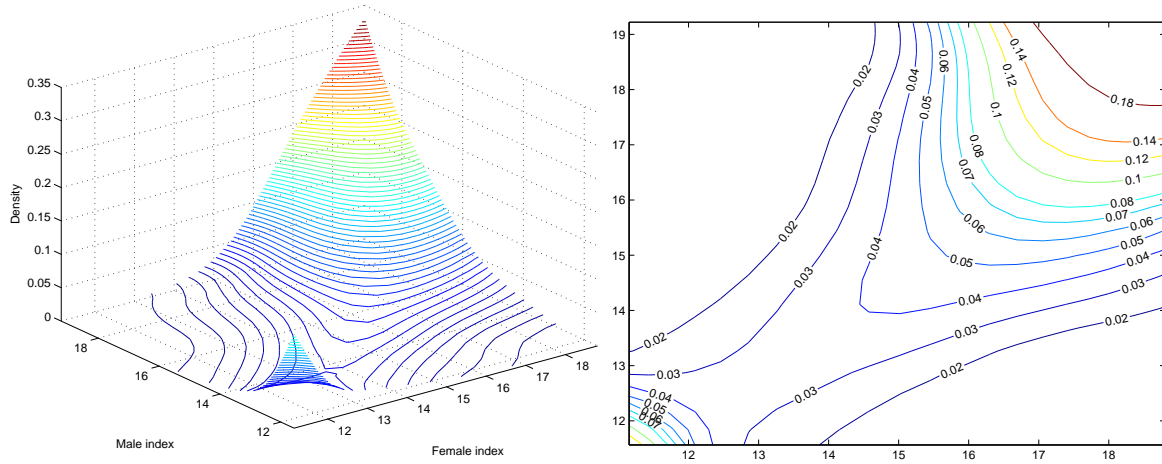


Figure 4 represents the marriage gain between two matched individuals¹⁰. We see that the gain is strictly supermodular with respect to the indexes for high indexes. For low indexes, it is also less clear.

⁹Remember that the index increases with wage, education and decreases with BMI.

¹⁰I compute $C(i, j)^{z > -s_{ij}}$ which can be negative as what only matters is $C(i, j) + z$.

Figure 4: Marriage gain

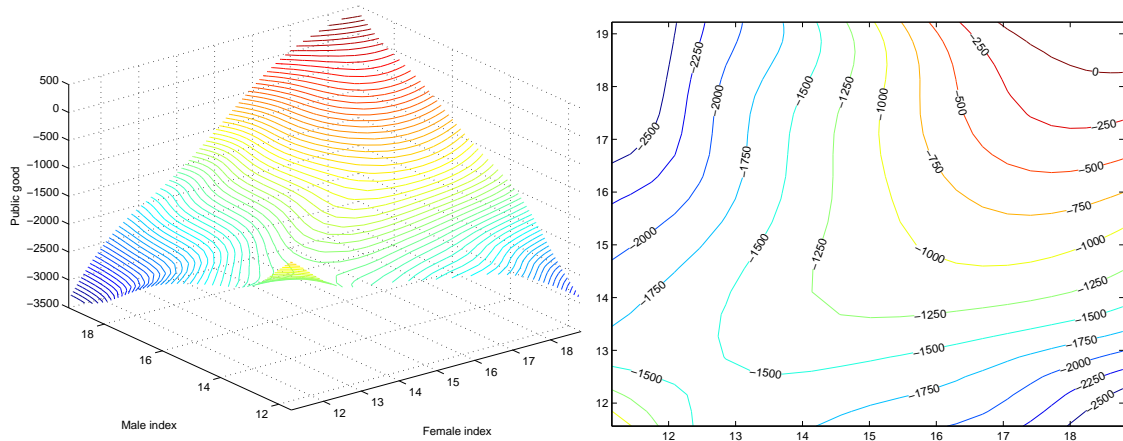
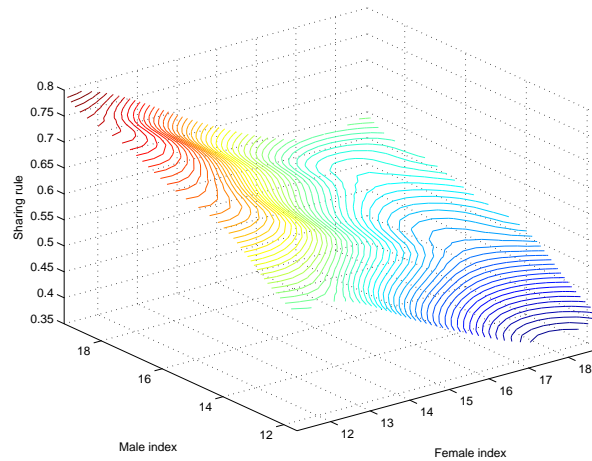


Figure 5 represents the share of the gain that the man receives ($\frac{t_i}{t_i+t_j}$). In general, the man gets an higher share of the surplus than the woman, his share is between 0.4 and 0.8. His share decreases with his partner's index and increases with his own index. The average share of men is equal to 56% of the total marriage gain.

Figure 5: Sharing rule



Multidimensional model

In this section, I present the results obtain with the multidimensional model. Figure 6 and 7 represent the marriage probability between two matched individuals. I represent these 4 variable functions on 3D graphs. Figure 6 shows the expected marriage probability when two singles meet conditional on their BMI whereas Figure 7 shows their expected marriage probability conditional on their wages. The match probability is higher for women with a low BMI (around 20) with men with average BMI (around 26). High-BMI women have very few chances to marry. We remark that when men and women have high BMI, the marriage probability depends less on the BMI of their partners (similar pattern is observed for the marriage gain on figure 10 in appendix). It works as if people with "ideal" physical norms have greater interest in finding the ideal partners with respect to physical attractiveness than the others.

Considering wages, positive assortative matching is still observed for high wages where the higher

the better. High-wage men and women look for high-wage partners (similar patterns are observed for the marriage gain represented on figure 11 in appendix). However, low-wage partners seem not to have interest in discriminating on wages.

Figure 6: Marriage probability conditional on BMI

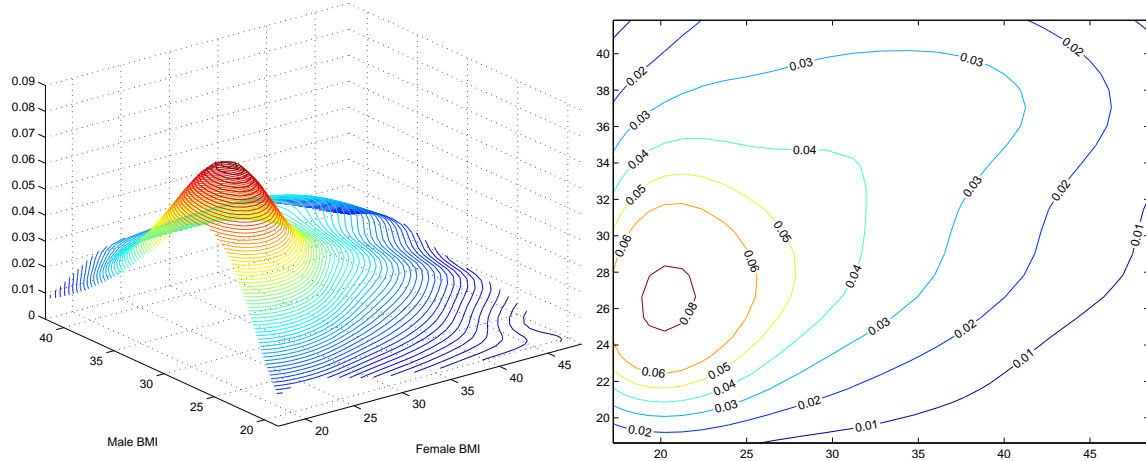


Figure 7: Marriage probability conditional on wages

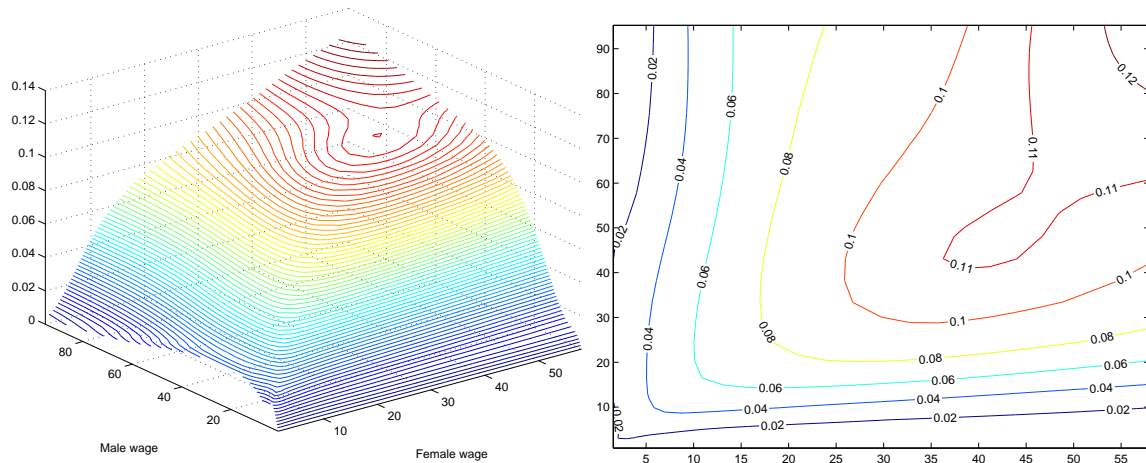
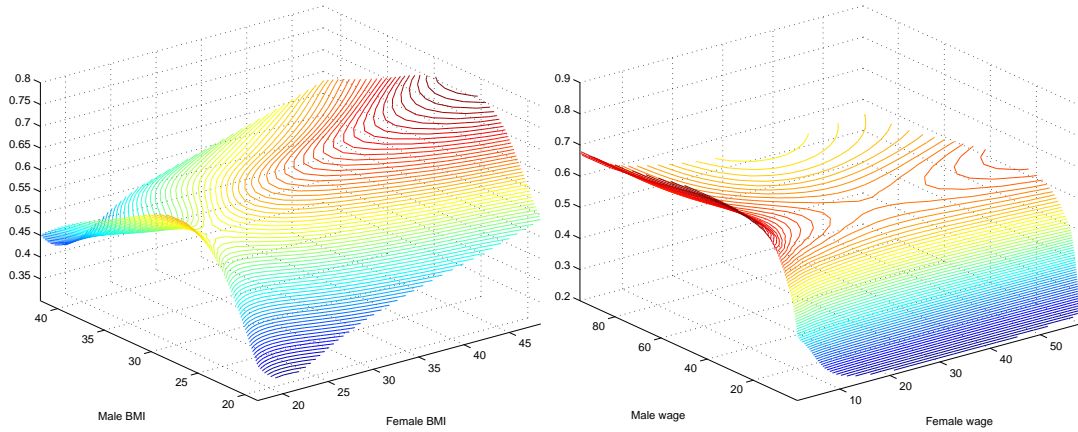


Figure 8 represents the share of the gain that the man receives. His share increases with her partner's Body Mass Index whereas his own Body Mass Index has a low impact on his share. His share also slightly decreases with the wage of his partner.

Figure 8: Sharing rule



The previous graphs show the shape of the marriage gain with respect to two out of the four characteristics : w_i, w_j, bmi_i, bmi_j . It would be interesting to better understand the contribution of each variable to the marriage gain and their interaction. To this end, I perform a nonparametric regression of the marriage gain using Generalized Additive Models (GAMs). I present in table 4 the generalized R^2 obtained for different specifications of the interactions of variables (I put a note on GAM in appendix). There exist complementarities between characteristics of the partners. The marriage gain is not a simple sum of the contributions of each partner. Table 4 shows that this sum would only reach 77% of the total variation. We remark that partner's wages plus their interaction account for 23.6% of the variance whereas the sum of each wage contribution would account for only 18.9%. Similarly, the interaction between partners' BMI accounts for 62.5% whereas the sum of their individual contribution is only 58.3%. There is also some interaction between the male wage and the woman BMI. Both variables plus their interaction explain 59% of the marriage gain, whereas without interaction, both variables would explain 55.7% of the marriage gain. However, there is apparently no interaction between the male BMI and his wage and no interaction between the female wage and the male BMI (including some interaction doesn't add more explanation power).

Characteristics of the women explain by themselves 53.5% of the variation whereas characteristics of the man explain 25.5 %. Besides, the physical attractiveness of the women by itself explains 44% of the variation of the marriage gain whereas the BMI of men only explains 14%.

Table 6: Comparisons of different models

Variables	Generalized R^2
(w_m)	11.4 %
(bmi_m)	14.0 %
(w_f)	7.5 %
(bmi_f)	44.3 %
(w_m, bmi_m)	25.5 %
(w_f, bmi_f)	53.5 %
(w_m, w_f)	23.6 %
(bmi_m, bmi_f)	62.5 %
(w_m, bmi_f)	59.0 %
(w_f, bmi_m)	21.3 %
(w_m, w_f, bmi_m, bmi_f)	92.3 %

6 Simulation of equilibrium

6.1 Equilibrium distribution of characteristics

In this section, I compute the equilibrium distribution of characteristics and match probabilities from the previous non-parametric estimates of the structural parameters, namely, the complementarities in characteristics $C(i, j)$ and the preference parameters. Let $n(i)$ denote the density of married men of type i then $n(i) = \int n(i, j) dj$. Besides, remember that

$$n(i, j) = \frac{\lambda U_m(i) U_f(j) a(i, j)}{N \lambda_z (1 - a(i, j))},$$

then we obtain the following equation for the distribution of type i among single men

$$U_m(i) = \frac{L_m(i)}{1 + \frac{\lambda}{\lambda_z} \int_j \frac{a(i, j) U_f(j)}{1 - a(i, j)} dj}.$$

Similarly, we obtain the expression for density of single women. Now, we will compute the equilibrium expression of the present value of single men and single women. We obtain the following Bellman equation for a single individual using equation (1).

$$\begin{aligned} rV_i^0 &= \lambda \iint \max(W_{i|j}(z) - V_i^0, 0) dj dG(z) \\ &= \frac{\lambda \beta}{r + \lambda_z} \iint \max(s(i, j) + z, 0) U_f(j) dj dG(z) \\ &= \frac{\beta \lambda}{r + \lambda_z} \int \left(\int \max(s(i, j) + z, 0) dG(z) \right) dU_f(j). \end{aligned}$$

We first compute the inside integral on z :

$$\begin{aligned}
 \int_z \max(s(i, j) + z, 0) dG(z) &= s(i, j)a(i, j) + \int_{-s(i, j)}^{+\infty} z dG(z) \\
 &= s(i, j)a(i, j) + \sigma \int_{-\frac{s(i, j)}{\sigma}}^{+\infty} v d\Phi(z) \\
 &= s(i, j)a(i, j) + \sigma \phi\left(\frac{s(i, j)}{\sigma}\right) \\
 &= \mu(a(i, j)),
 \end{aligned}$$

with Φ the Gaussian standard distribution function and ϕ its associate density. We obtain the following formula for the present value of a single man of type i and a woman of type j

$$\begin{aligned}
 rV_i^0 &= \frac{\beta\lambda}{r + \lambda_z} \int \mu(a(i, j)) dU_f(j) \\
 rV_j^0 &= \frac{(1 - \beta)\lambda}{r + \lambda_z} \int \mu(a(i, j)) dU_m(i).
 \end{aligned}$$

An equilibrium is a fixed point of (u_m, u_f, V_i^0, V_j^0) of the following system of equations

$$\begin{aligned}
 u_m(i) &= \frac{L_m(i)}{U_m + \frac{\lambda}{\lambda_z} U_m \int \frac{a(i, j) U_f(j)}{1 - a(i, j)} dj} \\
 u_f(j) &= \frac{L_f(j)}{U_f + \frac{\lambda}{\lambda_z} U_f \int \frac{a(i, j) U_m(i)}{1 - a(i, j)} di} \\
 rV_i^0 &= \frac{\beta\lambda}{r + \lambda_z} \int \mu(a(i, j)) dU_f(j) \\
 rV_j^0 &= \frac{(1 - \beta)\lambda}{r + \lambda_z} \int \mu(a(i, j)) dU_m(i),
 \end{aligned}$$

where $a(i, j)$ solves the following fixed point equation :

$$a(i, j) = 1 - G\left(-C(i, j) + rV_i^0 + rV_j^0 - \frac{\lambda_z}{r + \lambda_z} \mu(a(i, j))\right).$$

The first two equations determine equilibrium type distributions for singles and the last two equations determine equilibrium present values of singlehood. Then we can verify that we can go backward, that is, calculate the equilibrium index distributions and match probabilities from the previous non-parametric estimates of the structural parameters, namely, the marriage externality function $C(i, j)$ and the preference parameters. Despite the lack of a global contraction mapping property, I found that the standard fixed-point iteration algorithm, $x_{n+1} = Tx_n$ worked well in practice, even starting far from the equilibrium (like with $V_i^0 = 0$ and $u_m(i) = l_m(i)$).

To solve the whole model, I then need to choose the set of the following fixed parameters (r, β, σ_z) . As usual in search literature, I choose a bargaining power β equal to 0.5, assuming that men and women in developed societies have the same potential decision power (independently of their

outside option). I choose a variance of 1000 for the random variable z . This parameter actually doesn't have any impact on the shape of the results¹¹. Then I set the discount interest r at 5% a year. The algorithm converges to the equilibrium observed in the data. Indeed, we obtain the fixed point (u_m, u_f, V_i^0, V_j^0) which corresponds to the density and present value we observe in the data. The first two lines of table 7 show the convergence of the simulated equilibrium toward the empirical distributions.

6.2 Simulations of other equilibria

Taxation setting

For redistributive motives (vertically across households or horizontally between singles and couples), the social planner may want to tax the good produced by couples. As taxation can impact match formation and change within-household transfers, it is interesting to look at the effect of taxation on matching patterns. We could think of two types of taxation: taxation of the inputs or taxation on the output if the social planner could identify it. A tax on the output is similar to tax the transfer each member receives. We can first consider a flat tax τ on both transfers. Then the individuals receive the transfers $t_i^\tau = t_i(1 - \tau)$ and $t_j^\tau = t_j(1 - \tau)$. We should then expect that fewer marriages will form as the probability that a couple of type (i, j) matches is now lower with¹²

$$a(i, j)^\tau = \mathbb{P} \left(z > -C(i, j) + \frac{rV_i^0 + rV_j^0}{1 - \tau} - \frac{\lambda_z}{r + \lambda_z} \int \max(s_{ij} + z', 0) dG(z') \right) \leq a(i, j)^0$$

We can also consider different taxes on transfers for men and women. This could correspond to different valuations of resources for each gender. With a tax τ_i on a transfer of man of type i and a tax τ_j on the transfer of a woman of type j , the probability that a couple of type (i, j) matches is now

$$a(i, j)^{\tau_i, \tau_j} = \mathbb{P} \left(z > -C(i, j) + \frac{rV_i^0}{1 - \tau_i} + \frac{rV_j^0}{1 - \tau_j} - \frac{\lambda_z}{r + \lambda_z} \int \max(s_{ij} + z', 0) dG(z') \right) \leq a(i, j)^0$$

Finally, we can also consider taxation on inputs, as these ones may be directly observable. Using a progressive tax on the indexes of individuals, we can implement an individual taxation setting or a joint taxation setting where the tax rate depends on the mean of the indexes of the two partners. Then, we can simulate the matching equilibrium. The idea behind is as the inputs have been modified by the taxation, so is the produced good. Table 7 displays some features of the new equilibria obtained under these different scenarios. I consider the numbers of single men and single women, the average men's share of the marriage gain and the total social surplus of both singles and couples computed as follows¹³:

¹¹Indeed, $a(i, j)$ doesn't depend on σ_z . As $s(i, j)$ is proportional to s_z with $s(i, j) = -\sigma_z \Phi^{-1}(1 - a(i, j))$, rV_i^0 and rV_j^0 are proportional to σ_z and so are $t_{i|j}$, $t_{j|i}$ and $C(i, j)$. Finally $rW_{i|j}$ and $rW_{j|i}$ are also proportional to σ_z .

¹²I show the general demonstration in appendix.

¹³The last equation comes from equation 9 demonstrated in appendix.

$$\begin{aligned}
\mathcal{W} &= \iint \overline{W_{i|j}(z) + W_{j|i}(z)}^{z > -s(i,j)} n(i, j) di dj + \int V_j^0 u_f(j) dj + \int V_i^0 u_m(i) di + T \\
&= \iint (\overline{W_{i|j}(z)} - V_i^0) + (\overline{W_{j|i}(z)} - V_j^0) n(i, j) di dj + \int V_j^0 l_f(j) dj + \int V_i^0 l_m(i) di + T \\
&= \iint \overline{S_{ij}(z)} (\beta(1 - \tau_i) + (1 - \beta)(1 - \tau_j)) n(i, j) di dj + \int V_j^0 l_f(j) dj + \int V_i^0 l_m(i) di + T,
\end{aligned}$$

where T is the total amount of taxes collected by the social planner and redistributed uniformly across individuals (this transfer is not included in V_i^0 and $W_{i|j}$)¹⁴.

Numerical simulations

To get the results, I set a flat tax at 0.3 and a progressive tax on women transfers which depends on her index. I consider two thresholds. The tax rate is zero for indexes below 12, then it amounts at 0.80 and reaches 0.90 for indexes above 17. The progressive tax on inputs are sets at zero below an index of 12, at 0.3 above 12 and below 17 and 0.5 above 17. The tax rate is based on the individual's index for individual taxation and is based on the mean of the two indexes of the partners for joint taxation.

Table 7: Simulation exercises

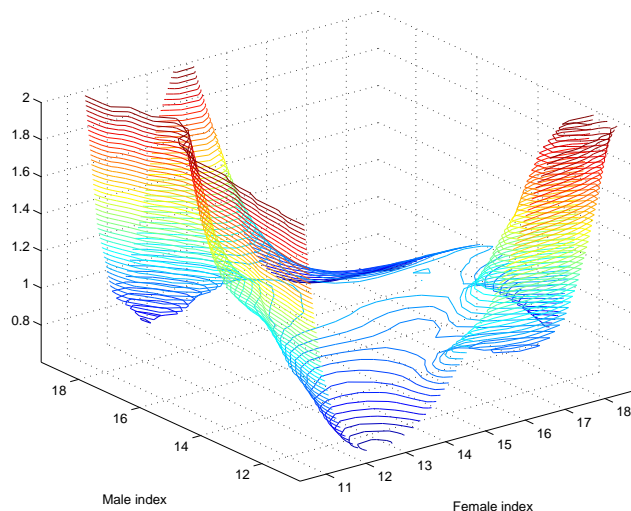
	U_m	U_f	Average man's share	Social surplus
1999 data	655	1119	0.5603	5.23×10^7
Simulated equilibrium	678	1064	0.5525	5.21×10^7
Tax on transfers				
Flat tax on transfers	777	1163	0.5440	4.44×10^7
Progressive taxes on women's transfers	697	1082	0.5276	5.01×10^7
Progressive taxes on men's transfers	706	1092	0.5757	4.96×10^7
Tax on inputs				
Individual progressive tax on inputs	735	1120	0.5528	4.49×10^7
Joint progressive tax on inputs	764	1149	0.5490	4.32×10^7

When you tax both transfers with a rate of 30%, about 3% of couples break up among the less productive (with lower indexes). The distribution of single women moves slightly to the right which improves a little bit their bargaining power within couples. The average share of men decreases by 2.5%. The total social welfare decreases also. Less couples form so less household good is created. When you tax only men or women transfers, the number of couples decreases also a little bit for similar reasons. When women's transfers are taxed, many couples where transfer is already low break up, the others stay together if she could increase her transfer. The average marriage gain of men decreases by 4.5%. For similar reasons, the average gain of men increases

¹⁴Present values are linear in transfers. I assume there is no cost in collecting the tax.

by 2.5% when men's transfers are taxed. When inputs are taxed, we also observe a decrease in the number of couples as the marriage gain decreases if the inputs decrease. Individual taxation decreases the output but doesn't change the bargaining terms of men and women, however joint taxation improves the marriage gain of women. I represent on figure 9 the ratio of the new joint density of indexes among couples over the previous one. It shows that couples with very different indexes have doubled their probability to marry under joint taxation. This makes the bargaining power of low index women increases and they get an higher share of the surplus.

Figure 9: Change in joint density of indexes among couples after joint progressive taxation on inputs



7 Conclusion

This paper proposes a multidimensional search and matching model of marriage which allows people to choose their partner according to their education level, their wage and their physical attractiveness. I find that wages account for 30% of men's attractiveness and 20% of women's attractiveness whereas physical characteristics account for 30% for men and 35% for women. Education accounts then for 40% for men and 45% for women. Considering efficiency, I find the marriage market of the young is more efficient than the one of older individuals: young individuals (below 35 years old) are more selective.

The setting allows me to identify the within-household transfers. Men get around 56% of the marriage gain. Taxation of the marriage gain through taxation of skills or taxation of transfers can modify the sharing rule. I compare matching equilibria obtained with individual and joint taxation of the household and I show that joint taxation slightly improves women's share of the surplus through less sorting on characteristics on the marriage market.

A natural extension of this model would be to allow individuals to work and consume goods. The additional revenue they get through marriage would then influence their consumption and labor supply. The ultimate goal of this type of model would be to evaluate the impact of a family policy program on men and women consumption and labor supply taking into account marital sorting and within-household transfers.

Bibliography

- S.L. Averett, A. Sikora, and L.M. Argys. For better or worse: relationship status and body mass index. *Economics & Human Biology*, 6(3):330–349, 2008.
- G.S. Becker. A theory of marriage: Part I. *Journal of Political economy*, 81(4), 1973.
- M.P. Bitler, J.B. Gelbach, H.W. Hoynes, and M. Zavodny. The impact of welfare reform on marriage and divorce. *Demography*, 41(2):213–236, 2004.
- M. Bozon and F. Héran. La formation du couple. *Population*, 1:152–159, 1991.
- M. Browning, P-A. Chiappori, and A. Lewbel. Estimating consumption economies of scale, adult equivalence scales, and household bargaining power. *The Review of Economic Studies*, 80(4):1267–1303, 2013.
- K. Burdett and M.G. Coles. Long-term partnership formation: marriage and employment. *Economic Journal*, pages 307–334, 1999.
- P-A. Chiappori, B. Fortin, and G. Lacroix. Marriage market, divorce legislation, and household labor supply. *Journal of political Economy*, 110(1):37–72, 2002.
- P-A. Chiappori, S. Oreffice, and C. Quintana-Domeque. Matching with a handicap: The case of smoking in the marriage market. 2010.
- P-A. Chiappori, S. Oreffice, and C. Quintana-Domeque. Black-white marital matching: Race, anthropometrics, and socioeconomics. 2011a.
- P-A. Chiappori, B. Salanié, and Y. Weiss. Partner choice and the marital college premium. *Discussion Papers*, 2011b.
- P-A. Chiappori, S. Oreffice, and C. Quintana-Domeque. Fatter attraction: anthropometric and socioeconomic matching on the marriage market. *Journal of Political Economy*, 120(4):659–695, 2012a.
- P-A. Chiappori, B. Salanié, F. Salanié, and A. Gandhi. From aggregate betting data to individual risk preferences. 2012b.
- E. Choo and A. Siow. Who marries whom and why. *Journal of Political Economy*, 114(1):175, 2006.
- H. Couprie. Time allocation within the family: Welfare implications of life in a couple. *The Economic Journal*, 117(516):287–305, 2007.
- A. Dupuy and A. Galichon. Personality traits and the marriage market. (*Forthcoming*) *Journal of Political Economy*.
- M. Francesconi, H. Rainer, and W. Van Der Klaauw. The effects of in-work benefit reform in britain on couples: Theory and evidence. *The Economic Journal*, 119(535):F66–F100, 2009.

- A. Galichon and B. Salanié. Cupid's invisible hand: Social surplus and identification in matching models. Technical report, mimeo, 2011.
- J. Greenwood, N. Guner, G. Kocharkov, and C. Santos. Marry your like: Assortative mating and income inequality. Working Paper 19829, NBER, January 2014. URL <http://www.nber.org/papers/w19829>.
- N. Jacquemet and J-M. Robin. Assortative matching and search with labour supply and home production. Technical report, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 2013.
- M. Manser and M. Brown. Marriage and household decision-making: A bargaining analysis. *International Economic Review*, 21(1):31–44, 1980. ISSN 0020-6598.
- R.D. Mare. Educational assortative mating in two generations. *Department of Sociology, University Of California Los Angeles*, 2008.
- M.B. McElroy. The empirical content of nash-bargained household behavior. *Journal of human resources*, pages 559–583, 1990.
- M.B. McElroy and M.J. Horney. Nash-bargained household decisions: Toward a generalization of the theory of demand. *International Economic Review*, 22(2):333–349, 1981.
- R. Shimer and L. Smith. Assortative matching and search. *Econometrica*, 68(2):343–369, 2000.
- J. Waldvogel. Fast construction of the Fejér and Clenshaw–Curtis quadrature rules. *BIT Numerical Mathematics*, 46(1):195–202, 2006. ISSN 0006-3835.
- L.Y. Wong. Structural estimation of marriage models. *Journal of Labor Economics*, 21(3):699–727, 2003.
- S. Wood. *Generalized additive models: an introduction with R*. CRC press, 2006.

1 Appendix

Table 8: Sample sizes

Age range*	Sample size	Single's share	Men share
22-30	2687	30 %	47 %
All	8734	20 %	47 %

* Age range of women (age range of men is shifted by 1 additional year).

1.1 two-components and one-component indexes

Table 9: Estimates from the wage matching model

Age range*	λ	λ_z	$\log L$
22-30	0.0035 (0.0016)	0.061 (0.0039)	-10037
All	0.023 (0.0041)	0.026 (0.00034)	-26337

* Age range of women (age range of men is shifted by 1 additional year).

Table 10: Estimates from the education matching model

Age range*	λ	λ_z	$\log L$
22-30	0.05 (0.013)	0.058 (0.0012)	-9950
All	0.010 (0.0029)	0.026 (0.00093)	-26355

* Age range of women (age range of men is shifted by 1 additional year).

Table 11: Estimates from the BMI matching model

Age range*	λ	λ_z	$\log L$
22-30	0.0015 (0.00020)	0.073 (0.00030)	-10088
All	0.0019 (0.00023)	0.028 (0.00036)	-26618

* Age range of women (age range of men is shifted by 1 additional year).

Table 12: Estimates from the two-component index : wage-education

Age*	α_m	α_f	λ	λ_z	$\log L$
22-30	0.01 (0.00)	0 (0.60)	0.044 (0.036)	0.058 (0.00079)	-9945
All	0.26 (0.00)	0.22 (0.00)	0.0039 (0.0074)	0.027 (0.0075)	-26312

* Age range of women (age range of men is shifted by 1 additional year).

Table 13: Estimates from the two-component index : wage-BMI

Age*	α_m	α_f	λ	λ_z	$\log L$
22-30	0.45	0.45	0.0049	0.056	-9877
	(0.00)	(0.00)	(0.00073)	(0.0011)	
All	0.50	0.48	0.0055	0.026	-26166
	(0.011)	(0.00)	(0.019)	(0.0043)	

* Age range of women (age range of men is shifted by 1 additional year).

1.2 Bidimensional model : Marriage gain

Figure 10: Marriage gain conditional on BMI

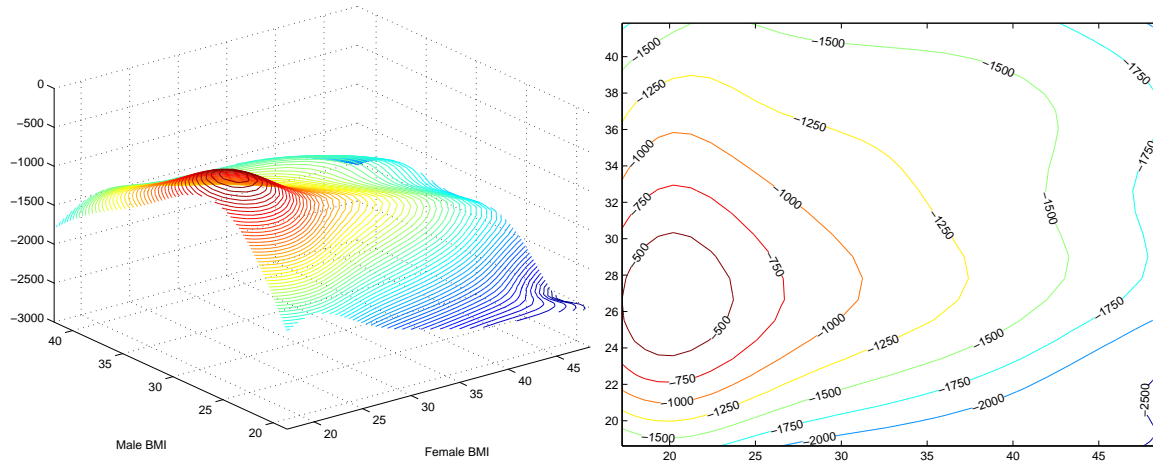
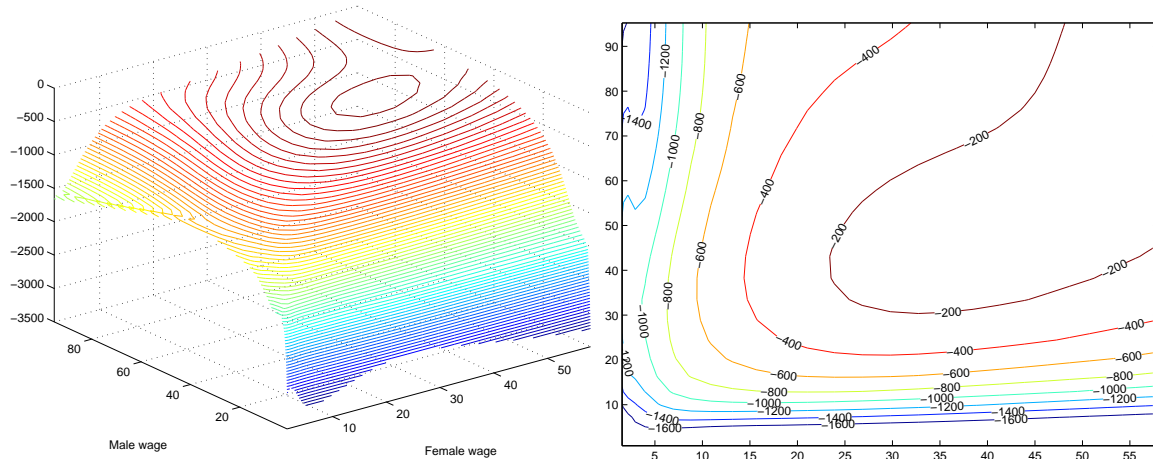


Figure 11: Marriage gain conditional on wages



1.3 Generalized Additive Models

This explanation has been developed in [Chiappori et al. \(2012b\)](#). Generalized additive models (GAM) were introduced by Hastie and Tibshirani (1986). They model a variable y_i by assuming that its distribution around its mean belongs to the exponential family and by modeling the mean as a sum of smooth functions of subvectors of the covariates (X_i). To estimate my GAM models, I use the methods described by [Wood \(2006\)](#); I use his implementation in the `mgcv`

package of R, which incorporates the improved algorithm of Wood (2008). More precisely, one writes

$$\mathbb{E}(y_i|X) = \sum_{j=1}^J f_j(X_i^j)$$

where each X_i^j is a user-defined subvector of X_i , and the f_j are to be estimated; and the user also chooses the distribution of the error term $(y_i - \mathbb{E}y_i)$ within the exponential family. Modeling starts by choosing a rich family of basis functions (typically splines) (b_{jk}) for $k = 1 \dots K_j$ with a maximal order K_j chosen large enough. Then

$$f_j(X_i^j) = \sum_{k=1}^{K_j} \beta_{jk} b_{jk}(X_i^j)$$

Finally, the generalized R^2 cited in the text are defined as the ratio $1 - \frac{\mathbb{E}V(y|X)}{V(y)}$

1.4 Modification with taxation

When transfers are taxed, equation 2 becomes

$$(r + \lambda_z)W_{i|j}(z) = (1 - \tau_i)t_{i|j}(z) + \lambda_z \int_{z'} \max(W_{i|j}(z'), V_i^0) dG(z'), \quad (7)$$

whereas we still have $t_{i|j}(z) + t_{j|i}(z) = C(i, j) + z$. Then the solution of the Nash bargaining gives :

$$\frac{W_{i|j}(z) - V_i^0}{\beta(1 - \tau_i)} = \frac{W_{j|i}(z) - V_j^0}{\beta(1 - \tau_j)} = S(i, j, z),$$

and then

$$\begin{aligned} W_{i|j}(z) - V_i^0 &= \beta(1 - \tau_i)S(i, j, z) \\ W_{j|i}(z) - V_j^0 &= (1 - \beta)(1 - \tau_j)S(i, j, z). \end{aligned} \quad (8)$$

Subtracting $(r + \lambda_z)V_i^0$ in Equation 7, we obtain

$$\begin{aligned} (r + \lambda_z)\beta(1 - \tau_i)S(i, j, z) &= (1 - \tau_i)t_{i|j}(z) - rV_i^0 + \lambda_z \int_{z'} \max(W_{i|j}(z') - V_i^0, 0) dG(z') \\ (r + \lambda_z)(1 - \beta)(1 - \tau_j)S(i, j, z) &= (1 - \tau_j)t_{j|i}(z) - rV_j^0 + \lambda_z \int_{z'} \max(W_{j|i}(z') - V_j^0, 0) dG(z'), \end{aligned}$$

then dividing each equation by $(1 - \tau_a)$, $a = i, j$ and summing the two we obtain

$$(r + \lambda_z)S(i, j, z) = C(i, j) + z - \frac{rV_i^0}{1 - \tau_i} - \frac{rV_j^0}{1 - \tau_j} + \lambda_z \int_{z'} \max(S(i, j, z'), 0) dG(z').$$

Moreover summing the two equations in 8 gives

$$W_{i|j}(z) - V_i^0 + W_{j|i}(z) - V_j^0 = (\beta(1 - \tau_i) + (1 - \beta)(1 - \tau_j))S(i, j, z) \quad (9)$$

1.5 Computational details

The computational method is the one used by [Jacquemet and Robin \(2013\)](#) and is adapted to fit a 4D-dimensional model. All functions are discretized on a compact domain using Tchebychev grids. For example, let $[\underline{x}, \bar{x}]$ denote the support of male wages, I construct a grid of $n + 1$ points as

$$x_j = \frac{\underline{x} + \bar{x}}{2} + \frac{\bar{x} - \underline{x}}{2} \cos\left(\frac{j\pi}{n}\right), \quad j = 1 \dots n$$

To estimate wage densities $n(x, y)$, $u_m(x)$ and $u_f(y)$ on those grids, I use kernel density estimators with twice the usual bandwidth to smooth the density functions in the tails. Indeed, additional smoothing is required to divide $n(x, y)$ by $u_m(x)u_f(y)$ to calculate $a(x, y)$. (In Matlab, I need $10^{-16} < a(i, j)$).

The Clenshaw-Curtis quadrature

Many equations involve integrals. Given Tchebychev grids, it is natural to use Clenshaw-Curtis quadrature to approximate these integrals. 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. The quadrature weights are

$$\begin{aligned} w_0 &= \frac{1}{N} \left(1 + \sum_{j=1}^{\frac{N}{2}} \frac{2}{1 - (2j)^2} \right) & w_{\frac{N}{2}} &= \frac{1}{N} \left(1 + \sum_{j=1}^{\frac{N}{2}} \frac{2(-1)^j}{1 - (2j)^2} \right) \\ w_k &= \frac{2}{N} \left(1 + \frac{(-1)^k}{1 - N^2} + \sum_{j=1}^{\frac{N}{2}-1} \frac{2}{1 - (2j)^2} \cos\left(\frac{2jk\pi}{N}\right) \right) \quad \forall k = 1, \dots, \frac{N}{2} - 1. \end{aligned}$$

I use the method of Jorg Waldvogel ([Waldvogel, 2006](#)) who derives a simple algorithm to obtain the weights of the Clenshaw-Curtis quadrature using matrices, Féjer'quadrature and Discrete Fourier Transform.

Interpolation

The fact that CC quadrature relies on Tchebychev polynomials of the first kind also allows us to interpolate functions very easily between points $y_0 = f(x_0), \dots, y_n = f(x_n)$ using Discrete Cosine Transform (DCT) such that

$$f(x) = \sum_{k=0}^n Y_k T_k(x) \quad (10)$$

where Y_k are the OLS estimates of the regression of $y = (y_0, \dots, y_n)$ on Tchebychev polynomials

$$T_k(x) = \cos \left(k \arccos \left(\frac{x - \frac{x+\bar{x}}{2}}{\frac{\bar{x}-x}{2}} \right) \right)$$

but are more effectively calculated using FFT. A MATLAB code for DCT is

```
Y = y([1:n+1 n:-1:2],:);
Y = real(fft(Y/2/n));
Y = [Y(1,:); Y(2:n,:)+Y(2*n:-1:n+2,:); Y(n+1,:)];
f = @(x) cos(acos((2*x-(xmin+xmax))/(xmax-xmin))*(0:n))*Y(1:n+1);
```

with $y = (y_0, \dots, y_n)$. A bidimensional version is

```
Y = y([1:n+1 n:-1:2],:);
Y = real(fft(Y/2/n));
Y = [Y(1,:); Y(2:n,:)+Y(2*n:-1:n+2,:); Y(n+1,:)];
Y = Y(:, [1:n+1 n:-1:2]);
Y = real(fft(Y'/2/n));
Y = [Y(1,:); Y(2:n,:)+Y(2*n:-1:n+2,:); Y(n+1,:)]';
f=@(x,y) cos(acos((2*x-(xmin+xmax))/(xmax-xmin))*(0:n))*Y(1:n+1,1:n+1)...
    *cos((0:n)'*acos((2*y'-(ymin+ymax))/(ymax-ymin))));
```

I also use a 4D dimensional version to evaluate transfers which depend on 4 variables : w_m, w_f, f_m, f_f . The fact that the grid (x_0, \dots, x_n) is not uniform and is denser towards the edges of the support interval allows to minimize the interpolation error and thus avoids the standard problem of strong oscillations at the edges of the interpolation interval (Runge's phenomenon). Another advantage of DCT is that, having calculated Y_0, \dots, Y_n , then polynomial projections of $y = (y_0, \dots, y_n)$ of any order $p \leq n$ are obtained by stopping the summation in (22) at $k = p$. Finally, it is easy to approximate the derivative f' or the primitive $\int f$ simply by differentiating or integrating Chebyshev polynomials using

$$\cos(k \arccos x)' = \frac{k \sin(k \arccos x)}{\sin(\arccos x)}$$

and

$$\begin{aligned} \int \cos(k \arccos x) &= x \text{ if } k = 0 \\ &= x^2/2 \text{ if } k = 1 \\ &= \frac{\cos((k+1)x)}{2(k+1)} - \frac{\cos((k-1)x)}{2(k-1)} \text{ if } k \geq 1. \end{aligned}$$

In calculating an approximation of the derivative, it is useful to smooth the function by summing over only a few polynomials. Derivatives are otherwise badly calculated near the boundary.

Chapter 2

Marriage Market and Intra-Household Allocation

Abstract

What is the impact of the marriage market on the labor supply of men and women ? Collective models have proved that opportunities of spouses outside marriage can influence the intra-household balance of power and ultimately the final allocation of resources. However these models consider couples as given and can't predict the impact of welfare policies on the sharing rule. In this paper, I model the intra-household allocation of resources jointly with the formation and separation of couples in a dynamic search and matching framework. Using the British Household Panel Survey (BHPS), I estimate the matching preferences of individuals over different characteristics such as wages and family values. Taking domestic production into account, I identify the within household transfers and show that they reduce labor supply of married women by 2 hours a week and increase married men labor supply by 1 hour a week.

1 Introduction

To evaluate the impact of taxation reforms and family policy programs on labor supply and inequalities, we need to understand the intra-household allocation of time and consumption. Classic economic theory considers the household as a unit and neglects bargaining issues within the household. The two individuals of the household are supposed to pool their income and maximize a neoclassical household utility function subject to the household's budget constraint. However, many empirical studies show that the income pooling hypothesis is rejected by the data¹. The pooling assumption at the aggregate level leads to the underestimation of income inequalities among individuals (Lise and Seitz, 2011) and to a bias in the estimation of labor supply trends (Knowles, 2013). Collective models² propose an identification strategy to recover the sharing rule from observed labor supplies of couples. However, whereas these models repeatedly show evidence that the within household sharing rule varies with the outside options of individuals (Chiappori et al., 2002), they consider couples as given and can't predict the impact of welfare policies on the sharing rule. Yet, such reforms may influence marital sorting through divorce and couples formation (Francesconi et al., 2009; Bitler et al., 2004).

Consequently, we need a model which could explain both the formation and separation of couples and the intra-household allocation. In this paper, I jointly model the marriage market and resource sharing within the household using the British Household Panel Survey (BHPS) where I observe wages, working hours, domestic work and marital history of each household member from 1991 to 2008. First I recover the matching patterns and the preferences for leisure, consumption and domestic production of men and women. Then, identifying the within household transfers, I show that changes in matching patterns have a significant impact on labor supply of men and women through changes in intra-household allocation of resources.

This paper bridges the gap between matching models which identify the matching preferences over different characteristics, and collective models which identify the within household transfers and their impact on economic outcomes. To my knowledge, the work of Jacquemet and Robin (2013) is the first attempt to link heterogeneity in marriage formation and intra-household allocation. They recover matching patterns from observed joint distributions of characteristics among couples and among singles³ and use a search framework to model frictions⁴. The present paper builds on their framework. It similarly includes a collective structure of labor supply in a search and matching model of marriage. Individuals are egoistic and enjoy their own leisure and consumption. Single individuals earn labour income whereas married individuals may also benefit from a transfer from their spouse on top of their labour income. Single individuals meet randomly and decide whether they marry. They evaluate the match anticipating what will be the surplus generated by the match and how it will be split. Individuals bargain *à la Nash*⁵ to

¹See (Bourguignon et al., 1994)

²These models developed by Chiappori (Chiappori, 1988) assume that the household members bargain over their resources and make Pareto-optimal agreement.

³Similar strategies are used in models with perfect information as in Choo and Siow (2006), Chiappori et al. (2011) and Galichon and Salanié (2011)

⁴Identification of matching patterns is obtained with the steady-state assumption of search models as Shimer and Smith (2000) and Wong (2003)

⁵Nash-Bargaining within household is a particular case of collective models: it gives Pareto efficient outcomes.

choose an optimal sharing rule. Either the surplus is high enough and both want to match, or it is not and both prefer to stay single. If they match, they first split the surplus and then choose separately their consumption and leisure according to their new budget constraint.

My paper extends the paper of Jacquemet and Robin (2013) in three ways. First, individuals directly enjoy the consumption of a domestically produced public good in addition to leisure and consumption. When two people decide to live together, joint domestic production can increase their utility in two ways. First their purchasing power increases due to economies of scale (sharing the rent, the electricity, increasing returns to scale in cleaning, meal preparation)⁶. Second, individuals may also directly enjoy the jointly produced public good which may consist in raising children or eating a home-made meal. Domestic production is crucial in analyzing household behavior. Omitting household production leads to a significant bias in the estimation of the sharing rule (Couprie, 2007). The use of the BHPS is particularly relevant here as we observe the time spent in housework by each household member. I specify a domestic production function in which public good production depends on three different inputs: the time spent in housework by each partner, the characteristics of each partner and some time-varying unobserved characteristics of the match.

This unobserved heterogeneity leads to my second contribution, I endogenise the separation of couples. Some shocks can hit the unobserved characteristics and lower the value of the match. In that case, the match breaks up. Only couples with high enough complementarity in observed characteristics will last. This new setting allows me to identify the instantaneous meeting probability which was not possible in the setting of Jacquemet and Robin. Using the panel structure of the BHPS, I observe partnership duration and singlehood duration and I can identify meeting parameters and quality shock parameters.

Finally, I extend the setting to multidimensional matching. As one household member's value on the marriage market has an impact on the sharing rule, all characteristics which are important in couple formation must have an impact on the sharing rule and then on labor supply. I allow people to choose their partners for two continuous different characteristics such as wages and family values⁷. Only a few papers consider multidimensional matching. Some build a marriageability index (Wong, 2003; Chiappori et al., 2012a), others match on two characteristics, one of which is a binary characteristic (Chiappori et al., 2010). Only the recent work of Dupuy and Galichon uncovers matching preferences over many different continuous characteristics.

The model fits the data well and provides several results. First, I identify the total surplus of couples for different characteristics of the partners. I find a positive assortative matching in wages and in family values. I disentangle what comes from preferences for the characteristics of

It is a simple way to include the outside option in the bargaining process. See Manser and Brown (1980), McElroy and Horney (1981) and McElroy (1990) for deeper discussion on Nash-Bargaining.

⁶Many studies have attempted to estimate the additional revenue generated by living in couples (Browning et al., 2013; Couprie, 2007). In collective models with domestic production, the intra-household allocation process (Chiappori, 1997) can also be decentralized. First, individuals decide on the level of domestic production they want. Then they define a conditional sharing rule that is how they will share the rest of the total income conditional on the chosen level of domestic production.

⁷I use an index representing family values. The higher this index the more traditional the individual is about family and gender roles. This index expresses opinion about divorce, marriage institution, etc.

the other partner and what comes from resource sharing and total income. I show that if the total surplus increases in wages of both members, complementarities in characteristics can be higher for same wage couples. I also find that women with traditional family values as strong beliefs in religion and marriage institution are more attractive on the marriage market. When married, these women get a higher share of the couple surplus both in terms of welfare and monetary resources. Furthermore, I show that matching patterns have slightly changed overtime and that high wage women have become more attractive on the marriage market.

Second, I identify the within household transfers and show they have a significant impact on working hours. As most wives get positive transfers from their male partner, they work less on average than if they didn't get any transfer but they work more than if the total resource of the household was shared equally between the two individuals. On the opposite side, husbands work more to compensate the revenue they give to their wife but less than if they had to split the total resource in two equal parts. In comparison to a situation where married individuals do not get any transfer from their spouse, I show that transfers reduce labor supply of married women by 2 hours a week and increase married men labor supply by 1 hour a week.

Finally, this model allows me to simulate the counterfactual equilibrium which would have obtained in 2008 if sorting patterns and/or preferences for consumption and leisure were still at their 1999 level. Among other results, I show that changes in sorting pattern have increased the number of couples with rich women, which leads to a decrease of 1 hour a week the average labor supply of single women and a small increase of 0.8 hours a week in labor supply of married women. I also present some simulation exercises where I simulate the equilibrium obtained with a change in wage distribution of men or women or if we gave a subsidy to low-wage single women. The ultimate goal of this model is to simulate the impact of a family taxation reform on within household allocations and labor supplies. I explain how the model could be extended to include taxation, children, or both.

The model is described in section 2 and the data in section 3. Section 4 presents the estimation strategy and section 5 the results. Simulation and equilibrium conditions are computed in section 6 and direct extensions of the model are proposed in section 7. Section 6 concludes.

2 Model

2.1 The marriage market

I consider a marriage market with L_m males and L_f females. The number of married couples is denoted by N and the respective numbers of single males and single females are $U_m = L_m - N$ and $U_f = L_f - N$. I assume that only singles search for a partner ruling out “on-the-marriage search”. I denote λ the instantaneous probability of a meeting between a random single woman and a random single man. Then $\lambda_m = \lambda U_f$ and $\lambda_f = \lambda U_m$ are the respective instantaneous probabilities of an agent among the population m and f of meeting a new person of the populations f and m .

2.2 Individual types and preferences

Individuals differ in a fixed set of characteristics i that contains their gender, their labor market ability or hourly wage, and other characteristics, which in my application will be reduced to a single index of how much they value the institution of family, denoted FVI for Family Value Index. So $i = (gender, wage, FVI)$ or (g_i, w_i, f_i) . In the case of couples, I will use i for husband's type and j for wife's type.

Individuals can be single or married (or cohabiting). Let $U(i)$ and $L(i)$ denote the distributions of $i = (gender, wage, FVI)$ in the population of singles and in the whole population, respectively. Let also $U_m(i)$ and $U_f(j)$ denote the distributions of wages and FVI for male and female singles separately. I similarly define distribution functions $L_m(i)$ and $L_f(j)$ for the whole populations of men and women. Lastly, let $N(i, j)$ be the distribution of matches with a male of type i and a female of type j . I also use lower-case notations, such as $u(i)$ or $u_m(i)$ or $n(i, j)$ to denote corresponding densities with respect to some adequate measure that is not necessarily the Lebesgue measure, but that we still simply denote as di or dj . Hence, $l_m(i) = \int n(i, j)dj + u_m(i)$, meaning that the number of males of type i in the economy is equal to the number of all married males and all single males of this type.

Individuals draw utilities from consumption $c \geq 0$, leisure $l \in [0, T]$ with T the total amount of time available to any individual, and a public good $Q \geq 0$. Being married or single determines the public good's production process. Let $U_i(c, l, Q)$ denote this utility function, which we index by the individual type i . For later use, we also define the conditional indirect utility function

$$v_i(R, Q) = \max_{c > 0, T > l > 0} U_i(c, l, Q) \text{ s.t. } c + w_i l \leq R,$$

for a given income R and public good Q and where w_i is the wage of individual i . Note that leisure follows from the indirect utility function by application of Roy's identity.

Singles have access to a household production technology that requires domestic time as single input. Let $Q = F_i^0(d)$ denote the home production function for singles of type i . The home production of married couples has the amount of time spent on domestic chores by both spouses (d_i, d_j) as factors, and it varies across matches by spouses' types (i, j) and a match specific component z : $Q = F_{ijz}^1(d_i, d_j)$ for a man i and a woman j . The match-specific component z is drawn from some distribution G at the time of the first meeting and is infrequently updated, with new draws z' from the same distribution G at random times following a Poisson process with parameter λ_z . The home production of married couples also requires some market good expenditure C_{ij} .

2.3 Marriage contracts

A marriage contract between a male of type i and a female of type j with a match-specific characteristic z specifies a utility flow for both spouses, u_i and u_j , and continuation values, $V_{i|j}^1(z')$ and $V_{j|i}^1(z')$, upon realization of the next match-specific shock z' . Let $W_{i|j}(u)$ denote the present value of marriage for an individual of type i receiving a flow utility u if wed to

an individual of type j . Any contingency to the match-specific component z is embodied in the utility flow u . Let V_i^0 denote the value of being single. The value of a marriage contract delivering u units of utility in the current period is defined by the equation,

$$(r + \lambda_z)W_{i|j}(u) = u + \lambda_z \int_{z'} \max(V_{i|j}^1(z'), V_i^0) dG(z'),$$

where r is the discount rate and the second term of the right-hand side is the option value of divorce or match continuation after a shock to the match-specific component.

For singles,

$$rV_i^0 = \max_{0 < d < T} v_i(w_i(T - d), F_i^0(d)) + \lambda_i \left(\iint \max(V_{i|j}^1(z) - V_i^0, 0) du_{g_j}(j) dG(z) \right), \quad (1)$$

where λ_i is either λ_m or λ_f depending on i 's gender $g_i = m, f$.

2.4 Bargaining

When a match is formed, the two members start to bargain on the level of production of the public good and on the repartition of the household resources. I model the decision process with a Nash bargaining where the threat point is to stay single. The respective threat points are then the respective outside options of the man and the woman and so their present value as singles. I denote β and $(1 - \beta)$ the respective bargaining power of the man and the woman. Then the Nash bargaining of the household is the maximization of the following program :

$$\begin{aligned} \max_{d_i, d_j, t_i, t_j} & (W_{i|j}(u_i) - V_i^0)^\beta (W_{j|i}(u_j) - V_j^0)^{1-\beta} \\ \text{s.t.} & t_i + t_j + C_{ij} = 0 \\ & u_i = v_i(w_i(T - d_i) + t_i, Q), u_j = v_j(w_j(T - d_j) + t_j, Q), Q = F_{ijz}^1(d_i, d_j). \end{aligned}$$

This model has the important property of transferable utility models : both surplus are simultaneously positive or negative. Either the match generates enough surplus to make both people want to marry or it doesn't and they both refuse to marry. The first-order condition with respect to transfers t_i, t_j writes as

$$\frac{\beta}{W_{i|j}(u_i) - V_i^0} \frac{\partial u_i}{\partial R} = \frac{1 - \beta}{W_{j|i}(u_j) - V_j^0} \frac{\partial u_j}{\partial R}$$

and after simple algebra, the conditions for d_i, d_j come out as

$$\frac{w_i d_i}{\epsilon_i} = \frac{w_j d_j}{\epsilon_j} = Q \left[\frac{\partial u_i / \partial Q}{\partial u_i / \partial R} + \frac{\partial u_j / \partial Q}{\partial u_j / \partial R} \right]$$

where $\frac{\partial u_i}{\partial R}, \frac{\partial u_j}{\partial Q}$ are the partial derivatives of the indirect utility function $v_i(R, Q)$, evaluated at $R_i = w_i(T - d_i) + t_i$ and $Q = F_{ijz}^1(d_i, d_j)$, and where $\epsilon_i = \frac{\partial \log F_{ijz}^1(d_i, d_j)}{\partial \log d_i}$ is the elasticity of

home production with respect to input d_i , and ϵ_j is the elasticity with respect to d_j . All these elasticities are in general functions of (d_i, d_j, z) . Finally, the promise-keeping constraint imposes that $W_{i|j}(u_i) = V_{i|j}^1(z)$, or equivalently,

$$(r + \lambda_z)[V_{i|j}^1(z) - V_i^0] = u_i + \lambda_z \int \max(V_{i|j}^1(z') - V_i^0, 0) dG(z') - rV_i^0, \quad (2)$$

with a symmetric expression for $V_{j|i}^1(z)$. We also have

$$\begin{aligned} (r + \lambda_z)(V_{i|j}^1(z) - V_i^0) &= u_i + v_{i|j} = \beta \frac{S_{ij}}{B_i} \\ (r + \lambda_z)(V_{j|i}^1(z) - V_j^0) &= u_j + v_{j|i} = (1 - \beta) \frac{S_{ij}}{B_j} \end{aligned} \quad (3)$$

Of course all these derivations hold only if there exists a feasible allocation (d_i, d_j, t_i, t_j) such that $W_{j|i}(u_j) - V_i^0 \geq 0$ and $W_{i|j}(u_i) - V_j^0 \geq 0$. This is a condition on z , i and j . Let $a(i, j) \in [0, 1]$ denote the probability of drawing z from distribution G such that there exists a mutually beneficial marriage contract.

2.5 Equilibrium flows

The characterization 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 overtime. [Burdett and Coles \(1999\)](#) review the different cases that have been considered in the literature. Here I suppose there is no entry of new singles, but the partnerships of type (i, j) are destroyed at rate $\lambda_z(1 - a(i, j))$ whereupon both return to the single market. At the equilibrium, there is equality between inflows and outflows for each type of marriage. Then we have for all couples of type (i, j) , the equality between the number of outflows and the number of inflows:

$$\lambda_z(1 - a(i, j))N(i, j) = U_m(i)\lambda_m \frac{U_f(j)}{U_f} a(i, j) = \lambda U_m(i)U_f(j)a(i, j), \quad (4)$$

The left-hand side is the flow of divorces. The right-hand side is the flow of new (i, j) marriages. It has three components: a single male of type i , out of the $U_m(i)$ ones, meets a single female with probability λ_m ; this woman is of type j with probability $U_f(j)/U_f$ the marriage is consummated with probability $a(i, j)$. Notice that, in this model, matches with a higher probability of marriage also have a lower probability of divorce.

The match probability $a(i, j)$ can be then computed using data after having estimated the parameters λ and λ_z .

3 Data

3.1 Data construction

I estimate the model using the British Household Panel Survey (BHPS) where I follow individual's marriage history from 1991 to 2008 using the family and individual samples. I merge the individual file with the marriage history file to obtain marriage history anterior to 1991 for

married people. I only use the original BHPS sample comprising 5,050 households and 9,092 adults interviewed at wave 1 (1991), whom I then follow yearly until 2008, even after separation from the original household⁸. The panel not only follows all individuals from the first wave (original sample members) but also all adult members of all households containing either an original sample member, or an individual born to an original sample member whether or not they were members of the original sample. The sample therefore remains broadly representative of the population of Britain as it changes over time. I only keep households composed of heterosexual couples and single member households who are between 22 and 40 years of age at the time of interview. I drop all observations with missing information on their usual gross pay per month, the number of hours normally worked per week (including paid and unpaid overtime hours) and the number of hours spent in a week doing housework. When married, both spouses have to work and declare their wage and hours to be included in the sample. The hourly wage is the usual gross pay per month divided by the number of hours normally worked per month (without overtime).

I consider two different variables to define the agent's type. Wages must be part of the type as most of the analysis is made on labor income and resource sharing. However, it is quite restrictive to assume that agents only differ by their productivity on the marriage market. Heterogeneity of individuals varies in an infinite number of dimensions which can be important in couples formation. One of the most important observable dimension of heterogeneity on the marriage market is education or social group (Mare, 2008; Bozon and Héran, 1991). The correlation of education among household members is around 0.6. However, some heterogeneity features of education or social group are already captured by the wage. It would be more interesting to study the impact of other variables on the match. The BHPS provides us with some alternatives. We could think of the Body Mass Index observed in 2004 and 2006 which could be a proxy for physical attractiveness. However, I prefer to use in this paper some information on family values available during the whole period. This information reflects how individuals value the marriage institution. I construct a Family Values Index (FVI) based on individuals' responses to various statements about family, cohabitation and divorce on a scale of 1 to 5: Strongly agree (1); Agree (2); Neither agree nor disagree (3); Disagree (4); Strongly disagree (5). Table 1 displays which statements are proposed each year. Let $A(1), \dots, A(9)$ denote the answers to questions 1 to 9. Then

$$I_{fv1991-1996} = \frac{6}{5}[A(1) + (6 - A(2)) + (6 - A(3)) + (6 - A(4)) + (6 - A(5))]$$

$$I_{fv1998-2008} = A(1) + (6 - A(2)) + A(6) + A(7) + (6 - A(8)) + (6 - A(9))$$

with values between 6 and 30. This index has some common features with the traditional versus secular-rational index of Inglehart and Baker (2000).

⁸The sample was a stratified clustered design with 250 Primary Sampling Units in England, Scotland and Wales and was designed to be representative of the population of Great-Britain (which excludes Northern Ireland and North of the Caledonian Canal)

Table 1: Family Value Statements. Do you agree with the following statements ?

Years 1992, 1994, 1996	Years 1998, 2000, 2002, 2004, 2006, 2008
1. <i>Divorce is better than unhappy marriage</i>	
2. <i>Adult children should take care of their parents</i>	
3. <i>Bible Gods word and true</i>	
4. <i>Man should be the head of the household</i>	
5. <i>Cohabiting is always wrong</i>	6. <i>Cohabitation is alright</i>
	7. <i>Marital status is irrelevant for children</i>
	8. <i>Homosexual relationships are wrong</i>
	9. <i>Parents ought stay together for children</i>

Finally, wages are deflated by the Consumer Price Index and computed in pounds of 2008. I trim the 1 % top and bottom tails of wage and time use variables. I thus obtain an unbalanced panel of 18 years (1991-2008), whose cross-sectional size varies between 2255 (in 1991) and 3456 individuals (in 1999).

My analysis has two important limits. First, I need to restrict my sample to working people as I do not model the extensive participation to the labor market jointly with the formation of couples. I am only considering the marriage market of working people composed of working singles and bi-working couples. However, married women's participation to the labor market has increased from 1991 to 2001 and my samples become more and more representative overtime. Since 1999, more than 75 % of married women between 22 and 40 years old are working and more than 90 % of men (These figures are presented in appendix on Figure 16). Second, I don't model the evolution of wages with age. Married individuals who are older in average than singles have then higher wages. This could lead to overestimate the attractiveness of high wage men on the marriage market. To limit the bias, I first restrict the sample to the age range between 22 and 40 years old⁹, second I define wages as the individual prediction of one's wage at 31 year old.

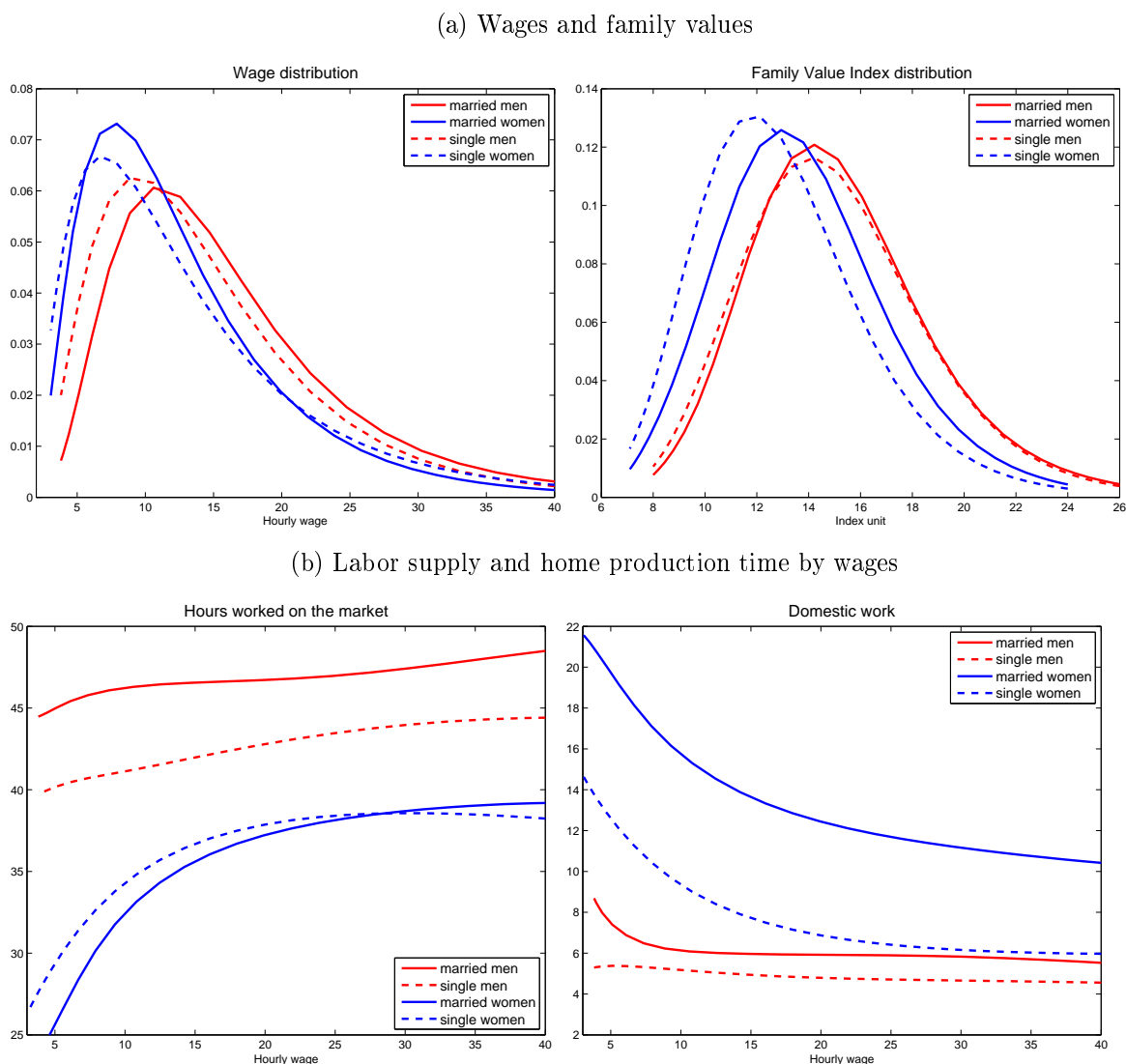
3.2 Wages and family values by gender and marital status

The left panel of Figure 1a represents the wage distribution for different marital status in 1999. Married people have higher wages, particularly men. The right panel represents the distribution of the Family Value Index. Men are more conservative than women and married women are more conservative than single women. In 1999, wage correlation among couples was around 0.32 and F.V.I correlation is around 0.44.

Figure 1b provides information on the link between wages and the amount of time spent in paid work and home production. Female labor supply is more wage-elastic than male labor supply. Married men work on average 3 hours more, in any given week, than single men, while married women work less than single women by about 2 hours. Conversely, married women work more hours at home than single women by about 5 hours per week. Moreover, female domestic time use steeply decreases with wage, whereas it is rather inelastic for men, with an average value of 6 hours per week.

⁹In my sample, married men are in average only 3 years older than single men. Married women are in average only 1.5 years older.

Figure 1: Distribution of wages, family values and time use. BHPS, 1999

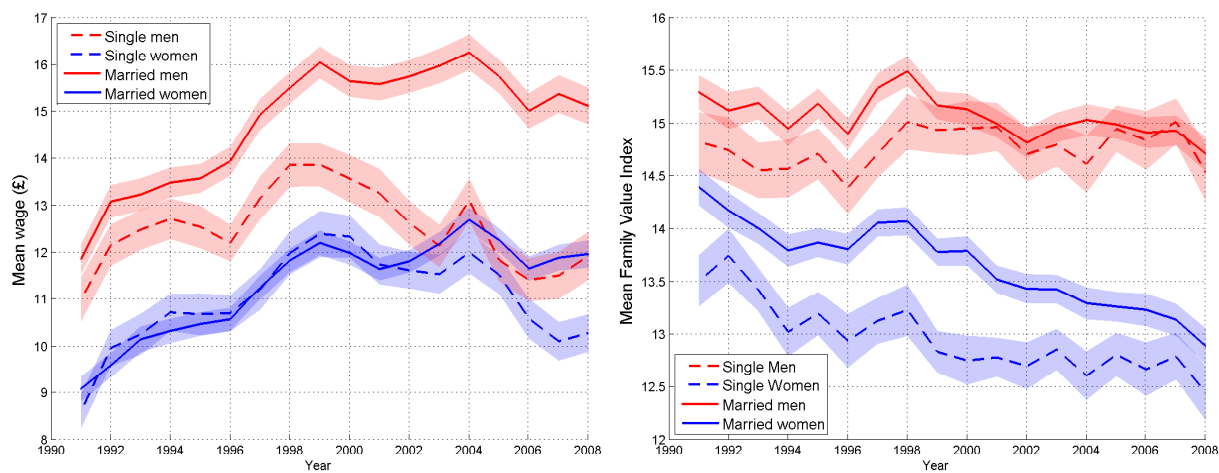


I have plotted in Figure 2a the trends in wages and FVIs for different marital statuses (the shaded area is the 90% confidence interval). The average wage of married men has increased from £12 in 2001 to £15 in 2008. Married women's wages have seen a parallel increase from £9 to £12.1. These regular increases contrast with the relative stability of single men's and single women's wages over the period. FVI trends also reveal sharp gender differences. Men are not only more conservative than women as far as family values are concerned, but the tendency for women over the period is towards more liberalism, whereas men's attitudes are stable by comparison. Next, we turn to the evolution of time use over the observation period (1991-2008). There is a clear tendency for single men to work less (a 4-hour reduction; see Figure 2b), that is also observable for married men, albeit to a lesser extent. In contrast, female labor supply remains constant over the period, while married women tend to work more (from about 30 hours per week in 1992 to 32.5 in 1999). The amount of time spent in home production follows a general decreasing trend for all categories of individuals, married or single, male or female (see Figure 2c). However, this negative trend is particularly pronounced for married women and mothers. They used to spend in housework 20 hours per week on average in 1991. In 2008, housework takes no more than 13

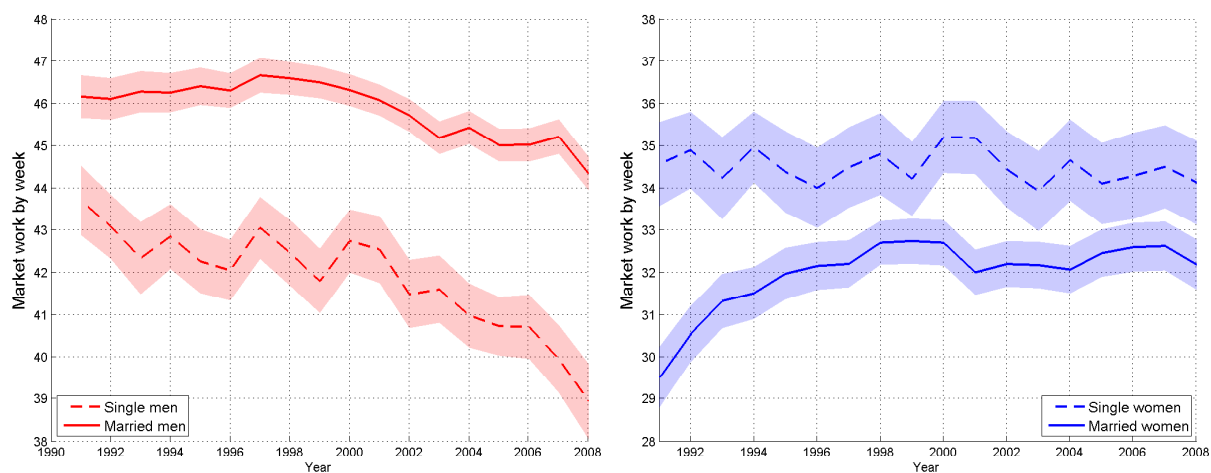
hours on average of women's time. These evolutions raise an interesting challenge for the model to explain. We shall ask the question of how much of the changes in time uses can be explained by exogenous changes to wages and family values, and which structural changes (i.e. changes to the parameters of the model) are necessary to account for the observed changes.

Figure 2: Evolution of wages, family values and time use. BHPS, 1991-2008

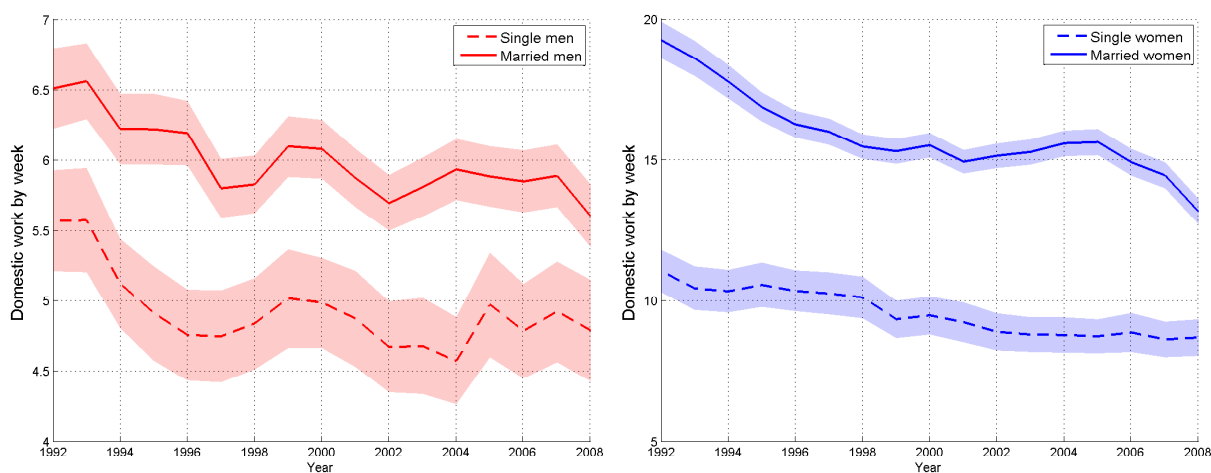
(a) Wages and family values



(b) Labor supply



(c) Home production



4 Specification and estimation

In this section I specify the utility functions and the production functions, and I develop the estimation procedure.

4.1 Parametric specification

Indirect utility and home production

For simplicity I assume as in [Jacquemet and Robin \(2013\)](#) that the indirect utility function has the following quasi-linear form

$$v_i(R, Q) = Q \frac{R - A_i}{B_i}$$

and that the domestic production functions are Cobb-Douglas:

$$\begin{aligned} [\text{couples}] \quad Q &= F_{ij}^1(d_i, d_j, z) = (\Phi_{ij} + z)(d_i - D_i^1)^{K_i^1} (d_j - D_j^1)^{K_j^1} \\ [\text{singles}] \quad Q &= F_i^0(d) = (d - D_i^0)^{K_i^0}. \end{aligned}$$

There is no multiplicative constant in front of $F_i^0(d)$ as it can be subsumed into parameter B_i . For singles, the maximization of their program leads to the following different expressions. Domestic time use d_i^0 is such that

$$w_i(d_i^0 - D_i^0) = \frac{K_i^0}{1 + K_i^0} (w_i(T - D_i^0) - A_i). \quad (5)$$

Leisure is

$$l_i^0 = A_i' + \frac{B_i'}{B_i} (w_i(T - d_i^0) - A_i) = A_i' + \frac{B_i' w_i(T - D_i^0) - A_i}{B_i (1 + K_i^0)}, \quad (6)$$

where A_i' and B_i' denote the derivatives of A_i and B_i with respect to w_i , the market wage of individual i . Labor supply is $h_i^0 = T - d_i^0 - l_i^0$.

Using equation 1 and 3, the present value of being single is

$$B_i r V_i^0 = \underbrace{\left(\frac{K_i^0}{w_i} \right)^{K_i^0} \left(\frac{w_i(T - D_i^0) - A_i}{1 + K_i^0} \right)^{1+K_i^0}}_{v_i^{0*}} + \beta_i \frac{\lambda}{r + \lambda_z} \iint S_{ij}(z)^+ dG(z) dU_f(z) \quad (7)$$

where $\beta_i = \beta$ or $1 - \beta$ according to i 's gender (male or female).

For married individuals, domestic time uses are d_i^1 and d_j^1 such that

$$\frac{w_i(d_i^1 - D_i^1)}{K_i^1} = \frac{w_j(d_j^1 - D_j^1)}{K_j^1} = R_{ij} \quad (8)$$

for

$$R_{ij} = \frac{w_i(T - D_i^1) + w_j(T - D_j^1) - C_{ij} - A_i - A_j}{1 + K_i^1 + K_j^1} \quad (9)$$

Leisure demands are

$$l_i^1 = A'_i + \frac{B'_i}{B_i}(w_i(T - d_i^1) + t_i - A_i), \quad l_j^1 = A'_j + \frac{B'_j}{B_j}(w_j(T - d_j^1) + t_j - A_j) \quad (10)$$

After some algebra detailed in appendix, we find that marriage is consummated if $S_{ij}(z) > 0$ where $S_{ij}(z)$ solves

$$S_{ij}(z) = \sigma_{ij}(s_{ij} + z), \quad (11)$$

with

$$\sigma_{ij} = \left(\frac{K_i^1}{w_i}\right)^{K_i^1} \left(\frac{K_j^1}{w_j}\right)^{K_j^1} R_{ij}^{1+K_i^1+K_j^1} (1 + K_i^1 + K_j^1) \quad (12)$$

and

$$s_{ij} = \Phi_{ij} - \frac{B_i V_i^0 + B_j V_j^0}{\sigma_{ij}} + \frac{\lambda_z}{r + \lambda_z} \int \max(s_{ij} + z', 0) dG(z'), \quad (13)$$

Transfers t_i and t_j are such that

$$w_i(T - d_i^1) + t_i - A_i = \frac{B_i r V_i^0}{Q} + \beta \left(R_{ij} - \frac{B_i r V_i^0}{Q} - \frac{B_j r V_j^0}{Q} \right), \quad (14)$$

and $t_j = -C_{ij} - t_i$

Heterogeneity

The dependency of the parameters to the exogenous variables (gender, wage and family values index) is specified as follows. First, the indirect utility for consumption and leisure is a smooth function of gender and wage with $A_i = A_i(g_i, w_i)$ and $B_i = B_i(g_i, w_i)$. Second, the domestic production functions vary with gender and family values, i.e.

$$\begin{aligned} D_i^0 &= D_i^1 = D_m, & K_i^0 &= K_i^1 = K_m, \\ D_j^0 &= D_f^0, & K_j^0 &= K_f^0 \\ D_j^1 &= D_f^1, & K_j^1 &= K_f^1 e^{\kappa_f^1 f_j} \end{aligned} \quad (15)$$

where f_i denotes the family values index for individual i . In addition I set C_{ij} equal to a constant: $C_{ij} \equiv C$. We observe in data on Figure 17 that domestic work of married women is elastic to their FVI¹⁰. To take this into account I assume that the domestic work elasticity of married women depends on her FVI. This flexibility has been chosen to match the observed domestic

¹⁰Data also show that domestic work of married women is elastic to their husband's FVI in a lesser extent.

work of married women who work much more at home than single women whereas domestic work of married men and single men are very similar. Finally Φ_{ij} is a smooth function of both spouses' wages and family values indexes: $\Phi_{ij} \equiv \Phi(w_i, f_i, w_j, f_j)$, where f_i, f_j denote the family values indexes of male spouse i and female spouse j . I thus use family values as an instrument or a measurement for the marriage public good above and beyond the externality generated by pooling time resources. Φ_{ij} will be named the affinity factor.

4.2 Identification and Estimation strategy

Estimation of λ and λ_z with a duration model

As in Wong (2003), we estimate "hazard rates" λ and λ_z by applying maximum likelihood to marriage/singlehood duration data. Agents can be single or married at the time of the interview. Let T_{0b} and T_{0f} be the elapsed and residual singlehood durations for those individuals who are single at the time of the interviews. Let C_{0b} and C_{0f} be corresponding censoring indicators. Let $(T_{1b}, T_{1f}, C_{1b}, C_{1f})$ be similar notations for marriage durations. We observe complete durations for about 10% of couples and 13% of singles. The mean marriage duration calculated over uncensored durations is equal to 13 years whereas the mean duration of singlehood is equal to 7 years. For any single, elapsed and residual durations are independent and have the same exponential distribution. Thus for a male single of type i , the individual contribution is

$$L_{0i} = \left(\lambda \int_j a(i, j) U_f(j) \right)^{1-C_{0b}+1-C_{0f}} e^{-\left(\lambda \int_j a(i, j) U_f(j) \right) (T_{0b}+T_{0f})} u_f(j)^{(1-C_{0f})}$$

Events occurring after exit from being single are independent of the events up to exit. Therefore, their probability is independent of the likelihood of being single. The event immediately following type i 's singlehood duration is the realization of whom to match with. This event is given by the density of accepted type $u_f(j)$. We define similarly an individual contribution for female singles.

For married couples of type (i, j) , the contribution of $(T_{1b}, T_{1f}, C_{1b}, C_{1f})$ to the likelihood is

$$L_{1ij} = (\lambda_z(1 - a(i, j)))^{1-C_{1b}+1-C_{1f}} e^{-\lambda_z(1-a(i, j))(T_{1b}+T_{1f})},$$

where $a(i, j)$ can be substituted out using the steady-state flow restriction 4 as

$$a(i, j) = \frac{N(i, j)}{\frac{\lambda}{\lambda_z} U_m(i) U_f(j) + N(i, j)}.$$

Densities $n(i, j)$ and $u_m(i), u_f(j)$ are estimated using Kernel estimators.

Estimation of domestic production parameters

I estimate the domestic production parameters using Non Linear Least Squares estimation simultaneously on the equations 5, 6, 8 and 9.

Preference functions : inference from hours

The identification of A_i and A_j follows from the identification of B_i and B_j and the parameters of domestic time use. Using Roy's identity, working hours of single and married people can be expressed

$$\begin{aligned} h_i^1 &= T - d_i^1 - A'_i(w_i) - \frac{B'_i(w_i)}{B_i(w_i)}(w_i(T - d_i^1) + t_i - A_i(w_i)) \\ h_i^0 &= T - d_i^0 - A'_i(w_i) - \frac{B'_i(w_i)}{B_i(w_i)}(w_i(T - d_i^0) - A_i(w_i)). \end{aligned} \quad (16)$$

Using the equation for singles, you obtain the following linear differential equation :

$$A'_i(w_i) - \frac{B'_i(w_i)}{B_i(w_i)}A_i(w_i) = T - h_i^0 - d_i^0 - \frac{B'_i(w_i)}{B_i(w_i)}w_i(T - d_i^0)$$

whose solution is

$$A_i(w_i) = B_i(w_i) \int_0^{w_i} \frac{T - h_i^0 - d_i^0 - \frac{B'_i(w)}{B_i(w)}w(T - d_i^0)}{P_i(w)} dw.$$

Then if the aggregate price indexes B_i and B_j are known, we can recover the functions $A_i(w_i)$ and $A_j(w_j)$. Besides, using (16), we write :

$$h_i^1 - h_i^0 = d_i^0 - d_i^1 - \frac{B'_i(w_i)}{B_i(w_i)}(t_i + w_i(d_i^0 - d_i^1)). \quad (17)$$

Then integrating the preceding equation, we get

$$\int_{z|z > -Sxy} (h_i^1 + d_i^1 - h_i^0 - d_i^0) dG(z) = -\frac{B'_i(w_i)}{B_i(w_i)} \left(\int_{z|z > -Sxy} t_i + w_i(d_i^0 - d_i^1) dG(z) \right).$$

We will consequently regress the ratio $\Delta \overline{H}_i = \overline{h}_i^1|z > -Sxy + d_i^1 - h_i^0 - d_i^0$ on $\overline{t}_i|z > -Sxy + w_i(d_i^0 - d_i^1)$ to obtain

$$\frac{B'_i(w_i)}{B_i(w_i)} = -\frac{\int_j (\Delta \overline{H}_i) (\overline{t}_i + w_i(d_i^0 - d_i^1)) n(i, j) dj}{\int_j (\overline{t}_i + w_i(d_i^0 - d_i^1))^2 n(i, j) dj}. \quad (18)$$

Then B_i can be recovered using transfers and the observation of domestic and market work. However, transfers are also function of A_i and A_j . Besides, the domestic work parameters are also estimated using A_i and A_j . These functions are then solved by iterations. Using initial values for these functions we estimate the parameters for domestic production, compute transfers, then B_i and B_j using (18) and we estimate new functions for A_i and A_j until convergence.

Bargaining power and distribution of z

The distribution of the match quality shock $G(\cdot)$ is modeled with a centered Gaussian distribution of variance σ_z^2 . I show in appendix 6 the impact of σ_z on single present value and affinity matrix.

I first show that single present values are linear expressions of σ_z . A high σ_z means an important contribution of the continuation value in the single present value. There exist two functions H_m and H_f such that

$$\begin{aligned} rB_iV_i^0 &= B_iv_i + \sigma_z H_m(a(i, j), \sigma_{ij}, u_f) \\ rB_jV_j^0 &= B_jv_j + \sigma_z H_f(a(i, j), \sigma_{ij}, u_m). \end{aligned}$$

Furthermore, there exist a function H such that the affinity matrix also depends on σ_z in the following way.

$$\Phi(i, j) = \frac{B_jv_j + B_iv_i}{\sigma_{ij}} + \sigma_z H(a(i, j), \sigma_{ij}, u_f, u_m).$$

The parameters σ_z and β are estimated by minimizing the errors in market hours prediction for married men and women with the whole model.

5 Estimation results

There are 12 parameters and 6 functions to estimate in the model. which are displayed in table 2.

Table 2: Parameters of the model

Discount rate	r
Quality shocks	$G(z), \lambda_z$
Meeting parameters	λ
Bargaining parameters	β
Domestic production	$K_m, K_f^0, K_f^1, \kappa_f^1, D_m, D_f^0, D_f^1$
Preference functions	$B_i(w_i), B_j(w_j), A_i(w_i), A_j(w_j)$
Affinity matrix	$\Phi(i, j)$
Cost	C

All parameters can't be estimated. The discount rate r is set at 3 % per year. The distribution of the match quality shock $G(\cdot)$ is modeled with a centered Gaussian distribution of variance σ_z^2 set to 0.12 to fit the best the data. The parameters λ and λ_z are estimated independently using the exponential duration model previously described. The domestic production parameters and the preference functions are estimated together as described in the previous section. I fix the cost C at 800 £ a month which is the minimum required to obtain realizable predictions of market hours. As 65 % of married couples in my sample have children, this must represent an average additional cost supported by parents to raise children¹¹. The bargaining parameter which fit the best the data is $\beta = 0.74$ in 1999.

Table 3 presents the parameter estimates in 1999 where the sample is the largest. The parameters of quality shocks and meeting give an average duration of singlehood of 10 years for single men

¹¹However, I should also take into account that many single women also incur an additional cost for raising children. 30 % of single women in my sample have children, which concerns less than 2 % of men.

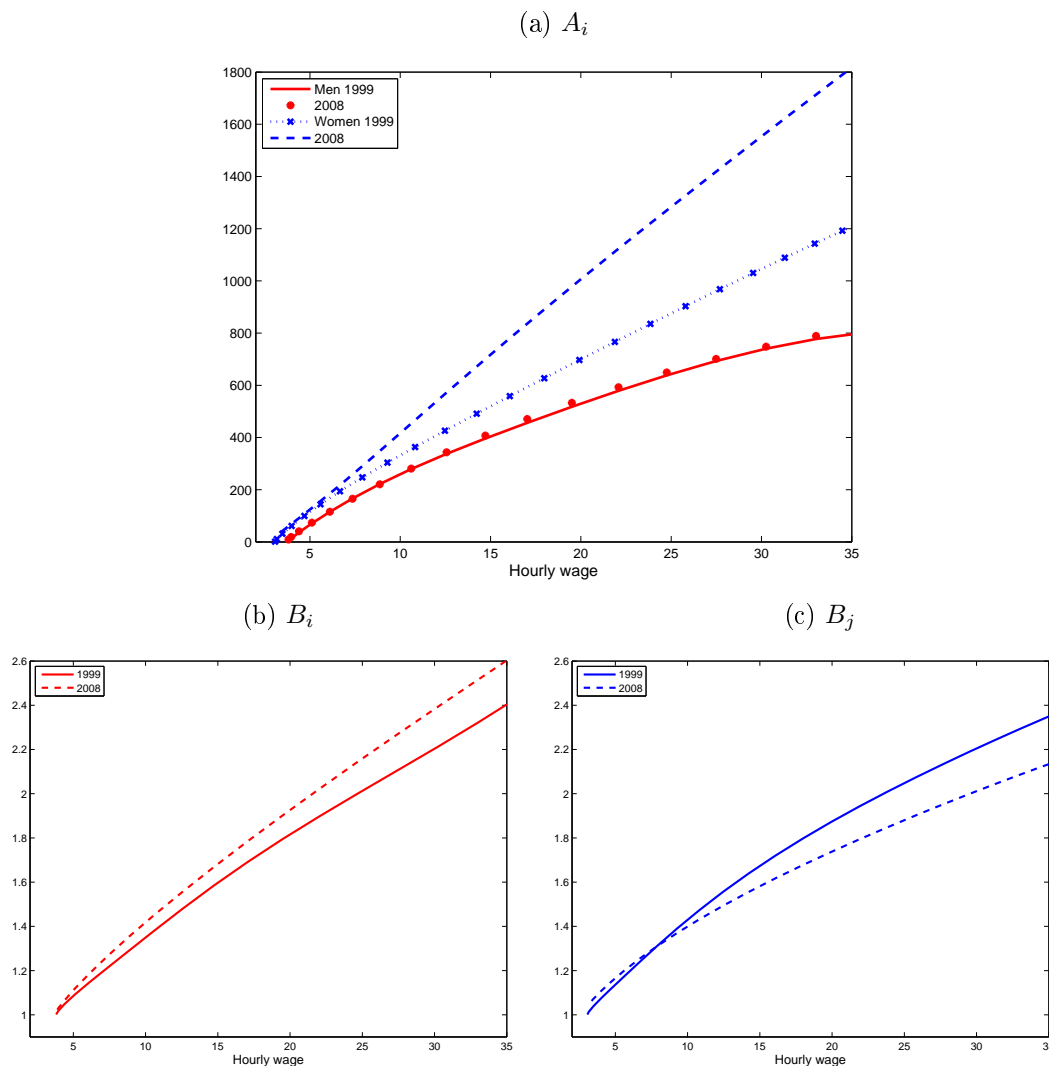
and 11 years for single women and a average duration of couples of 30 years.

Table 3: Parameters estimates

Quality shocks ^(a) λ_z	Meeting parameters ^(a) λ	Domestic production parameters						
		D_m	D_f^0	D_f^1	K_f^1	κ_f^1	K_f^0	K_m
0.0028 (0.00020)	0.00071 (0.0003)	4.04 (0.25)	4.25 (0.62)	8.01 (1.37)	0.017 (0.0084)	0.046 (0.011)	0.072 (0.009)	0.013 (0.0022)

The preference functions are represented on Figure 3. The upper panel represents the minimal amount of consumption and shows that it increases almost linearly with wages for both men and women. The slope is higher for women. The lower panel of Figure 3 represents aggregate price index. Preference for leisure increases with wage and is higher for men than for women. Figure 3 also represents the evolution of preferences for consumption and leisure over years. Men have decreased their minimum level of consumption and leisure and have increased their preferences for leisure relative to consumption. This results is consistent with the observation of Aguiar and Hurst (2007) who document a sharp increasing trend in men leisure over the last four decades which is not explained by changes in wages. On the contrary, women have increased their minimum level of consumption and leisure and have decreased their preference for leisure relative to consumption.

Figure 3: Evolution of preferences for consumption and leisure



The next two subsections will present the estimation of matching patterns, within household transfers and predictions of labor supply. Section 5.1 will presents the results obtained in 1999 with the whole multidimensional model. Section 5.2 will present how matching patterns and transfers have evolved over the 1999-2008 period focusing on a unidimensional model where individuals only match on wages.

5.1 Cross section analysis in 1999

Matching patterns

This section presents the estimation of the probability that a man of type $(wage_i, f_i)$ and a woman of type $(wage_j, f_j)$ match if they happen to meet. I represent this four variable function on a 3D graph. Figure 4 shows the expected match probability conditional on wages. The left panel is a 3D plot whereas the right panel represents the level curves. The left panel shows that the matching probability is strongly increasing in both wages. The probability that a man with a wage rate of 25£ matches with a woman of wage rate 20£ when he meets her is 0.04 whereas the probability that he marries a woman of wage rate 5£ when he meets her is 0.02. This figure

also shows a little dissymmetry more visible on the right panel. Women with low wages have higher chances to marry than men with low wage. Conditional on meeting, the probability that a rich man marries a low wage woman is higher than the probability that a rich woman marries a poor man. Figure 5 represents the match probability conditional on the Family Value Index. The expected probability of matching is lower in average meaning that the FVI explains less of the matching probability. The right panel shows that the matching probability is higher when the two FVI are high and close. The matching probability is very low for high FVI male with low FVI female.

Figure 4: Expected Marriage probability conditional on wages

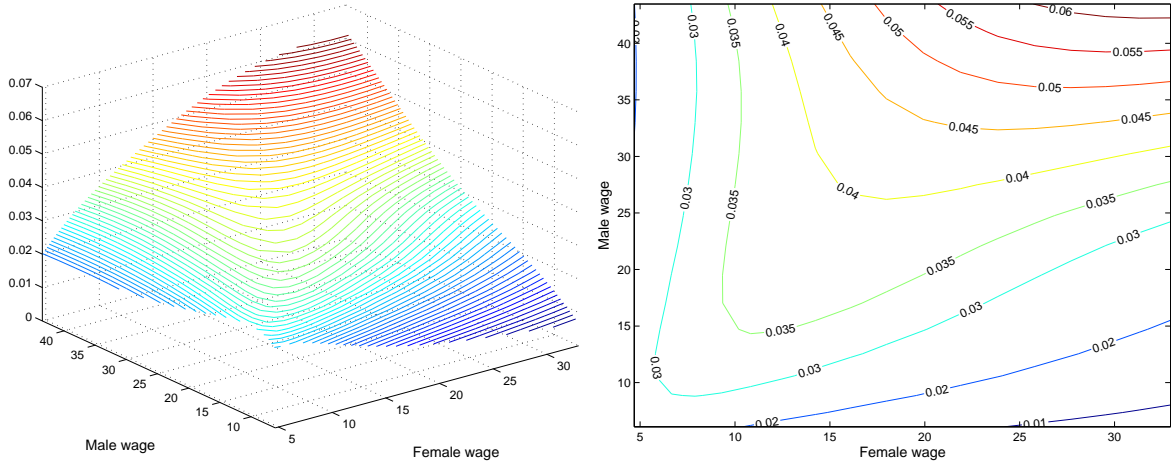
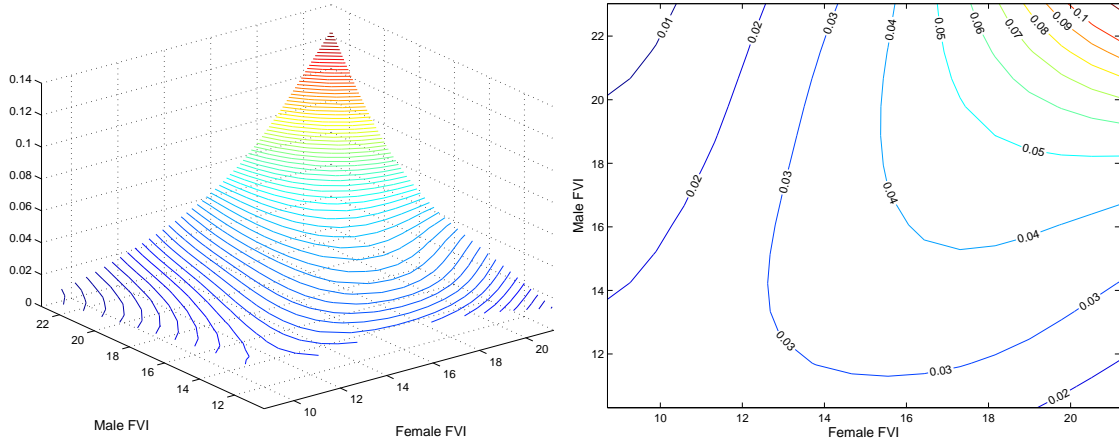


Figure 5: Expected Marriage probability conditional on Family Value Index



I represent the affinity factor Φ_{ij} conditional on wages on Figure 6. Its shape is different from the total surplus. Low wage women have affinity with low wage men. It is not the higher the wage the better anymore. Same wage partners have high complementarities in public good production, particularly low wage individuals. Moreover, we see that high wage women have very low complementarities with low wage men. When we look at the affinity factor conditional on the Family Value Index on Figure 7, the best match is reached for couples with more traditional family values. It is very low for low FVI female matched with high FVI men.

Figure 6: Affinity factor conditional on wages

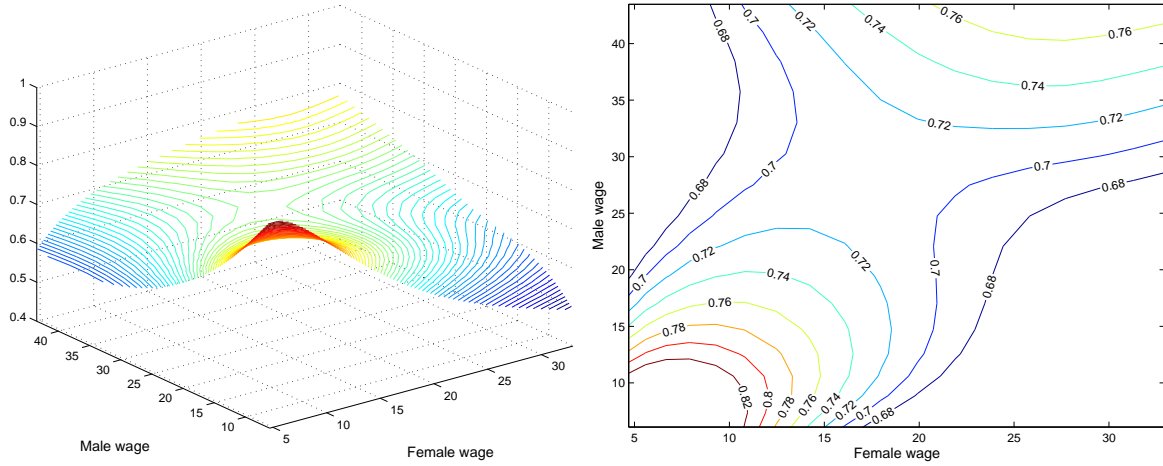
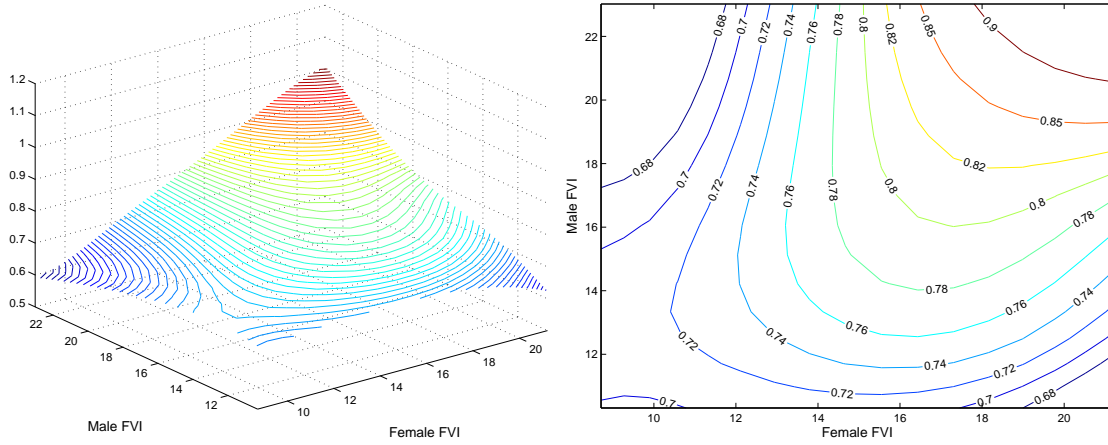


Figure 7: Affinity factor conditional on Family Value Index



The previous graphs show the shape of the affinity factor with respect to two out of the four characteristics : w_i, w_j, f_i, f_j . It would be interesting to better understand the contribution of each variable to this factor and their interaction. To this end, I perform a non-parametric regression of the affinity factor using Generalized Additive Models (GAMs). I present in table 4 the generalized R^2 obtained for different specifications of the interactions of variables (I put a note on GAM in appendix).

There exist strong complementarities between characteristics of the partners. The public good is not only a sum of the contributions of each partner. For instance, partner's wages plus their interaction account for 61.8% of the variance whereas the sum of each wage contribution would account for only 21.8%. Similarly, the partners' FVI plus their interaction account for 25.5% whereas the sum of their individual contribution is only 8.4%. However, there is apparently no interaction between the other characteristics (including some interaction doesn't add more explanation power).

The characteristics of the man contribute more to the affinity factor. His wage explains by itself 19% of the variance. The interaction of the partners wage plus the interactions of the partners FVI account for almost all the variance : 87.3%. The interaction of the male characteristics plus

the interaction of the women characteristics only account for 33% of the variance.

Table 4: Comparisons of different models

Variables	Generalized R^2
(w_m)	18.9%
(f_m)	6.2%
(w_f)	2.9%
(f_f)	2.2%
(w_m, f_m)	27.3 %
(w_f, f_f)	5.7 %
(w_m, w_f)	61.8 %
(f_m, f_f)	25.5 %
(w_m, f_f)	21.3 %
(w_f, f_m)	9.9 %
(w_m, w_f, f_m, f_f)	93.6 %

Transfers and Inequalities

When married, the individual's resource is changed by a transfer from his spouse (positive or negative). For now, we considered the generalized sharing rule t_i and t_j . Usually, collective models with domestic production are interested in the conditional sharing rule $t_i - w_i d_i^1$ and $t_j - w_j d_j^1$. However, this rule neglects public production. It doesn't represent how individuals contribute to domestic production. A woman can get a large part of the rest of the total income, but she may have also contributed much more to domestic production than her husband and she finally doesn't benefit so much from the couple surplus. The generalized sharing rule takes individual domestic work into account but neglects price of public consumption. Indeed, men and women could have different marginal propensity to spend time and money in public production, then it could be less costly for a woman to spend more time in housework than for a man. We can consider instead the measure developed in Chiappori, Meghir (2013), the Money Metric Welfare Index (henceforth MMWI) which corresponds to the monetary amount that one would need to reach alone the same utility level that she reaches when she is in couple. Chiappori and Meghir (2013) argue that the Money Metric Welfare Index fully characterizes the utility level reached by the agent. That is the MMWI is the amount M such that if we give M to a single individual of type i , he would reach $v_i^{0*} = v_{i|j}^1$ that is

$$F_i^0(d) \frac{w_i(T - d_i^0) + M - A_i}{B_i} = F_{ijz}^1(d_i, d_j) \frac{w_i(T - d_i^1) + t_i - A_i}{B_i}.$$

To sum up, I have available three instruments to measure within household inequalities, the generalized sharing rule, the conditional sharing rule and the MMWI. Using Discrete Cosine Transform, I compute the three measures of transfers for each men and women of each couple. In 1999, the median of the MMWI's share of the woman $\left(\frac{M_f}{M_f + M_m}\right)^{12}$ is 0.32 that is 50% of married women get less than 32% of the welfare surplus generated by the couple. The median of the woman conditional transfer is - 752£ which is the share of the children cost (here 800£)

¹²I transform the variable to obtain a strictly positive measure. For each household g , I use the value $M'_{mg} = M_{mg} - \min(0, M_{mg}, M_{fg})$ and $M'_{fg} = M_{fg} - \min(0, M_{mg}, M_{fg})$

supported by the woman plus the total opportunity cost of her domestic work. The median of the woman generalized transfer is equal to - 237£ which is only the share of the children cost supported by the woman.

Table 5: Median value of transfers to the women in 1999

MMWI woman's share ($\frac{M_f}{M_m+M_f}$)	0.25
Generalized Sharing Rule	-9 £
Conditional Sharing Rule	-700 £

It is more interesting to go beyond the median and to study these transfers for particular types of household. I classify each partner in 3 categories: those whose wages are below the first quartile ($w \leq q_{25}$), those whose wages are between the first quartile and the third quartile ($w \in [q_{25} - q_{75}]$) and those whose wages are over the third quartile ($w > q_{75}$). We obtain 9 different types of households. The average of transfers among each type of household are presented in table 6. A wife gets a higher share of the welfare when her husband has a lower wage than her. Her welfare share also increases with her wage. Things are different for monetary transfers. Higher her wage and lower her husband wage, higher she contributes to the public good and lower is her resulting transfer. A women in the top 25 % married with a man also in the top 25 % will get 7£ a month whereas if she's in the bottom 25 % she will get 266£.

Table 6: Average values of transfers to the women according to the household type

Woman Wage Quantile	$\leq q_{25}$			$[q_{25} - q_{75}]$			$> q_{75}$		
Man Wage Quantile	$\leq q_{25}$	$[q_{25} - q_{75}]$	$> q_{75}$	$\leq q_{25}$	$[q_{25} - q_{75}]$	$> q_{75}$	$\leq q_{25}$	$[q_{25} - q_{75}]$	$> q_{75}$
$(\frac{M_f}{M_m+M_f})$	0.33	0.24	0.18	0.35	0.27	0.18	0.34	0.27	0.20
Conditional Sh.Rule (£)	-643	-480	-281	-1018	-833	-593	-1541	-1398	-1133
Generalized Sh.Rule (£)	-112	54	266	-226	-61	181	-427	-269	7

To obtain a complete characterization of these transfers, I regress them on individuals characteristics. Remind the formula for the generalized transfer to the woman

$$t_j = -(w_j(T - d_j^i) - A_j) + (1 - \beta)R_{ij} + \left(\beta \frac{B_j r V_j^0}{Q} - (1 - \beta) \frac{B_i r V_i^0}{Q} \right).$$

Even if we have the exact formula for all these transfers, the impact of different variables is not obvious in the expression $\beta \frac{B_j r V_j^0}{Q} - (1 - \beta) \frac{B_i r V_i^0}{Q}$ which corresponds to a bargaining term due to outside options. To separate the effect of outside options from the effect of total resources, I also regress the outside option bargaining term on characteristics. Regression results are presented in table 7. The first column displays the results for the average welfare share of women that is $\frac{M_f}{M_f+M_m}$, the second column displays the results for the generalized transfer. The third column corresponds to the outside option term.

All things being equal, the outside option term increases in woman's wage and decreases in her husband wage. This results show that outside option depends positively on wages. It also increases in the woman FVI but decreases with the man FVI. Then FVI have also positive impact

on outside option for both men and women (remind that high FVI people are more attractive)¹³. Let us consider an average couple¹⁴. An increase of 1£ in man's wage leads to a decrease of 0.9% in her welfare share and an increase of 26£ in her transfer. Actually, the increase in man's wage has lowered the outside option bargaining term by 68£ but as it has also increased the total resources by 94£.

Table 7: Determinants of the MMWI and the conditional sharing rule in 1999

Variables	Woman share of MMWI (%)	Sharing rule t_f (£/mth)	$\beta \frac{B_j r V_j^0}{Q} - (1 - \beta) \frac{B_i r V_i^0}{Q}$ (£/mth)
Constant	21.97 (1.64)	-514 (36)	-780 (104)
Wage male	-1.53 (0.03)	35.3 (0.65)	-35.1 (1.86)
Wage male square	0.019 (0.0006)	-0.28 (0.014)	-1.02 (0.04)
Wage female	0.77 (0.039)	-18.42 (0.88)	152.4 (2.51)
Wage female square	-0.016 (0.0012)	-0.10 (0.026)	1.50 (0.07)
FVI male	-1.60 (0.16)	32.6 (3.7)	-66.1 (10.5)
FVI male square	0.05 (0.0005)	-0.79 (0.11)	1.85 (0.32)
FVI female	3.1 (0.17)	-17.3 (3.78)	169.2 (10.8)
FVI female square	-0.07 (0.006)	1.08 (0.13)	-6.3 (0.37)
R^2	86 %	93 %	98 %

Prediction of hours

Using the model, I compute the predicted working hours conditional on wages for married people. Figure 8 shows a very good prediction of conditional market hours for both men and women. I also compute what would be the working hours of individuals in two extreme cases. The first case is when there is no possibility of transfers. Then married individuals still benefit from complementarity in domestic production but each member keeps his own labor income and pays the half of the children cost. The second case is when married individuals share equally all their resources. More precisely the expression of men labor supply in these three cases are

¹³ An increase in FVI of the man leads to an increase in total resource and an increase in her domestic work so increases the transfer to the woman. This has been ruling out by the model but I use the observed working hours and domestic hours to compute the transfers. That's why the male FVI has a different impact on the outside option bargaining term and on the transfer.

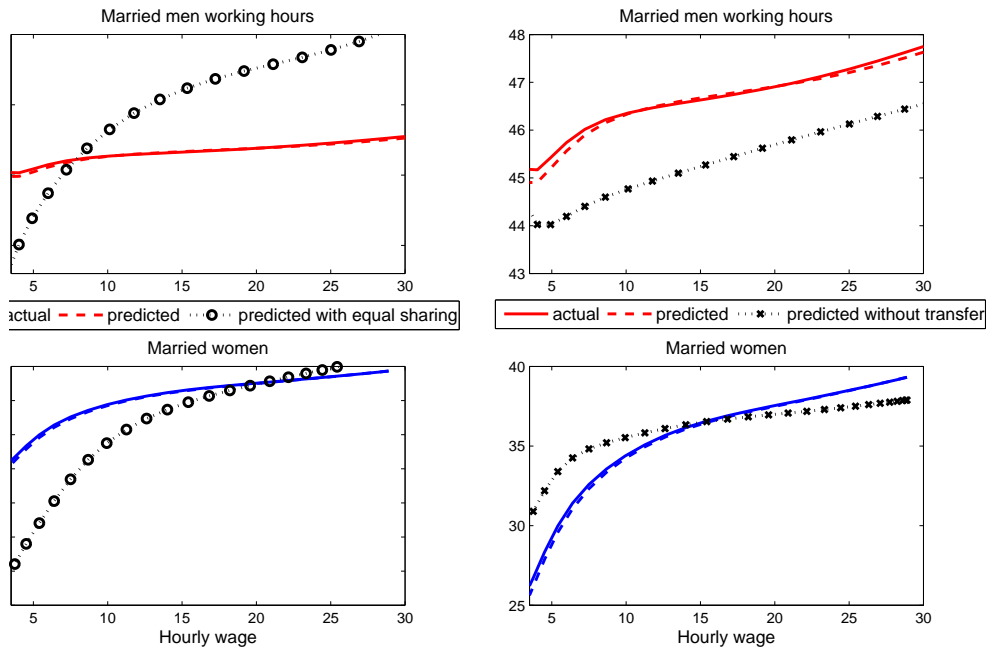
¹⁴ The man would have an FVI of 15.2 and an hourly wage of 16£ and the woman an hourly wage of 12.2£ and a FVI index of 13.8

$$\begin{aligned}
\text{Labor supply with transfers} \quad h_i^1 &= T - d_i^1 - A_i' - \frac{B_i'}{B_i}(w_i(T - d_i^1) + t_i - A_i) \\
\text{Labor supply without transfers} \quad h_i^1 &= T - d_i - A_i' - \frac{B_i'}{B_i}(w_i(T - d_i^1) - A_i - \frac{C}{2}) \\
\text{Labor supply with equal sharing} \quad h_i^1 &= T - d_i^1 - A_i' - \frac{B_i'}{B_i} \frac{w_i(T - d_i^1) + w_j(T - d_j^1) - A_i - A_j - C}{2}
\end{aligned}$$

In the last two equations, there is no “bargaining effects”. Labor supplies depend on the standard income effects and substitution effects. When man’s wage rises, his labor supply tends to increase through substitution effects with the decrease in $\frac{B_i'}{B_i}$ and the decrease in domestic work d_i^1 (and the increase in his wife’s domestic work for the model with equal sharing). It tends to decrease through the income effect with the rise in $w_i T - A_i$. In the first equation, there is also a “bargaining effect” which acts like an income effect through changes in the transfer t_i due to changes in wage.

Figure 8 is decomposed in four parts, the two upper parts represent labor supply of men and the two lower parts labor supply of women. The two left parts compare labor supply predicted by the model with labor supply predicted in the equal sharing case. The two right parts compare labor supply predicted by the model with labor supply predicted in the no transfer case. Resource sharing mostly benefit to low wage women who can work much less as if they didn’t get any transfer. On the contrary, married men should work less than they do to compensate the transfer they give to their wife. If individuals shared equally their resources, as men have generally higher wages, they would work much more as they would give more than half of their resources to their wife. On the contrary, women would work much less. The actual working hours lie between the two extreme cases.

Figure 8: Prediction of working hours in 1999



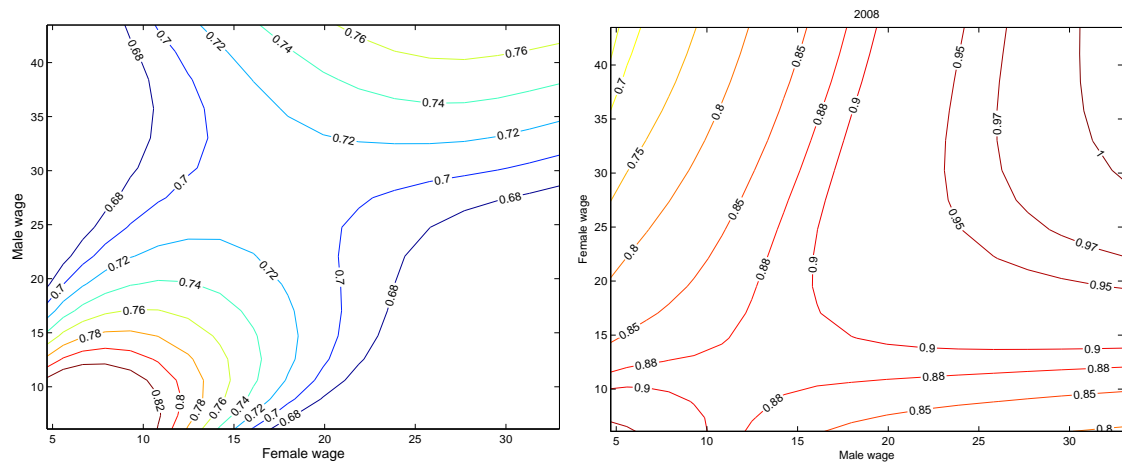
5.2 Evolution from 1991 to 2008

In this subsection I present how matching patterns and transfers have evolved over time.

Matching patterns

Figure 9 represents the evolution of the affinity matrix $\Phi(i, j)$ with respect to wages overtime. We observe that high wage women become more and more attractive whereas high wage men become less attractive. We always observe high complementarities on the diagonal which has moved to the right (toward high wage women) in 2008.

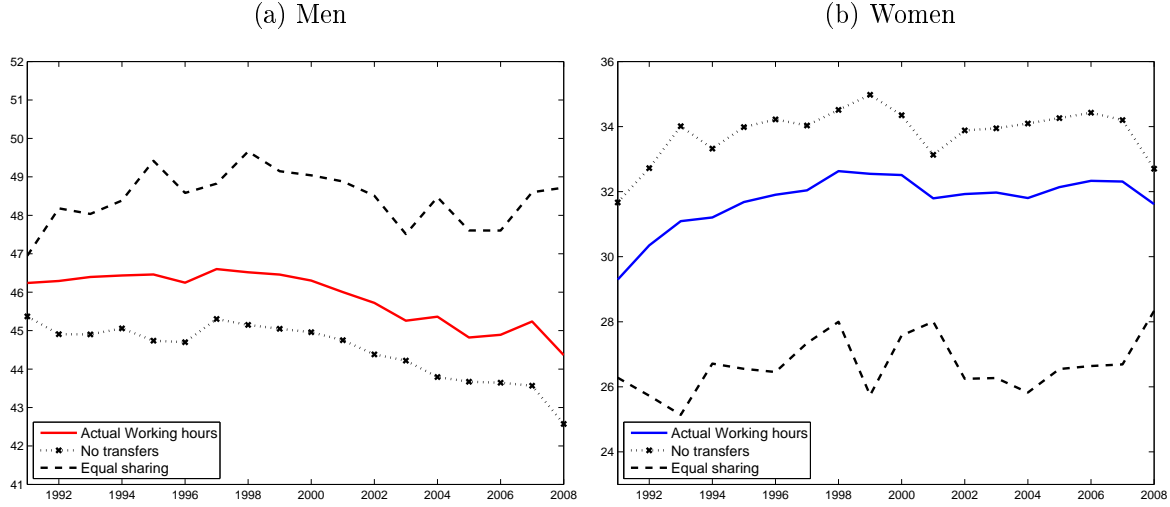
Figure 9: Evolution of complementarities in wages from 1999 to 2008



Prediction of hours

In a similar way than in section 5.1, I compute the predicted hours for each individual of each sample and the predicted hours with (i) no within household transfers and (ii) equal sharing. I present the evolution of the average of working hours by marital status on Figure 10. Without transfers, men would work 1 hour less in average (about 2.5 % less). To the contrary, married women would work more by 2 hours (about 4.5 % more). This gap seems constant overtime for both men and women. If there was equal sharing, men would have worked 2 hours more from 1991 to 1998, 3 hours more until 2006, then 4 hours more. On the opposite, with equal sharing, women would have worked 4 hours less from 1991 to 2002 then 4 hours from 2002 to 2007.

Figure 10: Evolution of working hours



6 Simulations

6.1 Characterization of the equilibrium

In this section I compute the equilibrium distribution of characteristics and match probabilities from the previous non-parametric estimates of the structural parameters, namely, the complementarities in characteristics $\Phi(i, j)$ and the preference parameters. Let $n(i)$ denotes the density of married men of type i , then $n(i) = \int_j n(i, j) dj$. Besides, remind that

$$N(i, j) = \frac{\lambda U_m(i) U_f(j) a(i, j)}{\lambda_z (1 - a(i, j))},$$

then we obtain the following equation for the distribution of type i among single men

$$U_m(i) = \frac{L_m(i)}{1 + \frac{\lambda}{\lambda_z} \int_j \frac{a(i, j) U_f(j)}{1 - a(i, j)} dj}$$

Similarly, we obtain the expression for density of single women. Now, we will compute the equilibrium expression of the present value of single men and single women. We obtain the following Bellman equation for a single individual using equation (1) and (3)

$$\begin{aligned} B_i(rV_i^0 - v_i^{0*}) &= \lambda \iint_{z,j} \max(B_i(V_{i|j}^1(z) - V_i^0), 0) U_f(j) dG(z) dj \\ &= \lambda \iint_{z,j} \left(\max \left(\frac{\beta}{r + \lambda_z} \sigma_{ij}(s(i, j) + z), 0 \right) U_f(j) dG(z) dj \right) \\ &= \frac{\beta \lambda}{r + \lambda_z} \int_j \sigma_{ij} \left(\int_z \max(s(i, j) + z, 0) dG(z) \right) U_f(j) dj. \end{aligned}$$

We first compute the inner integral on z :

$$\begin{aligned}
 \int_z \max(s(i, j) + z, 0) dF_z(z) &= s(i, j)a(i, j) + \int_{-s(i, j)}^{+\infty} z dG(z) \\
 &= s(i, j)a(i, j) + \sigma \int_{-\frac{s(i, j)}{\sigma}}^{+\infty} v d\Phi(z) \\
 &= s(i, j)a(i, j) + \sigma \phi\left(\frac{s(i, j)}{\sigma}\right) \\
 &= \mu(a(i, j)),
 \end{aligned}$$

with Φ the gaussian standard distribution function and ϕ its associate density. We obtain the following formula for the present value of a single man of type i and a woman of type j

$$\begin{aligned}
 B_i(rV_i^0 - v_i^{0*}) &= \frac{\beta\lambda}{r + \lambda_z} \int_j \sigma_{ij} \mu(a(i, j)) U_f(j) dj \\
 B_j(rV_j^0 - v_j^{0*}) &= \frac{(1 - \beta)\lambda}{r + \lambda_z} \int_i \sigma_{ij} \mu(a(i, j)) U_m(i) di
 \end{aligned} \tag{19}$$

An equilibrium is a fixed point of (u_m, u_f, V_i^0, V_j^0) of the following system of equations where the first two equations determine equilibrium wage distributions for singles and the last two equations determine equilibrium present values of single men and single women.

$$\begin{aligned}
 U_m(i) &= \frac{L_m(i)}{1 + \frac{\lambda}{\lambda_z} \int_j \frac{a(i, j) U_f(j)}{1 - a(i, j)} dj} \\
 U_f(j) &= \frac{L_f(j)}{1 + \frac{\lambda}{\lambda_z} \int_i \frac{a(i, j) U_m(i)}{1 - a(i, j)} di} \\
 B_i(rV_i^0 - v_i^{0*}) &= \frac{\beta\lambda}{r + \lambda_z} \int_{fj} \sigma_{ij} \mu(a(i, j)) U_f(j) dj \\
 B_j(rV_j^0 - v_j^{0*}) &= \frac{(1 - \beta)\lambda}{r + \lambda_z} \int_{mi} \sigma_{ij} \mu(a(i, j)) U_m(i) di,
 \end{aligned}$$

where $a(i, j)$ solves the following fixed point equation

$$a(i, j) = 1 - G\left(-\Phi(i, j) + \frac{B_i r V_i^0 + B_j r V_j^0}{\sigma_{ij}} - \frac{\lambda_z}{r + \lambda_z} \mu(a(i, j))\right).$$

Despite the lack of a global contraction mapping property, the standard fixed-point iteration algorithm, $x_{n+1} = T x_n$ works well in practice, even starting far from the equilibrium (for instance with $B_i(rV_i^0 - v_i) = 0$ and $u_m(i) = l_m(i)$). For each dataset, the algorithm converges to the equilibrium observed in the data. I obtain the fixed point (u_m, u_f, V_i^0, V_j^0) which corresponds to the density and present value we observe in the data.

6.2 Simulation : impact of preferences, matching patterns and wage distributions

I focus here on the 1999-2008 period where domestic work and employment rates have remained stable. On that period, we observe a decline in men labor supply whereas women labor supply

has remained stable. Moreover, we observe changes in wage distribution, consumption preferences and matching patterns. Particularly, wages have risen and preferences for consumption relative to leisure have increased for women and have decreased for men. Furthermore, high wage women have become more attractive on the marriage market. How do these changes can account for the observed labor supply trends ? To answer this question, I compute the counterfactual equilibrium obtained in 2008 by fixing some elements of the model at their 1999 level. I consider three different elements: the preferences for consumption and leisure represented by the functions A_i , A_j , B_i and B_j , the matching preferences represented by the function Φ , and the wage distribution. Table 8 presents the results. The first two columns describe the simulated results when all factors are those estimated for each year. The model fits the data well. The last three columns present the counterfactual equilibria obtained in 2008 by fixing an element at its 1999 level. The table presents the resulting number of singles, the quantiles of the resulting wage distribution among singles and the obtained average labor supply of men and women according to their marital status.

The results show that the sorting pattern has a strong influence on matching and resulting labor supply. If sorting preferences had remained the same as in 1999, a lot of couples would not have been formed. Many rich women would not have found a partner. Change in sorting pattern can account for an increase in 6% of couples in 2008. It lowers the wage distribution of single women, decreases their average work by 0.6 hour and increases the average labor supply of married women by 0.4 hours.

The increase in preference for leisure for men account for a decrease of 2.1 hours a week for single men and 1.4 hour a week among married men. Preferences changes account for a decrease in labor supply of married women by 2.8 hours.

Finally distribution has also a strong impact. The number of high wage single women has increased. Bargaining position of men seem to have increased with their wages and makes them give lower transfers. The change in wage distribution leads to an increase of 2 hours for married women.

Table 8: Labor supply trends under different scenarios

	Model (deviation** to the current value)		2008 Equilibrium		
	1999	2008	Matching pattern of 1999	Consumption preference of 1999	Wage distribution of 1999
Matching pattern					
U_m	423 (+1)	320 (-3)	341	317	400
U_f	402 (-17)	323 (-4)	345	321	376
Single men quantile*	[10.6 14.7 22.1]	[8.8 12.5 19.4]	[8.8 14.6 19.4]	[8.8 12.5 19.4]	[10.6 14.7 19.5]
Single women quantile*	[7.9 12.5 18.0]	[7.5 10.3 15.4]	[8.9 12.0 19.2]	[7.5 10.3 15.5]	[7.5 12.5 16.1]
Labor supply					
Married men	46.1 (-0.5)	44.4 (+0.1)	44.7	45.8	44.5
Married women	32.1 (-0.5)	31.9 (-0.4)	31.5	34.7	29.8
Single men	41.8 (=)	39.0 (+0.1)	39.1	41.8	39.0
Single women	34.4(+0.2)	34.3(+0.2)	34.9	34.4	33.9

* Quantiles $q25$, $q50$ and $q75$.

** Deviation to the actual value. For instance, in 1999 there are 422 single men who work 41.8 hours a week in average, and the model predicts 423 single men who work in average 41.8 hours a week.

I present on Figure 11 the evolution of the average labor supply of men and women for different marital status from 1999 to 2008. The solid lines represents the mean of observed values whereas the dash lines represent the counterfactual average we would obtain if the matching pattern had stayed at its 1999 level. As previously said, there would be more high wage single women which would mechanically increase labor supply of single women and decrease labor supply of married women. The average effect for men is nil but it hides different effects for different types of individuals. Figure 12 represents in 2008 the counterfactual conditional working hours obtained for married men and women if the matching pattern was still at its 1999 level. High wage women are less attractive then have a lower bargaining power and their husband (usually high wage men) would work less. On the contrary, with 1999 affinity, low wage women are more attractive and their husbands have to work more.

Figure 11: Counterfactual : labor supply with matching pattern of 1999.

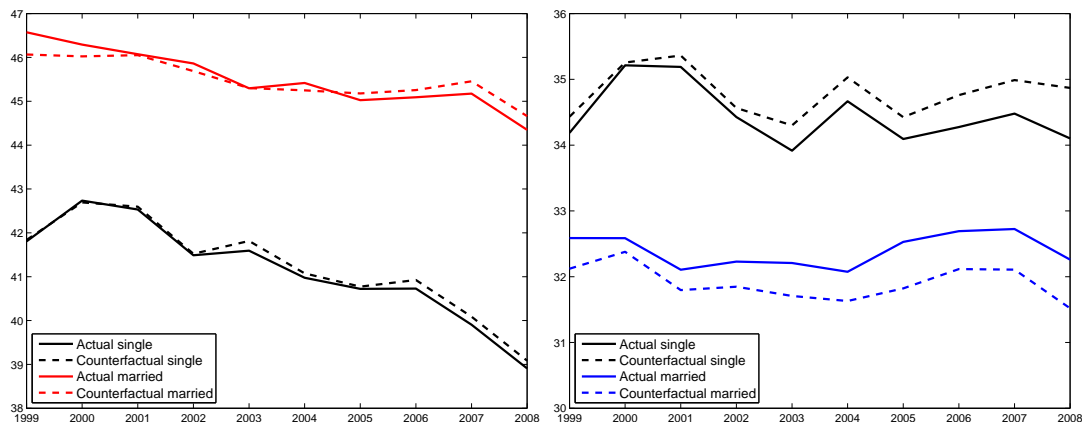


Figure 12: Counterfactual : labor supply with matching pattern of 1999. Cross section in 2008

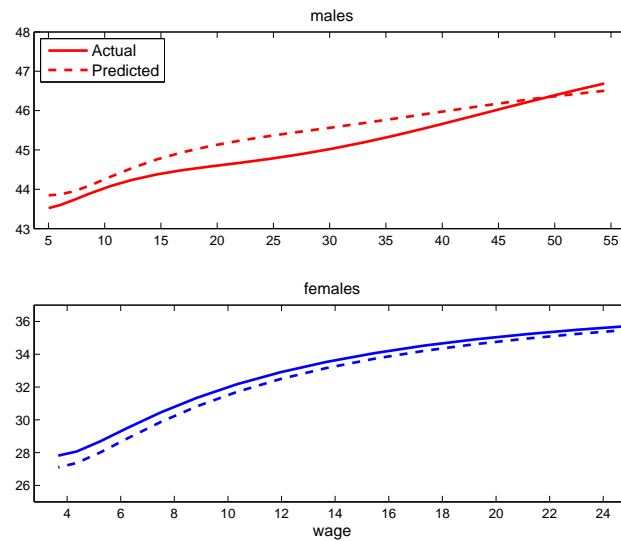
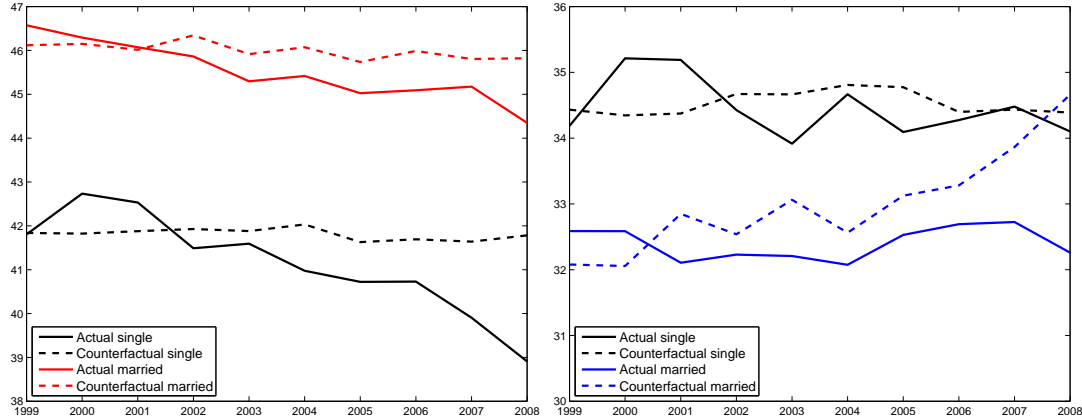


Figure 13 shows counterfactuals of labor supply trends with 1999 preference for consumption and leisure. It shows that at an aggregate level, labor supply of men would not have changed whereas labor supply of single women would have increased by 0.5 hours and labor supply of married women would have increased by almost 2.5 hours.

Figure 13: Counterfactual : labor supply with consumption and leisure preferences of 1999.



6.3 Simulation of other equilibria

What would change if all women had higher wages whereas men's wages stay the same ? Using my sample in year 1999, I slightly change the wage distribution of men and women and look at its impact on different outcomes. I consider the following different scenarios :

- **Scenario 1:** Women's distribution of wage is uniform on $[10\mathcal{L} - 20\mathcal{L}]$ and the distribution of men doesn't change.
- **Scenario 2:** Men's distribution of wage is uniform on $[10\mathcal{L} - 20\mathcal{L}]$ and the distribution of women doesn't change.

- **Scenario 3:** All single women with a wage inferior to 10 £ receive a transfer of 300 £ each month.

Table 9: Simulation exercises

	1999 data	Simulated equilibrium	Scenario 1	Scenario 2	Scenario 3
Matching pattern					
U_m	424	415	438	348	583
U_f	420	415	461	317	583
Social surplus*		1.71×10^6	1.99×10^6	1.91×10^6	1.75×10^6
Labor supply					
Married men	46.5	46.2	43.7	47.1	45.8
Married women	32.7	32.6	35.9	31.7	31.1
Single men	41.8	41.9	41.6	43.5	41.5
Single women	34.2	34.3	37.6	34.2	30.4

* The social surplus is computed as follows

$$\iint \overline{W_{i|j}(z) + W_{j|i}(z)}^{z > -s(i,j)} n(i,j) di dj + \int V_j^0 u_f(j) dj + \int V_i^0 u_m(i) di$$

Scenario 1 When women's wages are higher, there are more single people. Women would not like to match with men with lower wages than them and stay single. Single women are the richest and work more by 3.3 hours because of substitution effects. Single men work less by 0.3 hours. Married women work more by 3.3 hours. They now give a transfer to their husband and married men work less by 2.5 hours due to an income effect. The social surplus is higher.

Scenario 2 When men's wages are higher, there are less single people. All women want to match with higher wage men. Married women work 0.9 hours less because of two opposite effects, a negative income effect and a positive bargaining effect due to a decrease in their transfer. Married men work more by 0.9 hours due to a positive substitution effect which is reduced by a bargaining effect. Single women work the same and single men work more by 1.6 hours whereas they get lower wages than married men (through selection effects). The social surplus is higher but less than in Scenario 1.

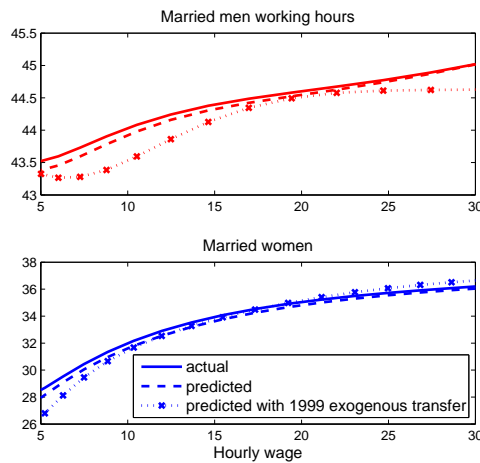
Scenario 3 When single women get high subsidies, they prefer to stay single than losing it. Lot of low wage women are single (consequently lot of low wage men are also single). Married men work 0.4 hours less and married women work 1.5 hour less because they increase their bargaining power. Single men work 0.4 less and single women work less by 3.9 hours due to a big income effect. The social surplus is higher (because money come from nowhere here) but much less than in the last two scenarios.

These simulation exercises would be very interesting to simulate the impact of taxation and family policy programs on matching patterns and labor supplies. This would require the introduction of taxation and children. I propose in section 7 a way to introduce these two important features.

6.4 Is it important to endogenise the sharing rule ?

The main contribution of this paper is to endogenise the within household sharing rule by taking into account the formation and separation of couples. What is the size of the error we make if we don't endogenise the sharing rule ? To assess this gain, I estimate the sharing rule in 1999 with respect to wages. Then for each year between 2000 and 2008, I apply the sharing rule of 1999 on existing married couples and compute their predicted labor supplies. I show on Figure 14 the current working hours conditional on wages in 2008, the labor supply simulated with the model and the labor supply obtained with the 1999 sharing rule. We see the error is significant. Not taking into account the new marriage market state, we would underestimate aggregate labor supply of married men by one hour. Similarly, we would underestimate labor supply of low wage married women by 0.5 hours and overestimate labor supply of high wage women by 0.5 hours.

Figure 14: Conditional labor supply with exogenous sharing rule



7 Extensions

7.1 Extension to taxation

Many countries use joint taxation: taxes are based on the household income level and not on the individual income. Even in countries which use individual taxation as a basis, there can be a bit of joint taxation to give some benefits to low income families. The estimation of collective models with taxation is a little bit trickier. Donni (2003) and Donni and Moreau (2002) showed that the decentralization process still applies but needs additional concepts as shadow wages and shadow non labor income. The household budget constraint with taxation is

$$c_i + c_j + C \leq g[w_i(T - d_i - l_i) + w_j(T - d_j - l_j)]$$

with g representing the total labor income revenue net of taxation. Donni defines shadow wages

ω_i^1 and ω_j^1 as

$$\begin{aligned}\omega_i^1 &= w_i g'[w_i(T - d_i^1 - l_i) + w_j(T - d_j^1 - l_j)] \\ \omega_j^1 &= w_j g'[w_i(T - d_i^1 - l_i) + w_j(T - d_j^1 - l_j)],\end{aligned}$$

And the shadow non labor income as

$$\eta = g[w_i(T - d_i^1 - l_i) + w_j(T - d_j^1 - l_j)] - \omega_i^1(T - d_i^1 - l_i) - \omega_j^1(T - d_j^1 - l_j).$$

The household decentralization process is the following. First the members bargain over the quantity of domestic production they want to produce and about the sharing rule such that $t_i + t_j = \eta$. Then each of them maximizes his own utility under his budget constraint

$$\begin{aligned}\max_{c_i, l_i} & u_i(c_i, l_i) \\ s. c & c_i \leq \omega_i^1(T - l_i - d_i^1) + t_i.\end{aligned}$$

If we consider income support for low income family, we could have a non convex budget set and it would be difficult to solve analytically the model (Salanié, 2003). However, if we consider a negative marginal tax rate (as for instance, we can consider the WFTC in the UK) for low income household when they are working, we may still have a convex budget set. In this case, the model can be derived similarly. Wages are replaced by their shadow wages. Equations of resulting surplus and transfers are derived in appendix.

7.2 Extension to children

Children are not taken into account in this setting. However, I propose a way to introduce children in that kind of model at the cost of two additional strong assumptions. First, I assume that when two people decide to match, they immediately and necessarily have children. Second when a couple separates, it is always the woman who keeps the children¹⁵. The marriage market is then composed of single men without children, single women with or without children and couples with children. The model is still identified.

The present value of single mothers is different from the present value of single women without children. The outside option for married women is now to be a single mother, which changes a little bit the bargaining terms. I present the program of a single woman in appendix as well the modified Nash bargaining expression and the resulting surplus.

The equilibrium on the market is also different. Single men can match with single women without children and with single women with children. Then there exist two different match probabilities. $a(i, j)$ is the match probability of a single woman of type j without children with a single man of type i when she meets him whereas $a(i, j)^c$ is the match probability of a single mother of type j with a single man of type i when she meets him. The steady-state equation condition becomes

$$\lambda_z(1 - a_{ij}^c)N_{ij} = U_m(i)\lambda(U_f(j)a_{ij} + U_f^c(j)a_{ij}^c)$$

¹⁵ In my data, around 65 % of couples have children, 25 % of single women have children whereas less than 2 % of single men have children.

Then, we can still recover s_{ij}^c from data by adding hypothesis on the distribution on z and derive the model. I present the solution in appendix. The estimation is however much more complex.

8 Conclusion

This paper proposes a model which identifies the impact of matching preferences and marital sorting on intra-household allocation and labor supply.

First, this model identifies the total surplus formed by a match. It disentangles what comes from preferences and complementarities of characteristics and what comes from resource sharing and productivity. If the total surplus increases in wages of both members of the household, complementarities in characteristics can be higher for same wage couples. Furthermore, it identifies changes in matching patterns overtime: high wage women have become more attractive and get married more easily.

Second this model shows that relative wages and family values have a large impact on the allocation of resources. High wage women with more traditional family values get higher share of the couple's surplus both in terms of welfare and monetary resources. I identify the within household transfers and show that labor supply of men lies between two extreme cases: one where household members share equally the total resource and one where each partner keeps his own labor income. The analysis of the evolution over 18 years on the BHPS shows that within household inequalities of resources have slightly decreased over these years.

Finally, simulations show that initial distributions of characteristics have a significant impact on matching patterns and resulting labor supplies which confirm the need to model the marriage market together with the sharing rule.

Bibliography

- M. Aguiar and E. Hurst. Measuring trends in leisure: The allocation of time over five decades. *The Quarterly Journal of Economics*, 122(3):969–1006, 2007.
- M.P. Bitler, J.B. Gelbach, H.W. Hoynes, and M. Zavodny. The impact of welfare reform on marriage and divorce. *Demography*, 41(2):213–236, 2004.
- F. Bourguignon, M. Browning, P-A. Chiappori, and V. Lechene. Intrahousehold allocation of consumption: A model and some evidence from french data. *Journal of Political Economy*, pages 1067–1096, 1994.
- M. Bozon and F. Héran. La formation du couple. *Population*, 1:152–159, 1991.
- M. Browning, P-A. Chiappori, and A. Lewbel. Estimating consumption economies of scale, adult equivalence scales, and household bargaining power. *The Review of Economic Studies*, 80(4):1267–1303, 2013.
- K. Burdett and M.G. Coles. Long-term partnership formation: marriage and employment. *The Economic Journal*, pages 307–334, 1999.
- P-A. Chiappori. Rational household labor supply. *Econometrica*, 56(1):63–90, 1988.
- P-A. Chiappori. Introducing household production in collective models of labor supply. *Journal of Political Economy*, 105(1):191–209, 1997.
- P-A. Chiappori, B. Fortin, and G. Lacroix. Marriage market, divorce legislation, and household labor supply. *Journal of Political Economy*, 110(1):37–72, 2002.
- P-A. Chiappori, S. Oreffice, and C. Quintana-Domeque. Matching with a handicap: The case of smoking in the marriage market. 2010.
- P-A. Chiappori, B. Salanié, and Y. Weiss. Partner choice and the marital college premium. *Discussion Papers*, 2011.
- P-A. Chiappori, S. Oreffice, and C. Quintana-Domeque. Fatter attraction: anthropometric and socioeconomic matching on the marriage market. *Journal of Political Economy*, 120(4):659–695, 2012a.
- P-A Chiappori, B. Salanié, F. Salanié, and A. Gandhi. From aggregate betting data to individual risk preferences. 2012b.
- E. Choo and A. Siow. Who marries whom and why ? *Journal of Political Economy*, 114(1):175, 2006.
- H. Couprie. Time allocation within the family: Welfare implications of life in a couple. *The Economic Journal*, 117(516):287–305, 2007.
- O. Donni. Collective household labor supply: nonparticipation and income taxation. *Journal of Public Economics*, 87(5):1179–1198, 2003.

- O. Donni and N. Moreau. Estimation d'un modèle collectif d'offre de travail avec taxation. *Annales d'Économie et de Statistique*, pages 55–83, 2002.
- A. Dupuy and A. Galichon. Personality traits and the marriage market. (*Forthcoming*) *Journal of Political Economy*.
- M. Francesconi, H. Rainer, and W. Van Der Klaauw. The effects of in-work benefit reform in britain on couples: Theory and evidence. *The Economic Journal*, 119(535):F66–F100, 2009.
- A. Galichon and B. Salanié. Cupid's invisible hand: social surplus and identification in matching models. Technical report, mimeo, 2011.
- R. Inglehart and W. E. Baker. Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*, pages 19–51, 2000.
- N. Jacquemet and J-M. Robin. Assortative matching and search with labour supply and home production. Technical report, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 2013.
- J. A. Knowles. Why are married men working so much? an aggregate analysis of intra-household bargaining and labour supply. *The Review of Economic Studies*, 2013.
- J. Lise and S. Seitz. Consumption inequality and intra-household allocations. *The Review of Economic Studies*, 78(1):328–355, 2011.
- M. Manser and M. Brown. Marriage and household decision-making: A bargaining analysis. *International Economic Review*, 21(1):31–44, 1980. ISSN 0020-6598.
- R.D. Mare. Educational assortative mating in two generations. *Department of Sociology, University Of California Los Angeles*, 2008.
- M. B McElroy. The empirical content of nash-bargained household behavior. *Journal of human resources*, pages 559–583, 1990.
- M. B McElroy and M. J. Horney. Nash-bargained household decisions: Toward a generalization of the theory of demand. *International Economic Review*, 22(2):333–349, 1981.
- B. Salanié. *The economics of taxation*. MIT press, 2003.
- R. Shimer and L. Smith. Assortative matching and search. *Econometrica*, 68(2):343–369, 2000.
- J. Waldvogel. Fast construction of the Fejér and Clenshaw–Curtis quadrature rules. *BIT Numerical Mathematics*, 46(1):195–202, 2006. ISSN 0006-3835.
- L.Y. Wong. Structural estimation of marriage models. *Journal of Labor Economics*, 21(3): 699–727, 2003.

1 Appendix

APPENDIX 1 :

Using equation (2) the surplus of a marriage for a i -type man with a j -type woman is

$$B_i(V_{i|j}^1(z) - V_i^0) = \frac{Q(w_i(T - d_i^1) + t_i - A_i)}{r + \lambda_z} + \frac{-rB_iV_i^0 + \lambda_z \int_{z'} \max(B_i(V_{i|j}^1(z) - V_i^0), 0) dG(z')}{r + \lambda_z},$$

then we obtain the following formulas for transfers using equation (3)

$$\begin{aligned} t_i Q &= (r + \lambda_z) \beta S_{ij}(z) + B_i r V_i^0 - (w_i(T - d_i^1) - A_i) Q \\ &\quad - \lambda_z \int_{z'} \max(B_i(V_{i|j}^1(z) - V_i^0), 0) dG(z') \end{aligned} \quad (20)$$

$$\begin{aligned} t_j Q &= (r + \lambda_z)(1 - \beta) S_{ij}(z) + B_j r V_j^0 - (w_j(T - d_j^1) - A_j) Q \\ &\quad - \lambda_z \int_{z'} \max(B_j(V_{j|i}^1(z) - V_j^0), 0) dG(z'). \end{aligned}$$

As we have $t_i + t_j = -C$, we can compute the total surplus by summing the last two equations

$$\begin{aligned} Q(w_i(T - d_i^1) + w_j(T - d_j^1) - A_i - A_j - C) \\ = (r + \lambda_z) S_{ij}(z) + B_i r V_i^0 + B_j r V_j^0 - \lambda_z \int_{z'} \max(S_{ij}(z'), 0) dG(z'). \end{aligned} \quad (21)$$

Using $R_{ij} = \frac{w_i(T - d_i^1) + w_j(T - d_j^1) - A_i - A_j - C}{1 + K_i^1 + K_j^1}$, in equation (21), we obtain

$$Q R_{ij} (1 + K_i^1 + K_j^1) = (r + \lambda_z) S_{ij}(z) + B_i r V_i^0 + B_j r V_j^0 - \lambda_z \int_{z'} \max(S_{ij}(z'), 0) dG(z').$$

where $Q = (\Phi(i, j) + z)(d_i^1 - D_i^1)^{K_i^1} (d_j^1 - D_j^1)^{K_j^1} = (\Phi(i, j) + z) \sigma_{ij}$ where σ_{ij} is defined in equation 12. We get

$$S_{ij}(z) = \frac{\sigma_{ij}}{r + \lambda_z} \left(\Phi(i, j) + z - \frac{B_i r V_i^0 + B_j r V_j^0 - \lambda_z \int_{z'} \max(S_{ij}(z'), 0) dG(z')}{\sigma_{ij}} \right).$$

Then using the formula (20), we obtain the expressions for transfers.

APPENDIX 2 : Additional data description

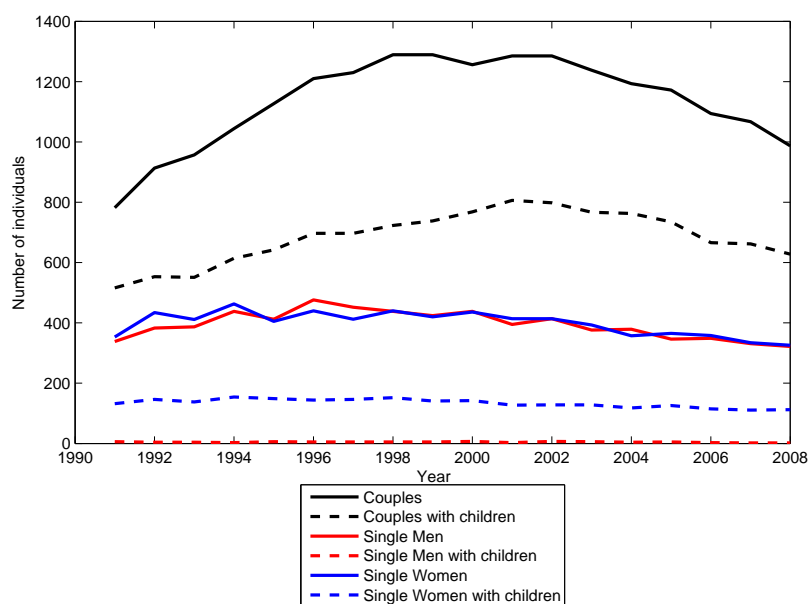


Figure 15: Evolution of sample size from 1991 to 2008. Author's computations from the BHPS population of employed people aged 22-40.

Lecture : in 1992, the sample is composed of 913 couples (of which 553 have children), 383 single men (of which 4 have children) and 434 single women (of which 146 have children)

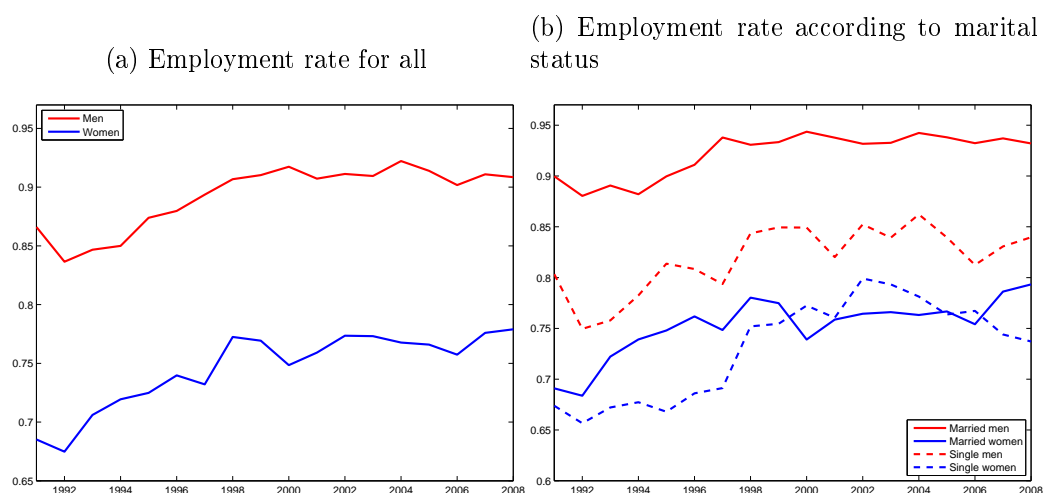
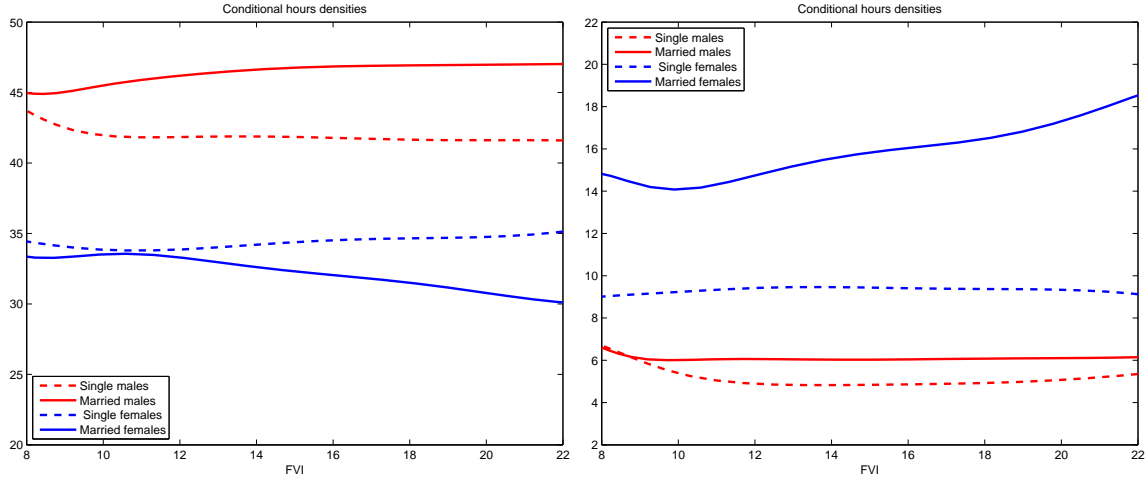


Figure 16: Evolution of the employment rate from 1991 to 2008. Author's computations from the BHPS population of people aged 22-40

Lecture : in 2000, 75 % of women and 90 % of men have a job. 85 % of single men and 93 % of married men have a job.

Figure 17: Average market and domestic work hours conditional on FVI in 1999. BHPS



APPENDIX 3 : Computational details

The computational method is the one used by [Jacquemet and Robin \(2013\)](#). It is adapted to fit a 4D-dimensional model. All functions are discretized on a compact domain using Tchebychev grids. For example, let $[\underline{x}, \bar{x}]$ denote the support of male wages, I construct a grid of $n + 1$ points as

$$x_j = \frac{\underline{x} + \bar{x}}{2} + \frac{\bar{x} - \underline{x}}{2} \cos\left(\frac{j\pi}{n}\right), j = 1 \dots n$$

To estimate wage densities $n(x, y)$, $u_m(x)$ and $u_f(y)$ on those grids, I use kernel density estimators with twice the usual bandwidth to smooth the density functions in the tails. Indeed, additional smoothing is required to divide $n(x, y)$ by $u_m(x)u_f(y)$ to calculate $a(x, y)$. (In Matlab, I need $10^{-16} < a(i, j)$).

1.1 The Clenshaw-Curtis quadrature

Many equations involve integrals. Given Tchebychev grids, it is natural to use Clenshaw-Curtis quadrature to approximate these integrals. 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. The quadrature weights are

$$\begin{aligned} w_0 &= \frac{1}{N} \left(1 + \sum_{j=1}^{\frac{N}{2}} \frac{2}{1 - (2j)^2} \right) & w_{\frac{N}{2}} &= \frac{1}{N} \left(1 + \sum_{j=1}^{\frac{N}{2}} \frac{2(-1)^j}{1 - (2j)^2} \right) \\ w_k &= \frac{2}{N} \left(1 + \frac{(-1)^k}{1 - N^2} + \sum_{j=1}^{\frac{N}{2}-1} \frac{2}{1 - (2j)^2} \cos\left(\frac{2jk\pi}{N}\right) \right) \quad \forall k = 1, \dots, \frac{N}{2} - 1. \end{aligned}$$

I use the method of [Waldvogel \(2006\)](#) who derives a simple algorithm to obtain the weights

of the Clenshaw-Curtis quadrature using matrices, Féjer'quadrature and Discrete Fourier Transform.

1.2 Interpolation

The fact that CC quadrature relies on Tchebychev polynomials of the first kind also allows us to interpolate functions very easily between points $y_0 = f(x_0), \dots, y_n = f(x_n)$ using Discrete Cosine Transform (DCT) such that

$$f(x) = \sum_{k=0}^n Y_k T_k(x) \quad (22)$$

where Y_k are the OLS estimates of the regression of $y = (y_0, \dots, y_n)$ on Tchebychev polynomials

$$T_k(x) = \cos \left(k \arccos \left(\frac{x - \frac{x+\bar{x}}{2}}{\frac{\bar{x}-x}{2}} \right) \right)$$

but are more effectively calculated using FFT. A MATLAB code for DCT is

```
Y = y([1:n+1 n:-1:2],:);
Y = real(fft(Y/2/n));
Y = [Y(1,:); Y(2:n,:)+Y(2*n:-1:n+2,:); Y(n+1,:)];
f = @(x) cos(acos((2*x-(xmin+xmax))/(xmax-xmin))*(0:n))*Y(1:n+1);
```

with $y = (y_0, \dots, y_n)$. A bidimensional version is

```
Y = y([1:n+1 n:-1:2],:);
Y = real(fft(Y/2/n));
Y = [Y(1,:); Y(2:n,:)+Y(2*n:-1:n+2,:); Y(n+1,:)];
Y = Y(:, [1:n+1 n:-1:2]);
Y = real(fft(Y'/2/n));
Y = [Y(1,:); Y(2:n,:)+Y(2*n:-1:n+2,:); Y(n+1,:)]';
f=@(x,y) cos(acos((2*x-(xmin+xmax))/(xmax-xmin))*(0:n))*Y(1:n+1,1:n+1)...
*cos((0:n)'*acos((2*y'-(ymin+ymax))/(ymax-ymin))));
```

I also use a 4D dimensional version to evaluate transfers which depend on 4 variables : w_m, w_f, f_m, f_f . The fact that the grid (x_0, \dots, x_n) is not uniform and is denser towards the edges of the support interval allows to minimize the interpolation error and thus avoids the standard problem of strong oscillations at the edges of the interpolation interval (Runge's phenomenon). Another advantage of DCT is that, having calculated Y_0, \dots, Y_n , then polynomial projections of $y = (y_0, \dots, y_n)$ of any order $p \leq n$ are obtained by stopping the summation in (22) at $k = p$. Finally, it is easy to approximate the derivative f' or the primitive $\int f$ simply by differentiating or integrating Chebyshev polynomials using

$$\cos(k \arccos x)' = \frac{k \sin(k \arccos x)}{\sin(\arccos x)}$$

and

$$\begin{aligned}
\int \cos(k \arccos x) &= x \text{ if } k = 0 \\
&= x^2/2 \text{ if } k = 1 \\
&= \frac{\cos((k+1)x)}{2(k+1)} - \frac{\cos((k-1)x)}{2(k-1)} \text{ if } k \geq 1.
\end{aligned}$$

In calculating an approximation of the derivative, it is useful to smooth the function by summing over only a few polynomials. Derivatives are otherwise badly calculated near the boundary.

APPENDIX 4 : Extension to children

1.3 Present value

In this variation, women's preferences for leisure and consumption are the same for women with children than for women without children. Having children incur a cost for both single mothers and couples with children. Singles without children do not pay this cost. Married and single mothers value similarly domestic production with the same preference parameter K_j^1 and need the same minimum quantity of housework D_f^1 . The indirect utility for a j -type single women with children is

$$v_j^c = \frac{(d_j^c - D_f^1)^{K_j^1} (w_j(T - d_j^c) - A_j - C)}{B_j}.$$

The single present value for a single women without children reads

$$rV_j^0 = v_j^{0*} + \lambda \iint \max(V_{j|i}^1(z) - V_j^0, 0) \mathbb{1}(V_{i|j}^1(z) > V_i^0) U_m(i) dG(z) di.$$

The single present value for a single women with children reads

$$rV_j^c = v_j^{c*} + \lambda \iint \max(V_{j|i}^1(z) - V_j^c, 0) \mathbb{1}(V_{i|j}^1(z) > V_i^0) U_m(i) dG(z) di.$$

A single man can now meet either a woman without children or a single mother. His present value is

$$\begin{aligned}
rV_i^0 &= v_i^{0*} + \lambda \iint \max(V_{i|j}^1(z) - V_i^0, 0) \mathbb{1}(V_{j|i}^1(z) > V_j^0) U_f(j) dG(z) dj \\
&+ \lambda \iint \max(V_{i|j}^1(z) - V_i^0, 0) \mathbb{1}(V_{j|i}^1(z) > V_j^c) U_f^c(j) dG(z) dj.
\end{aligned}$$

When the couple breaks up, the woman becomes a single mother and the man becomes a single man without children. The present values of members of a couple are:

$$\begin{aligned}
rV_{j|i}^1(z) &= v_j^{0*} + \lambda_z \int_z \max(V_{j|i}^1(z') - V_j^0, 0) \mathbb{1}(V_{i|j}^1(z) > V_i^0) U_m(i) dG(z) di \\
rV_{i|j}^1(z) &= v_i^{0*} + \lambda_z \int_z \max(V_{i|j}^1(z') - V_i^0, 0) \mathbb{1}(V_{j|i}^1(z) > V_j^c) U_f(j) dG(z) dj.
\end{aligned}$$

1.4 Surplus, Nash bargaining and transfers

When the two members of the couple bargain, the outside option for the man is still his single present value whereas the outside option of the woman is now a single mother present value. The Nash bargaining is now modeled by the following program

$$\begin{aligned} \max & (V_{i|j}^1(z) - V_i^0)^\beta (V_{j|i}^1(z) - V_j^c)^{1-\beta} \\ \text{s.t.} & t_i + t_j = -C, \end{aligned}$$

whose solution gives

$$\begin{aligned} B_i(V_{i|j}^1(z) - V_i^0) &= \beta S_{ij}(z) \\ B_j(V_{j|i}^1(z) - V_j^c) &= (1 - \beta) S_{ij}(z). \end{aligned} \tag{23}$$

where $S_{ij}(z)$ is still linear in z and equals $\frac{\sigma_{ij}}{r + \lambda_z}(s_{ij}^c + z)$.

When a single woman without children meets a man, she knows that she will get children if she marries him and that her outside option will be the one of a single mother. In this case, the surplus for a woman without children is different for a woman with children. And a match of a woman without children of type j with a man of type i under circumstance z can be valuable for the man and not for the woman, whereas it would have been valuable for a single mother of type j . The match probability between a single mother of type j who meets a single man of type i is

$$a_{ij}^c = \mathbb{P}(V_{j|i}^1(z) - V_j^c > 0) = \mathbb{P}(z > -s_{ij}^c),$$

whereas the probability that a woman type j without children match with a man of type i when she meets one can be written :

$$a_{ij} = \mathbb{P}(V_{j|i}^1(z) - V_j^0 > 0 \ \& \ V_{i|j}^1(z) - V_i^0 > 0).$$

After some algebra¹⁶ we can similarly write

$$a_{ij} = \mathbb{P}\left(z > -s_{ij}^c + \max\left(0, (V_j^0 - V_j^c) \frac{B_j(r + \lambda_z)}{\sigma_{ij}(1 - \beta)}\right)\right).$$

This last expression shows us that a match will be more valuable for a single woman without children if her single value with a child is large. It reminds us the result of the job market

¹⁶

$$\begin{aligned} a_{ij} &= \mathbb{P}(V_{j|i}^1(z) - V_j^0 > 0 \ \& \ V_{i|j}^1(z) - V_i^0 > 0) \\ &= \mathbb{P}(V_{j|i}^1(z) - V_j^c + V_j^c - V_j^0 > 0 \ \& \ V_{i|j}^1(z) - V_j^c > 0) \\ &= \mathbb{P}\left(\frac{(1 - \beta)\sigma_{ij}}{(r + \lambda_z)B_j}(s_{ij}^c + z) + V_j^c - V_j^0 > 0 \ \& \ z > -s_{ij}^c\right) \\ &= \mathbb{P}\left(z > (V_j^0 - V_j^c) \frac{B_j(r + \lambda_z)}{\sigma_{ij}(1 - \beta)} - s_{ij}^c \ \& \ z > -s_{ij}^c\right) \end{aligned}$$

search : the reservation wage of non-participants lowers when the unemployment benefits for the unemployed rise.

$(V_j^0 - V_j^c)$ is complex to derive and can be computed recursively as follows.

$$\begin{aligned}
rV_j^0 - rV_j^c &= v_j^{0*} - v_j^{c*} + \lambda \iint \max(V_{j|i}^1(z) - V_j^0, 0) \mathbb{1}(V_{i|j}^1(z) > V_i^0) U_m(i) di dG(z) \\
&\quad - \lambda \iint \max(V_{j|i}^1(z) - V_j^c, 0) \mathbb{1}(V_{i|j}^1(z) > V_i^0) U_m(i) di dG(z) \\
&= v_j^{0*} - v_j^{c*} + \lambda \iint (\max(V_{j|i}^1(z) - V_j^0, 0) - \max(V_{j|i}^1(z) - V_j^c, 0)) \mathbb{1}(V_{j|i}^1(z) > V_j^c) U_m(i) di dG(z) \\
&= v_j^{0*} - v_j^{c*} + \lambda \iint (\max(V_{j|i}^1(z) - V_j^0, 0) - V_{j|i}^1(z) + V_j^c) \mathbb{1}(V_{j|i}^1(z) > V_j^c) U_m(i) di dG(z) \\
&= v_j^{0*} - v_j^{c*} + \lambda(V_j^c - V_j^0) \int_i a_{ij} U_m(i) di - \lambda \iint (V_{j|i}^1(z) - V_j^c) \mathbb{1}(V_j^0 > V_{j|i}^1(z) > V_j^c) U_m(i) di dG(z),
\end{aligned}$$

then

$$\begin{aligned}
\left(r + \lambda \int a_{ij} U_m(i) \right) B_j(V_j^0 - V_j^c) &= B_j(v_j^{0*} - v_j^{c*}) \\
&\quad - \lambda \iint B_j(V_{j|i}^1(z) - V_j^c) \mathbb{1}(0 < V_{j|i}^1(z) - V_j^c < V_j^0 - V_j^c) U_m(i) dG(z) \\
&= B_j(v_j^{0*} - v_j^{c*}) \\
&\quad - \lambda \frac{1 - \beta}{(r + \lambda z)} \int \sigma_{ij} U_m(i) di \int_{z > -s_{ij}^c}^{z > -s_{ij}^c + \frac{B_j(V_j^0 - V_j^c)(r + \lambda z)}{(1 - \beta)\sigma_{ij}}} (s_{ij}^c + z) dG(z),
\end{aligned}$$

with

$$\begin{aligned}
\int_{z > -s_{ij}^c}^{z > -s_{ij}^c + \frac{B_j(V_j^0 - V_j^c)(r + \lambda z)}{(1 - \beta)\sigma_{ij}}} (s_{ij}^c + z) dF_z(z) &= \\
&= s_{ij}^c (a_{ij}^c - a_{ij}) + \sigma \left(\phi \left(\frac{s_{ij}^c}{\sigma} \right) - \phi \left(\frac{s_{ij}^c}{\sigma} + \frac{B_j(V_j^0 - V_j^c)(r + \lambda z)}{\sigma(1 - \beta)\sigma_{ij}} \right) \right).
\end{aligned}$$

We obtain a complex recursive formula for $V_j^0 - V_j^c$. I use a fixed point algorithm to estimate it. First I suppose that $V_j^0 - V_j^c = 0$ and I compute a_{ij} as follows

$$\begin{aligned}
a_{ij} &= \mathbb{P}(V_{j|i}^1(z) - V_j^0 > 0 \ \& \ V_{i|j}^1(z) - V_i^0 > 0) \\
&= \mathbb{P}(V_{i|j}^1(z) - V_i^0 > 0) \\
&= a_{ij}^c,
\end{aligned}$$

with

$$\lambda_z(1 - a_{ij}^c)N_{ij} = U_m(i)\lambda(U_f(j)a_{ij} + U_f^c(j)a_{ij}^c).$$

Then I find a new value for $V_j^0 - V_j^c$ using

$$V_j^0 - V_j^c = \frac{v_j^{0*} - v_j^{c*}}{r + \lambda \int_i a_{ij} U_m(i)}.$$

Finally I use my new estimate as an initial value and resume the process until convergence.

APPENDIX 5 : Extension to taxation

Let R denotes labor income and τ_1, τ_2, τ_3 denote the subsidies rates. Let us consider the function g which represents the total labor income net of transfer and taxes such that

$$g(R) = R + \tau_1 R \mathbb{1}_{R \leq A_1} + \tau_2 R \mathbb{1}_{A_1 < R \leq A_2} + \tau_3 R \mathbb{1}_{R > A_2},$$

with $\tau_1 > \tau_2 > \tau_3 > 0$ and $g'(R) = \frac{g(R)}{R}$. Then the shadow wages and income are

$$\begin{aligned}\omega_i^1 &= w_i(1 + \tau_1 \mathbb{1}_{R \leq A_1} + \tau_2 \mathbb{1}_{A_1 < R \leq A_2} + \tau_3 \mathbb{1}_{R > A_2}) \\ \omega_j^1 &= w_j(1 + \tau_1 \mathbb{1}_{R \leq A_1} + \tau_2 \mathbb{1}_{A_1 < R \leq A_2} + \tau_3 \mathbb{1}_{R > A_2}) \\ \eta &= 0.\end{aligned}$$

The indirect of singles remains

$$v_i^0 = \frac{(d_i - D_m)^{K_m}(w_i(T - d_m) - A_i(w_i))}{B_i(w_i)},$$

and the indirect utility when married becomes

$$v_i^1 = Q \frac{\omega_i^1(T - d_i) + t'_i - A_i(\omega_i^1)}{B_i(\omega_i^1)}.$$

Then the surplus equation becomes

$$s_{ij} = \Phi(i, j) - \frac{B_i r V_i^0 + B_j r V_j^0}{\sigma_{ij}(\omega_i^1, \omega_j^1)} + \frac{\lambda_z}{r + \lambda_z} \int_{z'} \max(s_{ij} + z', 0) dG(z'),$$

where the single present value reads

$$B_i(r V_i^0 - v_i^{0*}) = \frac{\beta \lambda}{r + \lambda_z} \int_{fj} \sigma_{ij}(\omega_i^1, \omega_j^1) \mu(a(i, j)) du_f(j) dj,$$

and the new transfer is

$$t'_i = \beta R_{ij}(\omega_i^1, \omega_j^1) - (\omega_i^1(T - d_i) - A_i(\omega_i^1)) + \frac{(1 - \beta)B_i r V_i^0 - \beta B_j r V_j^0}{Q}.$$

If we assume that A_i and A_j are linear in wages and the logarithm of B_i and B_j are linear in

the logarithm of the wage (Cobb Douglas specification of utility), then

$$A_i(w_i) = a_i^1 + a_i^0 w_i$$

$$\log B_i(w_i) = (1 - b_i^1) \log w_i.$$

In that case we obtain $\frac{B'_i}{B_i} = \frac{1-b_i^1}{w_i}$ and the labor supply equations for married men and single men rewrite

$$h_i^1 = T - d_i^1 - a_i^1 - \frac{1-b_i^1}{\omega_i^1} (\omega_i^1 (T - d_i^1) + t'_i - a_i^1 \omega_i^1 - a_i^0)$$

$$h_i^0 = T - d_i^0 - a_i^1 - \frac{1-b_i^1}{w_i} (w_i (T - d_i^0) - a_i^1 w_i - a_i^0).$$

h_i^1 can be rewritten

$$\begin{aligned} h_i^1 &= T - d_i^1 - a_i^1 - \frac{1-b_i^1}{\omega_i} (\omega_i (T - d_i^1) + t'_i - a_i^1 \omega_i - a_i^0) \\ &= T - d_i^1 - a_i^1 - \frac{1-b_i^1}{g' w_i} (g' w_i (T - d_i^1) + t'_i - a_i^1 g' w_i - a_i^0) \\ &= T - d_i^1 - a_i^1 - \frac{1-b_i^1}{w_i} (w_i (T - d_i^1) + \frac{t'_i - a_i^0}{g'} - a_i^1 w_i) \end{aligned}$$

and h_i^0 doesn't change. Then $a_m^1, a_m^0, a_f^1, a_f^0, b_m^1$ and b_f^1 are still identifiable.

APPENDIX 6 : Generalized Additive Models

This explanation has been developed in [Chiappori et al. \(2012b\)](#). Generalized additive models (GAM) were introduced by Hastie and Tibshirani (1986). They model a variable y_i by assuming that its distribution around its mean belongs to the exponential family and by modeling the mean as a sum of smooth functions of subvectors of the covariates (X_i). To estimate my GAM models, I use the methods described by Wood (2006); I use his implementation in the `mgcv` package of R, which incorporates the improved algorithm of Wood (2008). More precisely, one writes

$$\mathbb{E}(y_i|X) = \sum_{j=1}^J f_j(X_i^j)$$

where each X_i^j is a user-defined subvector of X_i , and the f_j are to be estimated; and the user also chooses the distribution of the error term $(y_i - \mathbb{E}y_i)$ within the exponential family. Modeling starts by choosing a rich family of basis functions (typically splines) (b_{jk}) for $k = 1 \dots K_j$ with a maximal order K_j chosen large enough. Then

$$f_j(X_i^j) = \sum_{k=1}^{K_j} \beta_{jk} b_{jk}(X_i^j)$$

Finally, the generalized R^2 cited in the text are defined as the ratio $\left(1 - \frac{\text{EV}(y|X)}{\text{V}(y)}\right)$.

Part III

Grade retention in French Junior High School

Chapter 3

Why do Inequalities in 15-year-old Cognitive Skills Increase in France between 2000 and 2009?

Abstract

PISA data show that inequalities in French student performances have increased between 2000 and 2009: the share of low-achieving students has increased while the share of high-achieving students has remained roughly the same. We consider that the French educational system has contemporaneously experienced two main changes: first new sorting practices triggered by a decline in grade retention, and second successive extensions of special education zones. Our study aims at assessing to what extent the change in the performance distribution over time can be attributed to those policy changes. We analyze the general decline in students' test scores by using distributional decomposition techniques. Score gaps are decomposed into their detailed components based on the multivariate regression of cognitive achievement on many observable variables related to student, family, school and institutional characteristics. The decomposition results show that most of the changes in performance distribution between 2000 and 2009 seem related to the declining quality of the French educational system rather than to shifts in students' composition. Our findings suggest that the increasingly negative effect of grade retention outweighs the positive evolution in students' composition due to the decline in retention practices. Secondly, the extension of areas targeted for special help and a lesser efficiency of that policy have contributed to the decline in reading performances of lowest-achieving students. Finally around two thirds of the decline in the score of the 20 percent lowest performers are related to changes in sorting practices and in special education policy.

1 Introduction

As reading literacy is the primary focus of the Program for International Student Assessment (PISA) in 2000 and in 2009, the publication of PISA 2009 outcomes has paved the way for particularly sharp comparisons of country's performance in reading over the 2000 decade. Among developed countries, France comes out as an interesting study case: whereas the average variation in student reading performance has decreased by 3% across OECD between 2000 and 2009 (OECD, 2010b), it has increased by more than 15% in France. Other European countries, namely Sweden, Italy, Iceland and Spain, also experienced a significant increase in performance dispersion but that increase did not exceed 15%. Among OECD countries, only Japan and Korea saw a larger rise in inequality than France. Like in Sweden, performance variation in France increased due to a decline in the performance of low-achieving students, while the score of the highest-achieving students remained roughly the same. The rise in performance variation is consequently associated with a decline, yet moderate, in the overall reading performance of 15-year-old students between 2000 and 2009 in France. Identifying the sources of change in France's reading performances can help policy makers design effective policies to overcome inequalities in learning opportunities and declining overall performance. Thus the question arises how the quality of the French educational system has changed across the 2000 decade and whether this can explain the rise in performance inequalities.

We consider that French educational system has contemporaneously experienced two main relevant changes: a decline in retention rates and successive extensions of areas targeted for special help in education. The decline in retention rates in France has been motivated by two factors. First, past international large-scale surveys, such as TIMSS¹, PIRLS² and more recently PISA, brought out France, together with Luxembourg, Spain, Portugal, Belgium and the Netherlands, as a country with high retention rates (OECD, 2004, 2010a). Second, French past research in education pointed out the lack of efficiency and of equity of grade retention (Levasseur and Seibel, 1985; Grisay, 1993). A change in students sorting practices has consequently been implemented since the beginning of the 1990s. To our knowledge, while short and long-term effects of grade retention on students' achievement have been considerably examined (see Jimerson et al. (2006) for a meta-analysis), there has not been any assessment of the decline in grade retention over the last years in France. Yet we expect the decline in grade retention to have affected student performance distribution. The French educational system has experienced a second major change: successive expansions of areas targeted for special help. Many OECD countries, such as the United States, Great Britain, Portugal and Belgium, have adopted compensatory education programs that channel supplementary teaching resources to disadvantaged schools (Bénabou et al., 2009). The program "Zones d'Education Prioritaires" (Priority Education Zones, henceforth ZEP) was launched in 1982 in France: it provides selected schools (mostly primary and middle schools but also few high schools which concentrate a high share of disadvantaged students with low levels of educational achievement) with additional funds, smaller class size and more teaching hours. Originally meant to be temporary, the program was instead maintained in the

¹Trends in International Mathematics and Sciences Study.

²Progress in International Reading Literacy Study.

initially targeted areas and it was successively extended to other areas. Assessments of the ZEP policy found mixed results regarding their impact on students' achievement and attitude towards school (Caille, 2001; Meuret, 1994; Piketty and Valdenaire, 2006; Bénabou et al., 2009). Besides, within the French education community, few researchers have argued that the extension of the program to other schools has led to diluting extra resources directed to selected schools (Maurin, 2004; Merle, 2012). We believe that this might have contributed to the increase of low-achievers intake.

This article assesses the extent to which the rise in performance inequalities over time can be attributed to those two policy shifts: the decline in grade retentions and successive extensions of compensatory education programs. Based on the investigation of French PISA 2000 and 2009 datasets, we analyze changes in students' test scores by using decomposition techniques of mean and quantile scores differential. The decomposition results show that 15-year-old students in France in 2009 might have slightly less favorable characteristics than their counterparts in 2000, but most importantly, that, except higher-achieving students, they experience lower returns to school's institutional characteristics than their counterparts in 2000. In fact most of the changes in performance distribution between 2000 and 2009 seem related to the declining quality of the French educational system rather than to shifts in students' composition. Our findings suggest that the increasingly negative effect of grade retention outweighs the positive evolution in students' composition due to the decline in retention practices. Secondly, the extension of areas targeted for special help and a lesser efficiency of that policy have contributed to the decline in reading performances of lowest-achieving students. Finally around two thirds of the decline in the score of the 20 percent lowest performers are related to changes in sorting practices and in special education policy.

The paper is organized as follows. In Section 2, we describe the main changes in French educational policies and we discuss their expected consequences on student reading performance. Section 3 introduces PISA 2000 and 2009 datasets and the variables selected for temporal comparisons. Section 4 describes the Oaxaca-Blinder and Firpo-Fortin-Lemieux decomposition techniques. Decomposition results and robustness checks are presented in Section 5. Section 6 summarizes and discusses our main findings and their political implications.

2 Main changes in French educational policies and their expected consequences on 15-year-old performances

Change in 15-year-old students' reading performances might result from changes in the production process of schooling within the French educational system. Since PISA tests assess students' achievement around the end of compulsory schooling, it is worth considering any change which could have affected students' learning conditions from their entry in primary school in September of the civil year they turn 6. The PISA 2000 cohort was born in year 1984 and entered primary school in 1990, whereas the PISA 2009 cohort was born nine years later in 1993 and entered primary school in 1999. We should consequently pay particular attention to the main changes which occur in the French educational system between 1990 and 2009. We reviewed the possible variations in the main organizational dimensions (Jonsson and Erikson, 2000; Le Donné, 2014)

of the French educational system. The reinforcement of the plan attributed catchment areas to schools implemented from 1997 up to 2007 has barely affected the allocation of students across schools and a fortiori students' learning conditions (Van Zanten and Obin, 2008). The size and the operating principles of the private paying sector have remained roughly the same across the last two decades (MEN, 1998, 2009). However a decline in retention rates and successive extensions of areas targeted for special help in education are two institutional shifts which deserve to be examined in depth.

2.1 The decline in retention rates

The aim of reducing retention rates in primary and lower-secondary education is one component of the "*politique des cycles*" ("cycles policy") launched by the 1989 Education Orientation Law in France. This political recommendation follows on from several French studies showing negative effects of grade retention on students' academic achievement (Levasseur and Seibel, 1985; Grisay, 1993) and from international surveys bringing out France as a country with very high retention rates. Since 1989 retention rates have effectively declined in France: the share of students having repeated a grade during primary school equaled 20%, respectively 16%, at the time where the PISA 2000 cohort, respectively the PISA 2009 cohort, reached the end of primary school (MEN, 1998, 2009). Retention rates during lower-secondary education have also decreased by around 3 percentage points within the nine years that separate the enrollment of PISA 2000 and 2009 cohorts in those grades (MEN, 1998, 2009).

The fall in retention practices has necessarily modified the allocation of 15-year-old students into the different grades and tracks available in the French educational system. In the 1990s like in the 2000s, primary and lower-secondary education corresponds to a comprehensive system (see Figure 1): pupils are firstly enrolled from the age of 6 in primary schools which last five grades, then from the age of 11 in the so-called "*collèges*" which are unified four-grade middle schools. At the end of middle school, students are allocated into two main tracks: either into the general and technological track which delivers "*baccalauréat*" and prepares to academic tertiary education, or into the vocational track which delivers vocational qualifications and possibly thereafter a vocational "*baccalauréat*". 15-year-old students who have never been retained are consequently enrolled in high school, either in the general or in the vocational track, whereas 15-year-old students who have repeated at least one grade are still enrolled in middle school. The decline in retention rates has led to a significant decline in the share of 15-year-old students enrolled in middle school and to an increase in the shares of 15-year-old students enrolled in high school, in the general or in the vocational tracks.

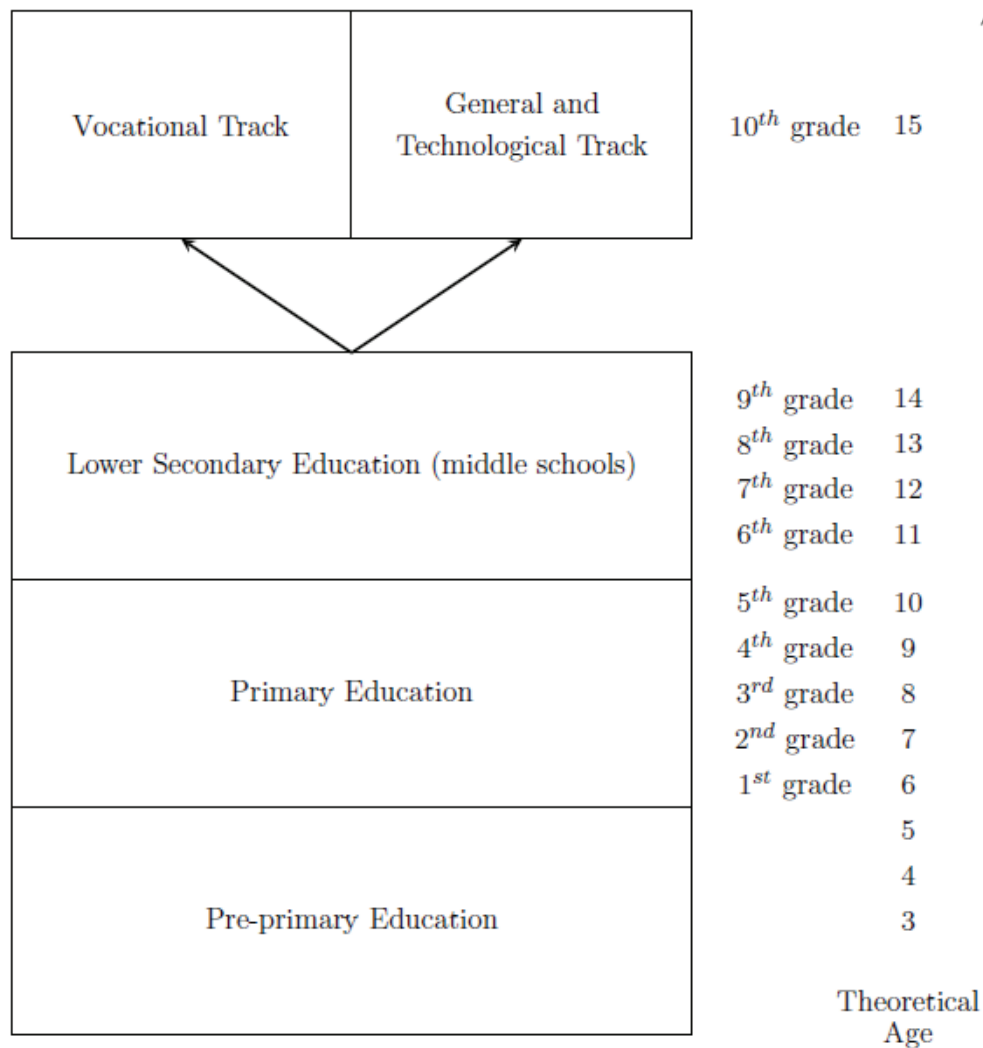


Figure 1: French school system

While many French and international studies have examined short and long-term effects of grade retention on students' achievement (for a meta-analysis see [Jimerson et al. \(2006\)](#)), as far as we know, there has not been any assessment of the decline in retention rates in France. The question arises yet how the more frequent promotion of students to next grades has affected student cognitive enhancement. To structure the investigation, we break down the whole student population into three notional groups: students "henceforth promoted", who would have probably been retained in 2000 but who are promoted in 2009 thank to new sorting recommendations; students who are "always promoted", whose risk of retention is low irrespective of retention practices; and students who are "always retained", whose probability of retention is high irrespective of retention practices. For those students who would have been retained in the past but who have "henceforth been promoted" to next grades, we expect that this promotion has been positive for their cognitive enhancement. Indeed, those students do not experience the stigma generally associated with grade retention ([Jackson, 1975](#); [Reynolds, 1992](#)) and, if their schooling level is sufficient, they might benefit from a more demanding and advantageous schooling environment, in particular if they are enrolled in the general track of high school.

The recent promotion of those students might also have negative effects on the performances of students who are "always promoted", irrespective of retention practices. Studies about the change from a selective educational system to a comprehensive unified system in the United Kingdom concluded that performances of high-achieving students have been weakened by the new presence of low-achieving peers within the unified general track (Galindo-Rueda and Vignoles, 2007; Manning and Pischke, 2006). But studies assessing comprehensive experiments in Europe and in the United States rather show that the positive effects of the reform for low-achieving students exceed the negative effects for high-achieving students (Gamoran, 1996; Figlio and Page, 2002; Palme and Meghir, 2005; Maurin, 2007; Pekkarinen et al., 2009; Jakubowski et al., 2010).

Lastly, the decline in retention rates might also affect the cognitive achievement of "always retained" students. Since retention rates have declined, retention decisions are likely to be based on different motives and generate different short-term effects on students' educational success. If grade repetition has recently been more intended as a measure of support (Kloosterman and de Graaf, 2010), *i.e.* as a real second chance for students coping with significant difficulties, we assume that the fall in grade retention would have more profited to retained students of recent birth cohorts than of previous ones. But if grade retention has more often been used as a mean to reduce schooling heterogeneity and if it is associated with a higher stigma in the recent context of declining retention practices, it might restrain even more than before students' cognitive enhancement. As grade retention is more often used as a mean to reduce schooling heterogeneity in France than in other OECD countries (Mons, 2007), we privilege the latter hypothesis over the former one.

2.2 The extension of "Zones d'Education Prioritaires"

The French educational system has experienced a second major change: the extension of "Zones d'Education Prioritaires" (ZEP). Launched in 1982, the ZEP policy aims at reducing schooling inequality by strengthening schooling action towards areas with a high concentration of disadvantaged populations. The ZEP program provides extra funding and teaching resources mainly to primary and middle schools (but also to few high schools) located in targeted areas and allow them to develop educational projects (for a description of this policy, see Bénabou et al. (2009)). Since its first implementation, the program has successively been expanded to other disadvantaged areas and has consequently concerned more and more students. The share of lower-secondary students enrolled in schools targeted for special help amounts to 10% in 1982, 14.3% in 1997 and to about 20% in 2009 (MEN, 1998, 2009). A French diagnosis report notes that many of the schools originally targeted by the program do not meet its criteria anymore since their composition and their overall achievement have improved. But none of them have been removed from the program yet, while newly disadvantaged schools have joined it. Despite successive extensions of the policy, social composition of ZEP schools has consequently not changed over time (Bénabou et al., 2009).

Past studies show mixed effects regarding the ZEP policy. First the ZEP program has resulted in smaller class sizes and in more teaching hours (Bénabou et al., 2009; Merle, 2012). However the additional resources channelled to ZEP schools have not always offset their initial lack of resources: some ZEP schools still have fewer funds than non-ZEP schools, in spite of the

policy (Moisan and Simon, 1998). Nonetheless additional teaching resources have led to positive short-term effects on cognitive achievement of students enrolled in ZEP primary schools (Piketty and Valdenaire, 2006; Bressoux et al., 2009) and on educational attitudes of ZEP middle school students (Meuret, 1994). The ZEP policy was found to have no long-term effect on students' educational attainment: at given social characteristics, ZEP students have as many chances as non-ZEP students to get any French degree, to reach the 8th and the 10th grades or to become "*baccalauréat*" incumbents (Caille, 2001; Bénabou et al., 2009).

The ZEP policy also triggered negative signalling effects. The ZEP status might have signalled low probability of educational success to students enrolled in targeted schools and it might have weakened student learning motivations (Merle, 2012). But, above all, the ZEP status had negative consequences on educational professionals. Despite financial incentives created to attract teachers, ZEP schools have experienced sharper teacher shortfall, higher turn-over rates, more temporary and delayed teacher affectations, and higher shares of less-experienced teachers (Bénabou et al., 2009). The negative signaling effects of the ZEP policy might have offset the positives effects related to smaller class size and more teaching hours.

We further assume that the geographical and numerical extension of the ZEP policy has affected 15-year-old students' performances. Existing research argues that the ZEP expansion has led to a dilution of the amount of extra resources per school (Maurin, 2004; Merle, 2012). As the ZEP status concerns more and more schools in 2009 than in 2000, while providing less financial and teaching resources to each targeted school, we expect the policy to be less efficient. Thus the ZEP extension could explain part of the decline in low-achieving student performance and consequently part of the rise in inequalities among 15-year-olds in France.

3 French PISA 2000 and 2009 data

The design of the Programme for International Student Assessment (PISA) does not only enable international comparisons of countries in terms of their learning outcomes, it also allows temporal comparisons of students' performance within countries. Conducted every 3 years since 2000, each PISA survey assesses one of three core domains in depth (considered the major domain) among reading, mathematics and sciences. PISA cycle started in 2000 with reading literacy as the major subject area. After a complete rotation of the three domains of assessment, reading is again the primary focus of PISA in 2009. For the first time since PISA launch, it is possible to obtain detailed comparison of how student performance in the major domain has changed³, that is the reason why this article is focused on inequalities in reading achievement and not in mathematics or in sciences. One should also keep in mind that PISA, like any assessment instrument, gives a partial and relative picture of students' skills. In particular, PISA does not test French or any national curriculum in reading but the ability of students to understand, use and reflect on written texts in order to achieve their goals, acquire new knowledge and participate in society (Schleicher, 1999). One should also acknowledge that PISA data are not longitudinal (Goldstein, 2004), and especially that they do not include measures of student past cognitive performances. This will make it impossible to disentangle what is related to student's academic

³41 out of the 130 PISA reading items used in the PISA 2009 reading test were taken from the PISA 2000 assessment (OECD, 2010b)

ability before grade repetition or before entering a ZEP middle school from the actual effect of being retained in the same grade or from the actual consequences of belonging to a ZEP school.

Reading performances in 2000 and 2009 are directly comparable across time: the PISA 2000 average score across 28 OECD countries was set at 500 and the standard deviation at 100, establishing the scale against which reading performance in PISA 2009 is compared (OECD, 2010b)⁴. France's average score in reading have substantially decreased over the 2000 decade by more than 10 points (10% of the OECD standard deviation): it amounts to 513.6 points in 2000 and only 503.4 points in 2009 (see Table 1)⁵. The decline in reading average performance results mainly from an increasing share of low-achieving students, although the share of high-achieving students has also slightly extended as well (see Figure 2). The performance dispersion is consequently larger in 2009 than in 2000 (the standard deviation equals 102.1 score points in 2009 and only 87.7 in 2000).

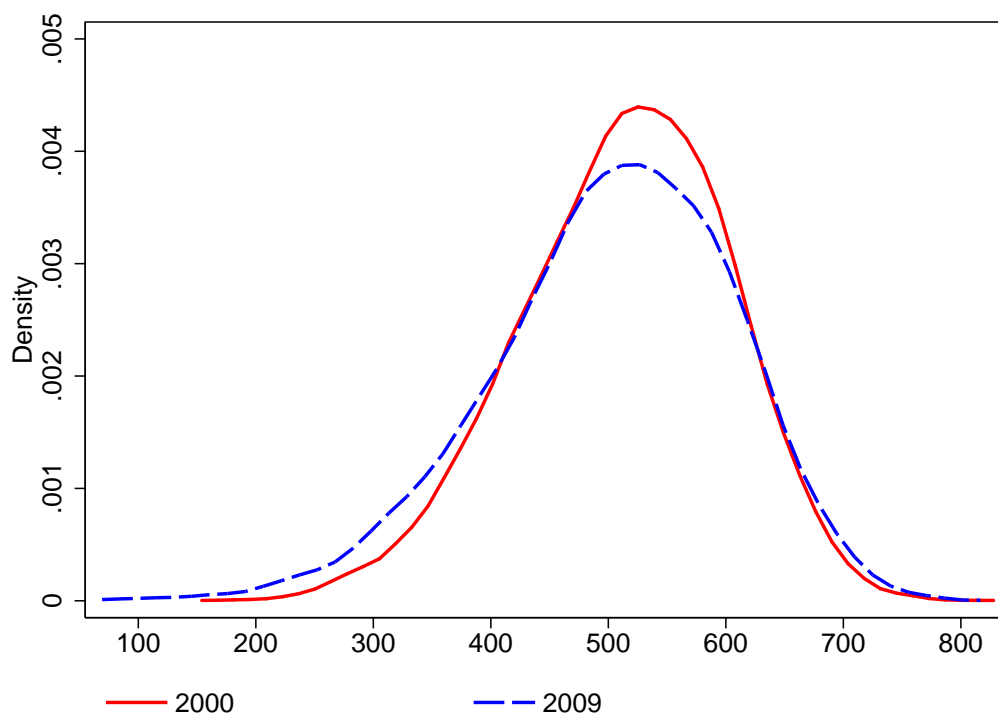
Table 1: Comparison of average test score in reading

	Average test score	
PISA 2000	513.58	(1.38)
PISA 2009	503.38	(1.67)
Difference	-10.21	(2.16)

⁴The PISA 2009 OECD average is 496 in 2009, while the reading performance scale remained unchanged.

⁵The average scores displayed are slightly greater than the ones mentioned in OECD publications, since we computed them using the first plausible value of student's performance, on PISA samples cleaned from missing values for the variables introduced in our decomposition analyses.

Figure 2: Reading Score density in 2000 and 2009



The sampling design used for PISA assessments is a two-stage stratified sample with the first-stage units consisting of schools having 15-year-old students and the second-stage units being students (Wu and Adams, 2002). PISA surveys targets students regardless of the type of educational institutions in which they are enrolled, during the spring of the civil year they turn 16. French PISA data presents two particularities one should be aware of. Firstly, the definition of PISA sampled schools in France generally differs from other countries. Thus 15-year-old students sampled in each school cannot be considered as representative of students enrolled in the same grade level in that school. It is particularly true for students enrolled in sampled middle schools: those students have been retained at least once in the past and they greatly differ in terms of social characteristics and past schooling outcomes from 14-year-old students enrolled in the same grade level. This specificity has some incidence on the estimation strategy. PISA data are generally analyzed with multilevel models allowing for random intercepts at the school level. But in the French case, those models are not necessary since, like linear regression models, they fail to take into account the potential effects that may arise from the way students are assigned to schools. That is the reason why we privilege linear model estimations for performing our decompositions but we also check for multilevel specifications (see Section 5.3).

Secondly, except in 2000, France did not implement the PISA school questionnaire so that the rich information collected on French schools in 2000 cannot be put in perspective with comparable data in 2009. Through the intermediary of the Evaluation, Forecasting and Performance Department of the French Ministry of National Education, we have two additional variables at our disposal for year 2009: school's sector (coded as a dummy variable which equals one if the student is enrolled in a private school) and, more interestingly, school's belonging to

ZEP program (coded as a dummy variable as well).

Table 2 provides descriptive statistics on the samples of students assessed in 2000 and 2009. It shows to what extent the population composition changes between 2000 and 2009. The handling of missing values for variables listed in the table resulted in a sample of 4149 students in 176 schools for year 2000 and of 3957 students in 168 schools in 2009. Observations are weighted with the final student weights.

Since our article aims at assessing whether main changes in the organisation of the French educational system have affected student reading performance, our main variables of interest relate to students' educational characteristics. Student's school track in the French educational system is coded with the help of dummy variables which describe a complete set of exclusive events: being enrolled in the general track of high school, in the vocational track of high school or in middle school. Since 15-year-old students enrolled in middle school who have never been retained are very rare (those students are generally recent migrant students), student's location in middle school is a very good proxy for past retention⁶. As expected, the share of students enrolled in lower-secondary education, that is to say, the share of students who have been retained has decreased by about 3%, but still remains high in 2009 (34.6%, see Table 2). Mechanically the shares of students enrolled in vocational and in general upper-secondary education have respectively increased by 1% and 2%. With respect to special education zones, PISA data clearly show their extension: in 2000 only 5.6% of 15-year-old students are enrolled in ZEP schools, while they are 8.5% in 2009. Accordingly average class size in schools having 15-year-old students has simultaneously decreased by 0.3 student. As expected, the relative sizes of the public and the private sectors have remained the same across the 2000 decade.

Performance distribution moves between 2000 and 2009 may not only be due to changes in the production process of education in France but also to changes in students' social characteristics. Thus we are also willing to account for modifications in student population composition. Among students' personal and familial characteristics that are known to play a major role in students' schooling achievement, we select PISA variables which are common to PISA 2000 and 2009 datasets and which are highly comparable across time: student's month of birth, gender and migration status (coded with the help of three dummy variables: 'both parents are both born abroad', 'one parent is born abroad', and 'language most often spoken at home is not French'). The share of students whose parents are born abroad has slightly increased (by 1%) between 2000 and 2009, and besides the share of students most often speaking another language than French with their family has more enlarged (by 2%).⁷ Conversely we have serious concerns about the temporal comparability of variables related to student's family structure⁸. However, given the important role of family composition in student's achievement (Sassler et al., 2013), we will take the liberty to cautiously introduce the dummy identifying single-parent families as a

⁶In 2000, students have not been interrogated about their past grade retention(s). In 2009, the share of students declaring having repeated a grade once in the past is slightly greater than the share of students enrolled in lower-secondary education. One advantage for considering student's enrollment in lower-secondary education instead of student's declaration of grade repetition is that it allows avoiding any desirability bias.

⁷This is consistent with changes observed by Prioux and Barbieri (2012) in migrant population size from French yearly censuses.

⁸The question statement related to student's home composition in 2000 is ambiguous and has probably led to an underestimation of single-parent families in 2000.

control variable in complementary analysis. We also consider two measures of social background⁹: parents' highest international socioeconomic index (HISEI)¹⁰ which takes values between 0 and 74 and number of books at home¹¹ gathered into four categories (0-10; 11-100; 101-500; more than 500 books). Quite surprisingly, the average socioeconomic index has decreased by 1.6 points in 9 years of time. This might be due to the increasing share of 15-year-old students living in single-parent and in migrant families, but we cannot exclude it to be due to some sampling bias. It is interesting to note that the average number of books at home has decreased as well, in accordance with the digital development.

It is wise not only considering changes in students' social characteristics (for instance an increasing share of migrant students) but also changes in the way students are allocated to schooling contexts according to their characteristics (for instance an increasing schooling segregation of migration students). We consequently built several school composition variables: the share of girls, the share of students with both parents born abroad, the share of students having more than 100 books at home and the average HISEI in the school. To compute the school-level indicators listed in Table 2, we use answers provided by 15-year-old students selected in the PISA sampled schools. As regards the aforementioned definition of PISA schools in France, these composition variables might significantly differ from the actual school composition. In the case of middle schools, composition variables are computed based on answers from retained students who represent a small and special segment of students enrolled in that type of schools. Following education research showing important school composition effects, we will allow ourselves to check the robustness of our results against the cautious introduction of those composition variables.

⁹The index of Economic, Social and Cultural Status which has been used in many secondary analyses of PISA datasets is not present in PISA 2000.

¹⁰The International Socioeconomic Index has been built by Ganzeboom et al. (1992). It assigns to each occupational category a score corresponding to the weighted average of education and income. It originally takes values between 16 and 90 but we rescaled it from 0 to 74 for decomposition purposes.

¹¹Cultural possessions such as books reflect both cultural and financial resources available in the household and are known to have a strong effect on student achievement (Van de Werfhorst and Mijs, 2010).

Table 2: Descriptive statistics

Variables	2000		2009	
Month of birth	5.5	(0.05)	5.7	(0.06)
Girl	52.1	(0.79)	51.9	(0.81)
Single-parent family	10.4	(0.48)	14.8	(0.57)
Immigration Background				
Both parents born abroad	10.8	(0.49)	11.8	(0.52)
One parent born abroad	12.8	(0.53)	12.6	(0.54)
Language most often spoken at home is not French	4.2	(0.32)	6.2	(0.39)
Socioeconomic background				
Parents' HISEI	32.5	(0.27)	30.9	(0.27)
11-100 books	43.8	(0.78)	47.3	(0.81)
101-500 books	36.2	(0.76)	31.6	(0.75)
> 500 books	8.8	(0.45)	7.6	(0.42)
School composition				
Share of students having more than 100 books	45.0	(0.30)	39.2	(0.38)
Average class size	27.5	(0.07)	27.2	(0.09)
School track				
Lower-secondary education	37.3	(0.76)	34.6	(0.78)
Vocational upper-secondary education	7.9	(0.4)	8.9	(0.48)
Special education zones				
	5.6	(0.37)	8.5	(0.46)
Private sector				
	22.2	(0.66)	21.8	(0.67)
Number of schools				
	176		168	
Number of students				
	4149		3957	

4 Decomposition techniques

We analyze changes in students' test scores by using decomposition techniques of mean and quantile scores differentials¹². Decomposition methods assign differences in test scores either to differences in students' and schools' characteristics over time (explained variation) or to changes in returns to those characteristics in terms of tests scores, *i.e.* in the quality of the educational system (unexplained variation). The explained part (composition effect) measures how much students of 2009 would have scored differently, if, given their own returns to characteristics in terms of performances, they had the same characteristics as students of 2000. The unexplained part (return effect) measures how much students of 2000 would have scored differently, if they experienced the production process of schooling of 2009, that is to say the same transformation of inputs into educational performances as students in 2009, given their own characteristics. We can further perform a detailed decomposition: we decompose score gaps into their detailed components based on the multivariate regression of cognitive achievement on many observable variables related to student, family, school and institutional characteristics. The next two subsections present formally the decomposition methods, the classical Oaxaca-Blinder decomposition

¹²Decomposition methods developed by [Oaxaca \(1973\)](#) and [Blinder \(1973\)](#) have already been applied to PISA data to study score differences between countries ([Ammermueller, 2008](#)) and also between different dates in a same country ([Barrera-Orsorio et al., 2011](#); [Gigena et al., 2011](#)).

of the mean and its extension to quantile decomposition.

4.1 The Oaxaca-Blinder decomposition of the mean

We want to decompose the difference in the mean of test score between 2000 and 2009. Let Y denote the test score and T the year of the evaluation with $T = 0$ in 2000 and $T = 1$ in 2009. We observe K characteristics X_1, \dots, X_K which have an impact on Y and we postulate a linear model for Y . Then, we have for each individual i , the following relation

$$Y_{Ti} = \beta_T^0 + \sum_{k=1}^K X_i^k \beta_T^k + \varepsilon_{Ti}, \quad T = 0, 1,$$

where we assume conditionally independent errors ($\mathbb{E}(\varepsilon|X) = 0$). The overall gap $\Delta_O = \mathbb{E}(Y|T = 1) - \mathbb{E}(Y|T = 0)$ can be decomposed as follows

$$\begin{aligned} \Delta_O &= \mathbb{E}(Y|T = 1) - \mathbb{E}(Y|T = 0) \\ &= \mathbb{E}(\mathbb{E}(Y|X)|T = 1) - \mathbb{E}(\mathbb{E}(Y|X)|T = 0) \\ &= \underbrace{(\beta_1^0 - \beta_0^0) + \sum_{k=1}^K \mathbb{E}(X^k|T = 1)(\beta_1^k - \beta_0^k)}_{\Delta_R} + \underbrace{\sum_{k=1}^K (\mathbb{E}(X^k|T = 1) - \mathbb{E}(X^k|T = 0))\beta_0^k}_{\Delta_X}. \end{aligned}$$

Δ_R is the "return effect" or the "unexplained part" which reflects differences in the returns of certain characteristics. Δ_X is the "composition effect" or the "explained part" which reflects differences in the distribution of the characteristics between the two years. We consider two levels of decomposition, the aggregate and the detailed levels of decomposition. The former only divides Δ_O into its two components, the return effect and the composition effect whereas the latter allows us to know the contribution of each individual covariate.

The case of the mean is simple because we can use the law of conditional expectations. When comparing means, we compare $\mathbb{E}(Y|T = 1)$ with $\mathbb{E}(Y|T = 0)$ and equivalently $\mathbb{E}(\mathbb{E}(Y|X)|T = 1)$ with $\mathbb{E}(\mathbb{E}(Y|X)|T = 0)$. To compare quantiles, we do not have this equivalence. We will use the RIF-regression approach developed in [Firpo et al. \(2007, 2009\)](#) to compute partial effects of changes in distribution of covariates on a given functional of the distribution $Y_t|T$. The outline of the method is to provide a linear approximation to a non-linear functional of the distribution. That approximation method allows one to apply the law of iterated expectations to the distributional statistic of interest and thus to compute approximate partial effects of a single covariate on the functional being approximated.

4.2 Quantile decomposition with RIF-regressions

Let ν be the distributional statistic of a distribution function F we are interested in (which can be quantiles, variance, Gini index, ...). We use a RIF (Recentered Influence Function) function whose main property is that its expectation yields the original ν .

$$\int RIF(y; \nu) dF(y) = \nu.$$

Letting $\nu_T = \nu(F_T)$, $T = 0, 1$ we can therefore write the distributional statistics ν_T as an expectation: $\nu_T = \mathbb{E}[RIF(y_T; \nu_T)|T]$. Using the law of iterated expectations, the distributional statistics can also be expressed in terms of expectations of the conditional recentered influence functions.

$$\nu_T = \int \mathbb{E}(RIF(y, \nu)|X = x, T) dF_{X_T}(x)$$

Then assuming that $\mathbb{E}(RIF(y_T, \nu)|X = x, T)$ is linear in X such that

$$\mathbb{E}(RIF(y_T, \nu)|X = x, T) = x\gamma_T^\nu,$$

we have

$$\nu_T = \int x\gamma_T^\nu dF_{X_T}(x) = \mathbb{E}(X|T)\gamma_T^\nu,$$

then the difference between distributional statistics in $T = 0$ and $T = 1$ can be decomposed as

$$\begin{aligned} \Delta_O^\nu &= \nu_1 - \nu_0 \\ &= \int \mathbb{E}(RIF(y_1, \nu_1)|X = x, T = 1) dF_{X_1}(x) - \int \mathbb{E}(RIF(y_0, \nu_0)|X = x, T = 0) dF_{X_0}(x) \\ &= \underbrace{\mathbb{E}(X|T = 1)(\gamma_1^\nu - \gamma_0^\nu)}_{\Delta_R^\nu} + \underbrace{(\mathbb{E}(X|T = 1) - \mathbb{E}(X|T = 0))\gamma_0^\nu}_{\Delta_X^\nu}. \end{aligned} \quad (1)$$

We obtain an expression similar to the classical Oaxaca-Blinder decomposition. However, here, the coefficients γ are the coefficients of the RIF functions and will be estimated with linear regressions of the RIF functions over the covariates¹³. The RIF-regressions of the τ -th quantile of the distribution q_t is $RIF(y, q_\tau) = q_\tau + (\tau - \mathbb{1}(y \leq q_\tau))/f(q_\tau)$ which gives

$$\mathbb{E}(RIF(Y, q_\tau)|X = x) = q_\tau - \frac{1 - \tau}{f(q_\tau)} + \frac{Pr(Y > q_\tau|X = x)}{f(q_\tau)}.$$

To compute the estimated RIF function, we first estimate \hat{q}_τ using its sample analog in the data. Second, we estimate the density at the sample quantile $\hat{f}(\hat{q}_\tau)$ using Kernel estimation. Then we regress the obtained RIF regressions on the covariate to get the OLS estimates $\hat{\gamma}_0^\nu$ and $\hat{\gamma}_1^\nu$ ¹⁴. Finally, we compute the estimation of the decomposition equation 1 using the Jann's Stata package (Jann, 2008).

¹³However, the difference $\gamma_1^\nu - \gamma_0^\nu$ may be contaminated by differences in the distribution of X between the two groups. In a robustness check, we will use a reweighted approach described in appendix to estimate this potential bias.

¹⁴We make the assumption that $Pr(Y > q_\tau|X = x)$ is linear in X and we regress the RIF functions on the observable characteristics (OLS-RIF method). We could also estimate $Pr(Y > q_\tau|X = x)$ with a logit model (RIF-Logit method) or Non-Parametrically (RIF-NP method). The three methods are compared in Firpo et al. (2009) and yield very similar results.

5 Results

5.1 Analyzing the decline in the average performance

Before examining the changes in the whole performance distribution, we pay attention to the decline in the average performance in France between 2000 and 2009. To help analyzing decomposition results, we successively show all the ingredients of the Oaxaca-Blinder decomposition: students' average characteristics have already been commented (see Table 2) and linear regression coefficients are now displayed in Table 3. In the estimated linear model, we set the reference to a male student whose both parents are born in France and who is enrolled in general upper-secondary education in a public high school. He has a low socioeconomic status and no more than 10 books at home and his school does not belong to a special education zone. Note that the average performance of the reference individual has decreased by around 10 points between 2000 and 2009 (so has the overall average reading score). Linear regression results show several differences in returns associated with students' social characteristics, *i.e.* in the educational value of students' social endowments. All other things being equal, girls are found to perform around 8 (=25.3-17.2) points higher than boys in reading in 2009 than in 2000. Whereas the return associated with parents' socioeconomic status has slightly decreased, the returns associated with different volumes of books at home have considerably diverged. It seems that the French educational system values even more classical humanities than before in spite of the digital development (Baudelot and Establet, 2009). Interestingly enough, the penalty associated with an immigrant background has decreased over time, but the prejudice related to speaking another language than French with one's parents has enlarged.

Turning on to returns associated with student educational characteristics, we find that being retained in lower-secondary education (rather than being in general upper-secondary education) is even more detrimental to students in 2009 (-98.4 points) than in 2000 (-87.3). This might result from a lower efficiency of grade retention across time but it can also be due to a selection bias. PISA 2009 students who have been retained, despite official recommendations towards reducing grade repetition, might be less proficient than their counterparts of PISA 2000. Interestingly, returns associated with being enrolled in vocational upper-secondary education are also much more negative in 2009 (-93.5) than in 2000 (-69.4). The promotion of "henceforth promoted" students, who are probably less able than students "always promoted" in the vocational track, might explain part of the decreasing return of vocational education. This could suggest that "henceforth promoted" students have not particularly benefited from their promotion or/and that "always promoted" students have suffered from it. Lastly, the returns associated with being enrolled in a school targeted for special education are more negative in 2009 (-36.1) than in 2000 (-18.9). Given that we control for students characteristics, this might suggest that the ZEP policy is less efficient in 2009 than it was in 2000.

Table 3: Linear regressions of students' performances in reading

Variables	2000		2009	
Month of birth	-0.89	(0.31)	0.01	(0.35)
Boy	<i>Ref.</i>		<i>Ref.</i>	
Girl	17.19	(2.07)	25.34	(2.38)
Immigration background				
Both parents born in France	<i>Ref.</i>		<i>Ref.</i>	
Both parents born abroad	-18.35	(3.95)	-6.20	(4.21)
One parent born abroad	-8.12	(3.18)	-7.05	(3.54)
Language most often spoken at home: French	<i>Ref.</i>		<i>Ref.</i>	
Other language	-13.16	(5.82)	-17.49	(5.89)
Socioeconomic background				
Parents' HISEI	0.70	(0.07)	0.60	(0.08)
<11 books	<i>Ref.</i>		<i>Ref.</i>	
11-100 books	39.44	(3.58)	33.72	(4.03)
101-500 books	54.29	(3.77)	62.05	(4.44)
> 500 books	57.01	(5.03)	84.66	(5.74)
School track				
General upper-secondary education	<i>Ref.</i>		<i>Ref.</i>	
Lower-secondary education	-87.27	(2.40)	-98.42	(3.01)
Vocational upper-secondary education	-69.35	(4.19)	-93.47	(4.74)
Special education zones (ZEP)				
Non-ZEP	<i>Ref.</i>		<i>Ref.</i>	
ZEP	-18.91	(5.30)	-36.06	(5.44)
School sector				
Public	<i>Ref.</i>		<i>Ref.</i>	
Private	4.22	(2.54)	6.44	(2.92)
Intercept	486.5	(4.62)	476.4	(5.22)
R^2	45.82 %		50.47 %	

We now turn on to the Oaxaca-Blinder decomposition results (see Table 4). Recall that the total gap between the average reading scores of 2000 and 2009 amounts to -10 points. This difference is moderate since it is 10% of the international standard deviation. The overall explained part (or the total "composition effect") accounts for -3.4 points in the total gap, meaning that student social and educational characteristics in 2000 were actually more advantageous than in 2009. The overall unexplained part (or the total "return effect") accounts for -6.8 points and explains most of the change in the average performance. It means that the production process of schooling was more efficient in 2000 than in 2009. The detailed decomposition provides us with more information. The negative composition effect is driven by increasing shares of students with disadvantaged socioeconomic background (-3.7), of students with immigration background (-0.4), and of schools targeted for special education (-1.1). The negative return effect is mainly driven by new sorting practices (-6.1) and then by the change in special education program (-1). The increasingly negative effects of grade retention and of vocational education (compared to general education) (-6.1) outweigh the positive composition effect due to the decline in grade

retention (+1,8). The total effect of new sorting policies is consequently negative, meaning that in average, sorting policies in 2000 were more efficient than in 2009. Regarding special education zones, the negative composition effect due to the program extension is reinforced by the negative return effect indicating a lower efficiency of the policy across time.

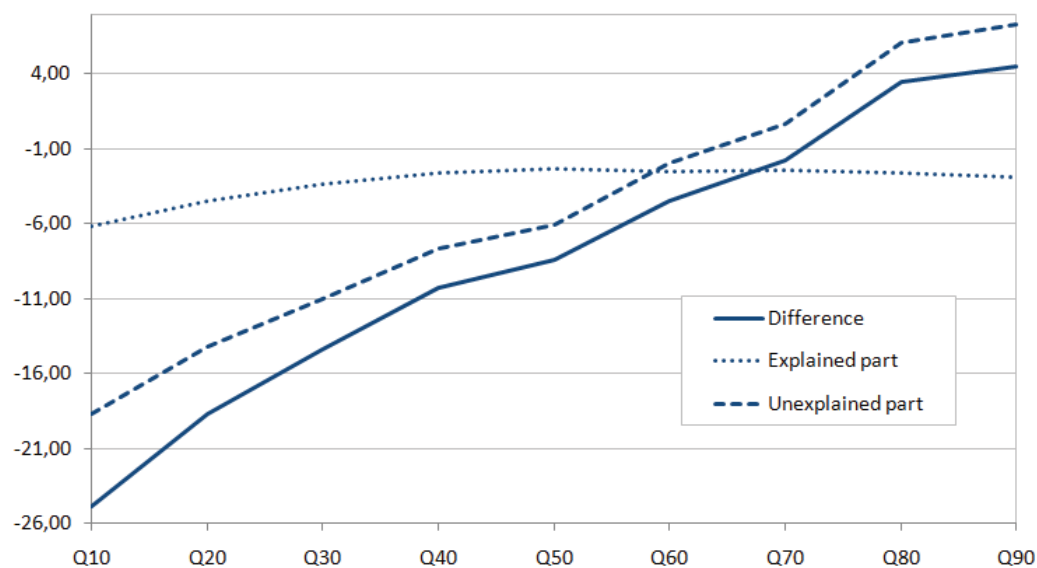
Table 4: Decomposition results

	"Explained" part		"Unexplained" part	
Month of birth	0.00	(0.05)	4.98	(2.57)
Girl	-0.06	(0.29)	4.25	(1.65)
Immigration background	-0.40	(0.17)	1.27	(0.90)
Socioeconomic background	-3.65	(0.66)	-0.73	(5.35)
School track	1.79	(1.10)	-6.07	(1.64)
Special education zones (ZEP)	-1.05	(0.26)	-0.96	(0.43)
Private sector	-0.03	(0.06)	0.49	(0.86)
Intercept			-10.05	(6.97)
Total effects	-3.39	(1.64)	-6.82	(1.60)
Total gap			-10.21	

5.2 Analyzing shifts in the whole performance distribution

So far we only considered the temporal difference in the average scores. However since the average score declined mainly because of a higher share of low-achievers, it is interesting to perform the decomposition along the entire score distribution, and in particular along its lower part. Figure 3 displays the reading score gap between 2000 and 2009 for each decile (solid line). The gap is declining along the deciles of the performance distribution. While it amounts to -25 points for the lowest performing 10 percent students, it equals +4 points for the highest performing decile. Performance inequalities are consequently larger in 2009 than in 2000. The overall difference is broken down into two components: the composition effect (the so-called "explained part") and the return effect (or "unexplained part"). The composition effect explains a small part of the total gap. It is slightly more negative for the lowest two deciles of the performance distribution, and it is quite stable (around -2 points) otherwise. This means that student characteristics have deteriorated comparatively more among the lowest performing 20 percents than among other students. The return effect represents the difference between the distributions that is only due to difference in the quality of the educational system, given the characteristics of students in 2000. It increases almost linearly over the distribution: it is negative along the lowest half of the performance distribution and slightly positive for the highest two deciles. The production process of schooling in 2009 is clearly less efficient for the lowest achieving students in 2009 than it was in 2000, and slightly more efficient for high proficient students. The French educational system seems to be more and more elitist.

Figure 3: Decomposition by decile



We now look at the detailed decomposition to examine the particular effects of our two educational policies of interest: the decline in grade retention and consequently changes in sorting practices on the one hand, and extensions of special education zones on the other hand (complete results of the detailed quantile decomposition are displayed in Table 5). Figure 4 shows the total effect attributed to student school track in the overall score gap and its components. The composition effect is positive and quite stable along the performance distribution, because of the decline in grade retention (see Table 7 in the appendix which displays separate effects of lower-secondary education and of vocational upper-secondary education). However, the return effect is negative for almost all students and in particular for the lowest deciles of the performance distribution. The positive effect of grade retention found for the highest performing 10 percent of the students might proceed from the fact that socially advantaged students and good achievers might use retention to increase their chances of being admitted in the general track (Duru-Bellat et al., 1993; Kloosterman and de Graaf, 2010). Apart from that grade retention seems to be lesser efficient in 2009 than in 2000. Finally, it appears that the total contribution of sorting policies to the decline in low-achieving performances is considerable: Almost one half of the overall 21 points decline in the score of the 20 percent lowest performers can be attributed to changes in sorting practices.

Figure 4: School track effect on reading performance by decile in 2000 and in 2009

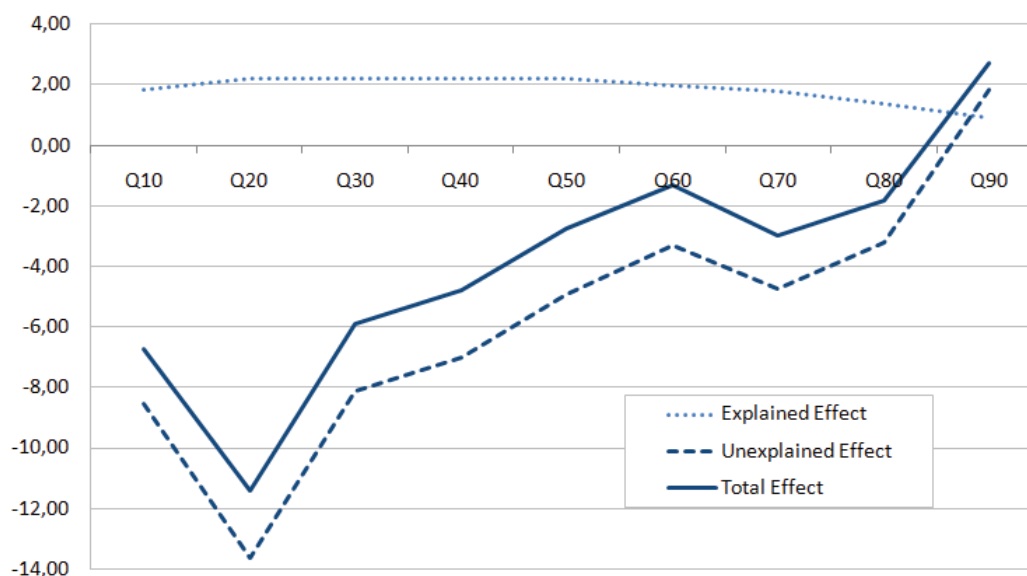


Figure 5 shows the total effect of special education zones and its components. Both the composition and the return effects are particularly negative for the lowest two deciles of the performance distribution: The total effect of the policy is particularly negative for the students who are actually targeted by it! The extension of the policy has probably firstly spread the negative effects related to the ZEP status and secondly led to diluting the resources allocated to disadvantaged schools. Changes in the dimension and the efficiency of the ZEP policy are responsible for almost one third of the total decline in the performances of the lowest two deciles of students (around 6 points out of 21). Finally, around two thirds $((9+6)/21)$ points) of the decline in the score of the 20 percent lowest performers are related to changes in sorting practices and in special education policy.

Figure 5: Special Education Zones effect on reading performance by decile in 2000 and in 2009

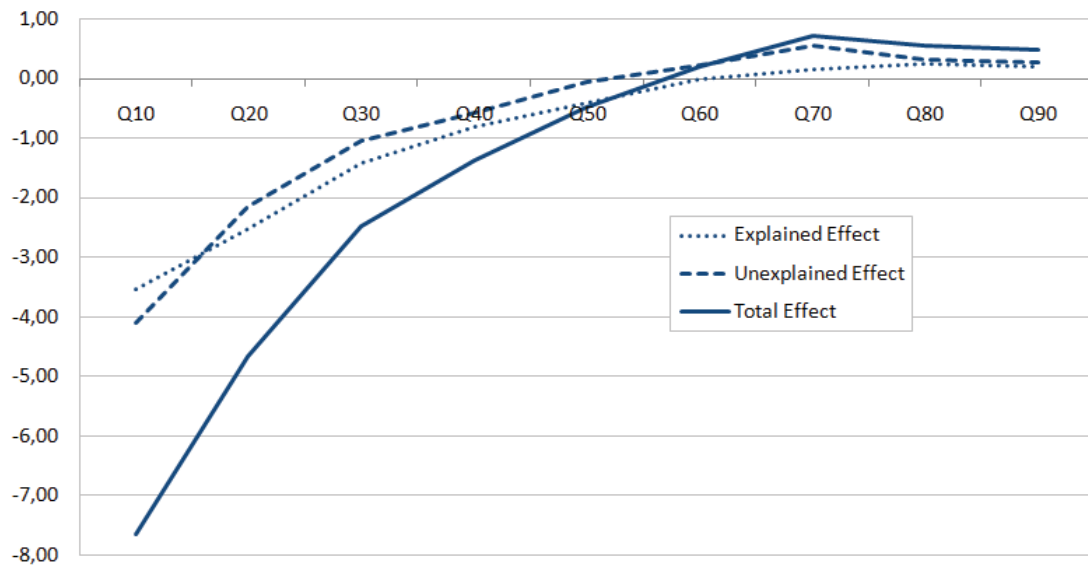


Table 5: Quantile Decomposition without reweighting

	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90									
Prediction 2009	369.47	3.40	419.87	2.69	456.41	1.85	486.48	2.12	510.45	1.96	536.88	1.84	563.35	1.76	592.47	2.40	626.01	2.05
Prediction 2000	394.32	2.61	438.61	2.11	470.84	2.14	496.74	1.86	518.80	1.94	541.37	1.70	564.94	1.66	589.03	1.65	621.60	2.03
Difference	-24.89	4.62	-18.74	3.34	-14.38	2.91	-10.31	2.54	-8.38	2.47	-4.44	2.52	-1.76	2.39	3.43	3.00	4.47	2.83
Explained																		
Month of birth	-0.02	0.17	-0.07	0.12	0.01	0.09	0.00	0.09	0.05	0.10	-0.01	0.08	0.08	0.10	0.04	0.10	-0.02	0.09
Girl	-0.11	0.45	-0.09	0.40	-0.07	0.33	-0.06	0.27	-0.06	0.31	-0.06	0.28	-0.06	0.24	-0.05	0.22	-0.04	0.17
Immigration background	-1.09	0.49	-0.54	0.29	-0.62	0.25	-0.39	0.21	-0.30	0.19	-0.25	0.18	-0.18	0.14	-0.07	0.15	0.02	0.17
Socioeconomic background	-3.20	0.90	-3.48	0.80	-3.46	0.71	-3.54	0.68	-3.78	0.77	-4.10	0.76	-4.21	0.75	-4.17	0.89	-3.85	0.83
School degree and track	1.82	1.26	2.22	1.47	2.22	1.46	2.22	1.53	2.19	1.94	1.97	1.21	1.77	1.07	1.39	0.84	0.89	0.56
Specialized Education	-3.53	0.95	-2.52	0.63	-1.42	0.40	-0.80	0.24	-0.41	0.18	-0.02	0.15	0.16	0.13	0.25	0.12	0.21	0.10
Private sector	-0.02	0.09	-0.01	0.07	-0.02	0.05	-0.03	0.08	-0.01	0.05	-0.04	0.09	-0.04	0.08	-0.02	0.08	-0.07	0.15
Total	-6.16	2.33	-4.49	2.21	-3.36	2.09	-2.61	2.04	-2.32	1.93	-2.49	1.67	-2.46	1.47	-2.64	1.48	-2.86	1.16
Unexplained																		
Month of birth	6.53	6.51	4.47	4.96	6.70	4.07	8.41	3.57	6.68	3.95	3.34	3.45	4.33	3.84	3.97	3.96	3.44	4.51
Girl	7.81	4.13	4.92	3.59	4.70	2.73	3.16	2.43	5.35	2.44	6.51	2.40	3.19	2.43	1.82	2.61	0.71	3.12
Immigration background	4.61	2.58	0.36	1.83	2.10	1.49	1.56	1.27	0.97	1.26	1.30	1.32	0.84	1.23	0.85	1.31	-0.35	1.35
Socioeconomic background	-8.58	19.99	-9.47	12.71	-4.89	10.77	-4.05	7.90	3.82	7.61	7.90	7.18	6.83	6.23	10.74	6.58	8.35	5.78
School degree and track	-8.56	4.33	-13.64	3.98	-8.11	3.89	-7.02	3.57	-4.95	3.31	-3.29	2.84	-4.76	2.68	-3.22	2.34	1.85	2.02
Specialized Education	-4.11	1.48	-2.15	1.09	-1.06	0.71	-0.58	0.53	-0.07	0.41	0.22	0.37	0.55	0.32	0.32	0.32	0.26	0.21
Private sector	-1.06	1.75	-0.88	1.61	-0.55	1.41	0.68	1.28	-0.41	1.30	0.96	1.29	1.76	1.46	0.71	1.40	2.42	1.54
Intercept	-15.38	22.59	2.14	16.51	-9.91	11.51	-9.85	10.28	-17.46	10.02	-18.89	9.37	-12.05	8.58	-9.12	9.22	-9.34	9.03
Total	-18.73	3.86	-14.24	3.06	-11.02	2.34	-7.71	2.16	-6.06	2.36	-1.95	2.09	0.70	2.15	6.07	2.67	7.33	2.54

Standard errors are computed with bootstrap (200 replications)

5.3 Robustness checks

We perform several robustness checks to take into account potential biases resulting from the specification of our decomposition analyses.

First, we check whether decomposition results change if we use five plausible values of student's score instead of the first one, then average the five corresponding decomposition results. Such results are showed for the Oaxaca-Blinder decomposition of the average gap in Table 8. As expected, the aggregate and the detailed decomposition results based on five plausible values are very close to the ones computed with the first plausible value. This allows us to use the first plausible value in our decomposition analyses.

Second, although we argued that multilevel models are not necessary to analyze French PISA data, we check whether performing the mean decomposition based on multilevel models (allowing for random intercepts at the school level) rather than on linear regression models modify our results. As can be seen in Table 9, using multilevel models does not change the sign and the approximate size of the decomposition results. Our results are robust against a multilevel specification of the regression analysis.

Third, we introduce additional variables, which are of lower quality than the ones selected for the main decomposition analyses, but which might play an important role in the performance shift. We first add a dummy indicating the structure of the family (single-parent or nuclear), because the share of single-parent families has increased during the 2000s in France (Chardon et al., 2008). Given that the share of 15-year-old students living in single-parent families is probably overestimated in 2000, the rise in single-parent families is likely underestimated in PISA data. This might downwardly bias the contribution of this variable to the decomposition analysis. We also take into account school composition variables (displayed in Table 2), which are poor proxies of the actual school composition, in order to control for possible changes in the degree of school segregation in France and their consequences on performance distribution. Since ZEP status is attributed to schools on the basis of their population composition, we are also interested in checking how the contribution of special education zones to the performance shift is affected by the introduction of composition variables. In the same vein, we also introduce the average class size of the school. Given that one component of the ZEP policy consists in lowering class size, we check whether the contribution of special education zones is altered when we purge the ZEP policy from one of its component. We perform two additional quantile decompositions based on two extended sets of variables (see Table 6) and we compare them with the quantile decomposition ("Decomposition A") displayed in the core of the article. "Decomposition B" is nothing else than "Decomposition A" with the inclusion of the dummy variable indicating students living in single-parent families and of the average class size in the school. "Decomposition C" is "Decomposition B" extended with school composition variables.

Results of Decompositions B and C are respectively displayed in Table 11 and Table 12. The composition and the return effects as well as the explained and unexplained parts attributed to school track and special education zones are very close to the ones estimated for Decomposition A. Our results are robust against the introduction of those additional variables.

Table 6: Comparisons of decompositions with different sets of variables

Decomposition A	Decomposition B	Decomposition C
Month of birth	Decomposition A	Decomposition B
Girl	+ Single-parent family	+ School composition
Immigration background	+ Class size	
Socioeconomic background		
School track		
Special Education Zones		
Private sector		

Fourth, we perform a threefold decomposition, that is we decompose the overall gap in three terms : a coefficient effect (another form of the "return effect"), the endowment effect (another form of the "composition effect") and an interaction term as follows

$$\begin{aligned}
\Delta_O &= (\beta_1^0 - \beta_0^0) + \underbrace{\sum_{k=1}^K \mathbb{E}(X^k|T=0)(\beta_1^k - \beta_0^k)}_{\Delta_R} + \underbrace{\sum_{k=1}^K (\mathbb{E}(X^k|T=1) - \mathbb{E}(X^k|T=0))\beta_0^k}_{\Delta_X} \\
&= + \underbrace{\sum_{k=1}^K (\mathbb{E}(X^k|T=1) - \mathbb{E}(X^k|T=0))(\beta_1^k - \beta_0^k)}_{Interaction}.
\end{aligned}$$

Results of the threefold decomposition are displayed in Table 10. The interaction term only accounts for -1 point in the total decline in the average score. This confirms us in using twofold decompositions.

Finally, to estimate our model we assumed a linear specification for the schooling process. If this specification is not correct, our estimators may be biased and may depend on the distributions on the covariates in 2000 and in 2009. To limit this bias, we build counterfactual distributions of the scores of individuals in 2000 if they had the characteristics of the individuals of 2009. More formally, let us denote Y_0 the average score in 2000, Y_1 the average score in 2009 and Y^C the counterfactual average score in 2009 if the students of 2009 were evaluated in 2000. Then $\Delta_R = Y_1 - Y^C$ is the return effect, it measures how much students would have performed differently if they experienced the same learning conditions as in 2000 with their own characteristics of 2009. The composition effect is then $\Delta_X = Y^C - Y_0$ which measures how much students in 2009 would score differently if they had the characteristics of the students in 2000 given their estimated returns. We have $\Delta_0 = (Y_1 - Y^C) + (Y^C - Y_0) = \Delta_R + \Delta_X$. So with an estimate of the counterfactual Y^C , one can compute the decomposition. This is a general point that holds for all decompositions, and not only for the mean. The identification of the return effect Δ_R requires some assumptions presented in [Firpo et al. \(2007\)](#) and [Fortin et al. \(2011\)](#) : overlapping support and ignorability. In our setting, it seems reasonable that these assumptions hold. First, there are no observable characteristics which are specific to one sample. Second, the assumption that conditionally on observables, the unobservables are independent of the sample is weaker

than the usual conditional independence assumption. Unobservables characteristics (such as the motivation or the effort of the student) can be correlated with the observable characteristics (such as social background) as long as the correlation is the same in 2000 and in 2009.

We describe in Appendix 1.6 how we build the countrefactual Y_C and how we decompose Δ_0 using reweighted RIF-regressions. We show the results in table 13. Our main conclusions remain valid.

6 Discussion

PISA data show that inequalities in French student performances have increased between 2000 and 2009. Our study aimed at assessing to what extent the rise in inequalities can be attributed to two main educational policies: namely new sorting practices triggered by the decline in grade retention and successive extensions of areas targeted for special help in education. Since the rise in inequalities results from the decline in low-achievers performances, we analyzed the change in PISA test scores by using distributional decompositions and in particular quantile decompositions. We assigned the change in test scores either to differences in students' and schools' characteristics over time (explained variation) or to changes in returns to those characteristics in terms of tests scores *i.e.* in educational system quality (unexplained variation). Our decomposition results show that, French 15 year-old students in 2009 have slightly less advantageous characteristics, and that, except for the highest deciles, they experience lower returns in terms of cognitive performance than their counterparts in 2000. Our findings provide some descriptive evidence that increasing inequalities in 15-year-old cognitive skills in France are related to change in education policies. Besides, around two thirds of the decline in the score of the 20 percent lowest performers are related to changes in sorting practices and in special education policy.

As regards special education zones, we found that both the composition and the return effects are negative for low-achievers. We believe that the extension of the policy has spread the negative signalling effects of the ZEP status and diminished the positive effects of channeled resources because of the resources dilution. Since the ZEP policy mostly concerns middle schools (85% of the ZEP students in PISA data) rather than high schools, the negative contribution of the ZEP policy we found mainly applies to middle school students, that is to students having been retained before the age of 15. It would be interesting to check with another data source whether our finding holds for ZEP students who have not been retained.

Examining more thoroughly the contribution of sorting practices, we found that the returns associated with grade retention are more negative for students aged 15 in 2009 than in 2000. This could be due to a selection effect: PISA 2009 students who have been retained despite of retention restrictions are probably less proficient than PISA 2000 students who have been retained. But, at the same time, the return to being enrolled in the vocational track is also increasingly negative over time. In total, the negative return effect of student's school track outweighs the positive evolution in student composition due to the decline in grade retention. This means that new sorting practices were not beneficial neither to retained students nor to promoted students. It is another step yet to conclude that they were detrimental to them. To do so, we would need measures of students past performances at different times of their schooling trajectory. It would also be interesting to know which grade has been repeated by the

student, since we expect that repeating a grade during primary school has a different impact than repeating a grade during secondary education. We face here one important limit of PISA: the fact that they are cross-sectional only and not longitudinal. In any case we have good reasons to think that the decline in grade retention has not triggered the expected effects. In this respect it seems that the aim of reducing grade repetition has not been associated with another way of managing schooling heterogeneity within classes and in particular of managing students coping with learning difficulties. This could explain the relative failure of this policy.

Bibliography

- A. Ammermueller. Pisa: What makes the difference? In *The Economics of Education and Training*, pages 241–265. Springer, 2008.
- F. Barrera-Orsorio, V. Garcia-Moreno, H. Patrinos, and E.E. Porta. Using the oaxaca-blinder decomposition technique to analyze learning outcomes changes over time: an application to indonesia’s results in pisa mathematics. *World Bank Policy Research Working Paper Series, Vol*, 2011.
- C. Baudelot and R. Estabiet. L’élitisme républicain. *Paris: Seuil*, 2009.
- R. Bénabou, F. Kramarz, and C. Prost. The french zones d’éducation prioritaire: Much ado about nothing? *Economics of Education Review*, 28(3):345–356, 2009.
- A.S. Blinder. Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455, 1973.
- P. Bressoux, F. Kramarz, and C. Prost. Teachers’ training, class size and students’ outcomes: Learning from administrative forecasting mistakes. *The Economic Journal*, 119(536):540–561, 2009.
- J-P. Caille. Les collégiens de zep à la fin des années quatre-vingt-dix: caractéristiques des élèves et impact de la scolarisation en zep sur la réussite. *Éducation et formations*, (61):111–140, 2001.
- O. Chardon, F. Daguet, and E. Vivas. Les familles monoparentales. des difficultés à travailler et à se loger. *Insee première*, 1195, 2008.
- J. Di Nardo, N. Fortin, and T. Lemieux. Labor market institutions and the distribution of wages 1973-1992: a semiparametric approach. *Econometrica*, 64, 1996.
- M. Duru-Bellat, J-P. Jarousse, and . Mingat. Les scolarités de la maternelle au lycée. étapes et processus dans la production des inégalités sociales. *Revue française de sociologie*, 34(1): 43–60, 1993.
- D.N; Figlio and M.E. Page. School choice and the distributional effects of ability tracking: does separation increase inequality? *Journal of Urban Economics*, 51(3):497–514, 2002.
- S. Firpo, N. Fortin, and T. Lemieux. Decomposing wage distributions using recentered influence function regressions. *University of British Columbia (June)*, 2007.
- S. Firpo, N.M. Fortin, and T. Lemieux. Unconditional quantile regressions. *Econometrica*, 77 (3):953–973, 2009.
- N. Fortin, T. Lemieux, and S. Firpo. Decomposition methods in economics. *Handbook of labor economics*, 4:1–102, 2011.
- F. Galindo-Rueda and A. Vignoles. The heterogeneous effect of selection in u.k secondary schools. *Schools and the equal opportunity problem*, pages 103–128, 2007.

- A. Gamoran. Curriculum standardization and equality of opportunity in scottish secondary education: 1984-90. *Sociology of Education*, pages 1–21, 1996.
- H.B.G. Ganzeboom, P.M. De Graaf, and D.J. Treiman. A standard international socio-economic index of occupational status. *Social science research*, 21(1):1–56, 1992.
- M. Gigena, M.L. Vera, R. Giuliadori, and H. Gertel. Exploring the gap difference in 2000-2009 pisa test scores between argentina, chile and mexico. *Regional and Sectoral Economic Studies*, 11(3), 2011.
- H. Goldstein. International comparisons of student attainment: some issues arising from the pisa study. *Assessment in Education: principles, policy & practice*, 11(3):319–330, 2004.
- A. Grisay. *Le fonctionnement des collèges et ses effets sur les élèves de sixième et de cinquième*. Université de Liège and Ministère de l'éducation nationale (1993-1995). Direction de l'évaluation et de la prospective. France, 1993.
- G.B. Jackson. The research evidence on the effects of grade retention. *Review of Educational Research*, pages 613–635, 1975.
- M. Jakubowski, H.A. Patrinos, E.E. Porta, and J. Wisniewski. The impact of the 1999 education reform in poland. 2010.
- B. Jann. The blinder-oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4):453–479, 2008.
- S.R. Jimerson, S.M.W. Pletcher, K. Graydon, B.L. Schnurr, A.B. Nickerson, and D.K. Kundert. Beyond grade retention and social promotion: Promoting the social and academic competence of students. *Psychology in the Schools*, 43(1):85–97, 2006.
- J.O. Jonsson and R. Erikson. Understanding educational inequality: the swedish experience. *L'Année sociologique*, 50(2):345–382, 2000.
- R. Kloosterman and P.M. de Graaf. Non-promotion or enrolment in a lower track? the influence of social background on choices in secondary education for three cohorts of dutch pupils. *Oxford Review of Education*, 36(3):363–384, 2010.
- N. Le Donné. European variations in socioeconomic inequalities in students' cognitive achievement: The role of educational policies. *European Sociological Review*, page jcu040, 2014.
- J. Levasseur and C. Seibel. les apprentissages instrumentaux et le passage du cours préparatoire au cours élémentaire. *Education et Formations*, (83), 1985.
- A. Manning and J.S. Pischke. *Comprehensive versus Selective Schooling in England and Wales: What Do We Know?* CEE DP 66. ERIC, 2006.
- E. Maurin. *Le ghetto français: enquête sur le séparatisme social*. Seuil, 2004.
- E. Maurin. La nouvelle question scolaire. *Les bénéfices de la démocratisation*, Paris, Le Seuil, 2007.

- MEN. Repères et références statistiques sur les enseignements, la formation et la recherche. 1998.
- MEN. Repères et références statistiques sur les enseignements, la formation et la recherche. 2009.
- P. Merle. *La ségrégation scolaire*. Découverte, 2012.
- D. Meuret. L'efficacité de la politique des zones d'éducation prioritaire dans les collèges. *Revue française de pédagogie*, pages 41–64, 1994.
- C. Moisan and J. Simon. *Les déterminants de la réussite scolaire en zone d'éducation prioritaire*. Institut national de recherche pédagogique, 1998.
- N. Mons. *Les nouvelles politiques éducatives: la France fait-elle les bons choix?* Presses Universitaires de France-PUF, 2007.
- R. Oaxaca. Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709, 1973.
- OECD. Pisa 2003 key findings. what makes a school system perform? seeing school system through the prism of pisa. 2004.
- OECD. Pisa 2003 data analysis manual: Sas users. 2005.
- OECD. Pisa 2009 key findings. what makes a school successful?, vol.4. 2010a.
- OECD. Pisa 2009 key findings. learning trends: Changes in student performance since 2000, vol.5. 2010b.
- M. Palme and C. Meghir. Assessing the effects of schooling on wages using a social experiment. *American Economic Review*, 95(1):414, 2005.
- T. Pekkarinen, R. Uusitalo, and S. Kerr. School tracking and intergenerational income mobility: Evidence from the finnish comprehensive school reform. *Journal of Public Economics*, 93(7): 965–973, 2009.
- T. Piketty and M. Valdenaire. *L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français: estimations à partir du panel primaire 1997 et du panel secondaire 1995*. Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche, Direction de l'évaluation et de la prospective. France, 2006.
- F. Prioux and M. Barbieri. L'évolution démographique récente en france: une mortalité relativement faible aux grands âges. *Population (00324663)*, 67(4), 2012.
- A.J. Reynolds. Grade retention and school adjustment: An explanatory analysis. *Educational Evaluation and Policy Analysis*, 14(2):101–121, 1992.
- S. Sassler, K. Williams, F.R. Addo, A.M. Frech, and E.C. Cooksey. Family structure and high school graduation: how children born to unmarried mothers fare. *Genus*, 69(2), 2013.
- A. Schleicher. *Measuring student knowledge and skills: a new framework for assessment*. OECD Publishing, 1999.

H.G. Van de Werfhorst and J.J.B. Mijs. Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology*, 36:407–428, 2010.

A. Van Zanten and J-P. Obin. *La carte scolaire*. Presses universitaires de France, 2008.

M. Wu and R. Adams. Pisa 2000 technical report, 2002.

1 Appendix

1.1 Grade and track : separate effect of retention and vocational track

Table 7

	Q10		Q20		Q30		Q40		Q50		Q60		Q70		Q80		Q90	
Explained																		
Low-sec track	2.69	<i>1.10</i>	3.47	<i>1.42</i>	3.49	<i>1.42</i>	3.46	<i>1.38</i>	3.37	<i>1.32</i>	3.08	<i>1.20</i>	2.73	<i>1.05</i>	2.14	<i>0.81</i>	1.40	<i>0.58</i>
Voc. track	-0.88	<i>0.61</i>	-1.25	<i>0.81</i>	-1.28	<i>0.89</i>	-1.24	<i>0.82</i>	-1.18	<i>0.80</i>	-1.10	<i>0.62</i>	-0.96	<i>0.60</i>	-0.76	<i>0.49</i>	-0.50	<i>0.34</i>
Unexplained																		
Low-sec track	-5.77	<i>3.76</i>	-8.91	<i>3.35</i>	-4.41	<i>3.04</i>	-4.43	<i>3.14</i>	-3.11	<i>2.76</i>	-1.84	<i>2.39</i>	-3.27	<i>2.36</i>	-2.55	<i>1.91</i>	1.67	<i>1.79</i>
Voc. track	-2.79	<i>1.97</i>	-4.73	<i>1.33</i>	-3.70	<i>1.02</i>	-2.59	<i>0.99</i>	-1.84	<i>0.81</i>	-1.46	<i>0.67</i>	-1.49	<i>0.63</i>	-0.67	<i>0.45</i>	0.17	<i>0.40</i>

1.2 Robustness check 1: five plausible values

Table 8: Decomposition results with five plausible values of student's score

	"Explained" part	"Unexplained" part
Month of birth	-0.01	5.33
Girl	-0.06	4.47
Immigration background	-0.38	1.41
Socioeconomic background	-3.60	-1.35
School track	1.84	-6.73
Special Education Zones	-1.04	-0.95
Private sector	-0.03	-0.65
Intercept		-9.73
Total effects	-3.27	-6.90
Total gap		-10.18

1.3 Robustness check 2: multilevel model

Table 9: Decomposition results with multilevel regressions

	"Explained" part	"Unexplained" part
Month of birth	-0.12	5.26
Girl	-0.03	4.10
Immigration background	-0.42	0.74
Socioeconomic background	-2.47	-10.55
School track	1.09	-21.21
Special Education Zones	-1.09	-0.05
Private sector	-0.04	-0.16
Intercept		18.02
Total effects	-3.08	-5.48
Total gap		-8.56

Note: The total gap estimated with the multilevel model (-8.56) differs from the total gap estimated with the linear model (-10.21). This is due to the shrinkage factor introduced in the estimation of multilevel models OECD (2005).

1.4 Robustness check 3: threefold decomposition

Table 10: Decomposition results with threefold decomposition

	Endowments		Coefficients		Interaction	
Month of birth	-0.14	(0.08)	4.98	(2.57)	0.14	(0.10)
Girl	-0.04	(0.19)	4.25	(1.65)	-0.02	(0.09)
Immigration background	-0.43	(0.20)	1.27	(0.90)	0.03	(0.17)
Socioeconomic background	-2.94	(0.55)	-0.73	(5.35)	-0.71	(0.31)
School track	1.72	(0.95)	-6.07	(1.64)	0.07	(0.20)
Special education zones (ZEP)	-0.55	(0.19)	-0.96	(0.43)	-0.50	(0.24)
Private sector	-0.02	(0.04)	0.49	(0.86)	-0.01	(0.03)
Intercept			-10.05	(6.97)		
Total effects	-2.39	(1.40)	-6.82	(1.60)	-1.00	(0.05)
Total gap			-10.21			

1.5 Robustness check 4: additional variables

Table 11: Decomposition B

	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90									
Prediction 2009	369.44	3.59	419.87	2.81	456.43	2.01	486.45	1.92	510.40	2.04	536.91	2.03	563.36	1.69	592.46	2.25	626.02	2.09
Prediction 2000	394.34	2.68	438.60	1.96	470.81	2.09	496.77	1.74	518.78	1.47	541.35	1.54	565.12	1.63	589.03	1.84	621.55	2.08
Difference	-24.89	4.57	-18.73	3.49	-14.38	2.94	-10.31	2.51	-8.38	2.46	-4.44	2.75	-1.76	2.38	3.43	2.75	4.47	2.95
Explained																		
Month of birth	-0.01	0.16	-0.04	0.14	0.04	0.08	0.03	0.10	0.08	0.09	0.02	0.09	0.11	0.10	0.06	0.08	-0.01	0.10
Girl	-0.10	0.46	-0.08	0.32	-0.07	0.32	-0.05	0.22	-0.06	0.26	-0.05	0.24	-0.05	0.23	-0.04	0.21	-0.04	0.16
Single-parent family	0.68	0.41	0.56	0.32	0.51	0.27	0.49	0.22	0.50	0.20	0.63	0.23	0.49	0.21	0.19	0.23	0.19	0.23
Class size	-0.60	0.38	-1.23	0.45	-1.61	0.50	-1.85	0.54	-2.04	0.64	-1.83	0.60	-1.49	0.51	-1.41	0.49	-1.15	0.44
Immigration background	-1.07	0.53	-0.52	0.33	-0.60	0.25	-0.37	0.21	-0.27	0.17	-0.23	0.19	-0.16	0.15	-0.06	0.16	0.03	0.16
Socioeconomic background	-3.04	0.85	-3.12	0.69	-2.98	0.59	-2.98	0.53	-3.16	0.55	-3.55	0.67	-3.76	0.73	-3.73	0.84	-3.50	0.83
School track	1.74	1.07	2.06	1.23	2.00	1.22	1.97	1.08	1.92	1.06	1.73	0.93	1.57	0.92	1.19	0.64	0.74	0.39
Special Education Zones	-3.26	0.88	-1.95	0.64	-0.67	0.32	0.07	0.20	0.55	0.21	0.84	0.23	0.86	0.25	0.91	0.24	0.75	0.20
Private sector	-0.03	0.07	-0.01	0.07	-0.02	0.06	-0.03	0.09	-0.01	0.05	-0.04	0.10	-0.04	0.09	-0.03	0.08	-0.07	0.16
Total	-5.69	2.23	-4.35	2.23	-3.39	2.04	-2.73	1.83	-2.49	1.84	-2.48	1.75	-2.47	1.69	-2.92	1.54	-3.05	1.23
Unexplained																		
Month of birth	6.96	6.08	5.12	4.79	7.53	4.00	9.34	3.95	7.68	3.49	4.32	3.70	5.12	3.48	4.62	3.70	3.97	4.64
Girl	8.47	4.27	5.44	3.49	5.03	2.70	3.16	2.51	5.25	2.30	6.32	2.39	3.08	2.52	1.62	2.26	0.49	2.75
Single-parent family	2.01	1.24	1.45	0.91	2.02	0.84	1.33	0.73	0.71	0.64	1.32	0.67	1.08	0.67	0.34	0.74	-0.35	0.80
Class size	-83.08	34.11	-53.03	27.78	-28.56	24.71	11.73	19.74	26.94	21.87	27.84	20.42	18.50	18.96	35.83	17.71	36.72	19.01
Immigration background	4.69	2.60	0.43	1.97	2.12	1.52	1.52	1.36	0.93	1.23	1.22	1.24	0.79	1.21	0.81	1.37	-0.40	1.48
Socioeconomic background	-5.34	19.72	-8.76	12.68	-5.95	9.98	-7.45	7.29	-0.65	7.48	3.82	6.95	3.70	6.57	6.78	6.53	4.74	5.93
School track	-16.43	5.45	-18.88	5.09	-11.35	4.31	-6.55	4.08	-3.03	3.62	-1.31	3.24	-3.54	3.09	-0.35	2.68	4.99	2.71
Special Education Zones	-4.22	1.62	-1.76	1.04	-0.36	0.77	0.45	0.56	1.16	0.46	1.34	0.43	1.43	0.35	1.25	0.36	1.08	0.29
Private sector	-2.32	1.80	-2.26	1.67	-1.95	1.44	-0.47	1.46	-1.54	1.28	-0.01	1.21	0.93	1.28	0.11	1.31	2.01	1.63
Intercept	70.03	42.60	57.88	35.37	20.48	29.41	-20.65	23.13	-43.34	24.88	-46.81	22.86	-30.38	21.55	-44.67	19.49	-45.74	23.51
Total	-19.20	4.29	-14.38	3.13	-10.99	2.30	-7.58	2.14	-5.88	2.04	-1.96	2.23	0.71	2.05	6.35	2.49	7.52	2.72

Table 12: Decomposition C

	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90	
Prediction 2009	369.44	419.87	456.43	486.45	510.40	536.91	563.36	592.46	626.02	2.09
Prediction 2000	394.34	438.60	470.81	496.77	518.78	541.35	565.12	589.03	621.55	2.08
Difference	-24.89	-18.73	-14.38	-10.31	-8.38	-4.44	-1.76	-3.43	-4.47	2.98
Explained										
Month of birth	-0.02	0.14	0.02	0.01	0.08	0.08	0.09	0.03	-0.03	0.10
Girl	-0.10	0.47	-0.07	-0.06	-0.06	-0.05	-0.05	-0.04	-0.04	0.18
Single-parent family	0.63	0.43	0.31	0.44	0.45	0.59	0.44	0.13	0.13	0.26
Class size	-0.17	0.41	0.39	-1.38	-1.39	-1.13	-0.90	-0.71	-0.41	0.27
Immigration background	-1.09	0.47	0.34	-0.35	-0.24	-0.17	-0.11	-0.03	0.06	0.20
Socioeconomic background	-2.64	0.90	0.65	-2.37	-2.41	-2.82	-3.12	-2.99	-2.70	0.72
School track	1.67	0.97	1.14	1.88	1.01	1.57	1.44	1.06	0.57	0.28
Special Education Zones	-3.45	1.05	-1.88	0.10	0.63	0.95	0.94	0.84	0.76	0.20
Private sector	-0.01	0.08	0.08	-0.01	0.03	0.07	-0.01	0.00	-0.03	0.08
School composition	-2.88	1.09	-4.30	-3.99	-5.06	-5.22	-4.64	-5.36	-5.47	1.10
Total	-8.07	2.35	2.17	-5.73	-6.23	1.90	2.07	-7.06	-7.17	1.48
Unexplained										
Month of birth	6.50	6.03	5.31	8.71	6.79	3.41	4.30	3.73	3.13	4.67
Girl	7.12	3.99	3.67	4.91	6.32	7.26	3.33	1.15	0.28	2.99
Single-parent family	1.77	1.40	1.00	1.22	0.59	1.25	1.00	0.20	-0.50	0.92
Class size	-110.34	41.70	36.04	15.39	12.14	17.74	19.58	25.80	24.36	24.42
Immigration background	2.76	2.58	1.79	1.50	1.17	1.85	1.35	0.60	-0.86	1.52
Socioeconomic background	-7.74	19.41	12.62	-8.54	-3.61	2.02	4.93	6.71	2.82	7.02
School track	-16.84	5.52	4.81	-6.69	-2.43	-0.13	-3.74	0.36	6.05	2.83
Special Education Zones	-5.46	1.80	1.22	0.38	1.27	1.53	1.46	0.78	0.53	0.34
Private sector	-1.53	2.13	2.01	0.42	-1.41	0.29	2.06	1.38	2.83	1.80
School composition	32.43	29.15	23.76	-27.00	0.63	-6.34	-24.95	-15.85	-1.43	22.92
Intercept	74.50	51.78	37.52	5.12	-23.60	-27.02	-5.18	-14.36	-25.58	22.15
Total	-16.83	3.72	3.03	-4.58	-2.15	1.86	4.14	10.49	11.63	3.07

1.6 Robustness check 5: reweighed RIF-regressions

Principle

In this subsection we are interested in computing the counterfactual Y^C to estimate $\Delta_0 = (Y_1 - Y^C) + (Y^C - Y_0) = \Delta_R + \Delta_X$.

Let ν be the distributional statistic of a distribution function F we are interested in. We denote F_{Y_T} the distribution function of Y in year T where $T = 0$ in 2000 and $T = 1$ in 2009. We denote $F_{Y_0^C}$ the counterfactual distribution function which corresponds to the distribution function of scores in group 0 if individuals of group 0 had the characteristics of group 1.

Proposition (Identification of the aggregate composition)

Under the assumptions of a simple counterfactual, overlapping support and ignorability, for every distributional statistic ν , the overall ν -score gap, Δ_O^ν can be written as

$$\Delta_O^\nu = \Delta_R^\nu + \Delta_X^\nu$$

where

- (i) The return effect term $\Delta_R^\nu = \nu(F_{Y_1}) - \nu(F_{Y_0^C, X=X_1|T_1})$ solely reflects difference between the production process of schooling in $T = 1$ and $T = 0$.
- (ii) The composition effect term $\Delta_X^\nu = \nu(F_{Y_0^C, X=X_1|T_1}) - \nu(F_{Y_0})$ solely reflects the effect of differences in the distribution of characteristics (observable (X) and unobservable (ε)) between the two groups.

There are three reasons why scores can be different between group 0 and group 1 : the score setting can be different in $T = 1$ and $T = 0$, the distributions of X can be different or the distributions of ε are different. The ignorability assumption only states that conditional on X , the distribution of ε is the same in both groups. So once we control for differences between the X 's in the two groups, we also control for differences in the ε 's.

Constructing a counterfactual distribution with reweighting

The distribution of Y_T is defined using the law of iterated probabilities, that is after we integrate over the observed characteristics we obtain

$$F_{Y_T}(y) = \int F_{Y_T|X_T}(y|X = x).dF_{X_T}(x), \quad T = 0, 1.$$

We can construct the counterfactual distribution as follows

$$F_{Y_0^C}(y) = \int F_{Y_0|X_0}(y|X = x).dF_{X_1}(x).$$

The idea is to integrate group 0's conditional distribution of Y given X over group 1's distribution of X . We will follow the reweighting approach of [Di Nardo et al. \(1996\)](#). They compute a

reweighting factor $\Psi(X) = \frac{dF_{X_1}(X)}{dF_{X_0}(X)}$ to obtain

$$F_{Y_0^C}(y) = \int F_{Y_0|X_0}(y|X)\Psi(X)dF_{X_0}(x)$$

Let $p = Pr(T = 1)$ and $p(x) = Pr(T = 1|X = x)$. The reweighting factor can be written

$$\Psi(X) = \frac{Pr(X|T = 1)}{Pr(X|T = 0)} = \frac{Pr(T = 1|X)/Pr(T = 1)}{Pr(T = 0|X)/Pr(T = 0)} = \left(\frac{p(X)}{1 - p(X)} \right) \left(\frac{1 - p}{p} \right)$$

We will use the RIF-regressions for the detailed decomposition. Letting $\nu_T = \nu(F_T)$ and $\nu_C = \nu(F_C)$, we can therefore write the distributional statistics ν_0 , ν_1 , and ν_C as expectations: $\nu_T = E[RIF(y_T; \nu_T)|T]$, and $\nu_C = E[RIF(y_0; \nu_C)|T = 1]$. Then we obtain

$$\begin{aligned} \Delta_R^\nu &= \nu_1 - \nu_C \\ &= \int E(RIF(y_1, \nu_1)|X = x, T = 1)dF_{X_1}(x) - \int E(RIF(y_0, \nu_C)|X = x, T = 1)dF_{X_1}(x) \\ &= E(X|T = 1)(\gamma_1^\nu - \gamma_C^\nu), \end{aligned}$$

$$\begin{aligned} \Delta_X^\nu &= \nu_C - \nu_0 \\ &= \int E(RIF(y_0, \nu_C)|X = x, T = 1)dF_{X_1}(x) - \int E(RIF(y_0, \nu_0)|X = x, T = 0)dF_{X_0}(x) \\ &= E(X|T = 1)\gamma_C^\nu - E(X|T = 0)\gamma_0^\nu \end{aligned}$$

and

$$\Delta_O^\nu = \underbrace{E(X|T = 1)(\gamma_1^\nu - \gamma_C^\nu)}_{\Delta_R^\nu} + \underbrace{E(X|T = 1)\gamma_C^\nu - E(X|T = 0)\gamma_0^\nu}_{\Delta_X^\nu}.$$

One difference with the Oaxaca-Blinder decomposition is that the coefficient γ_C^ν (the regression coefficient when the group 0 data is reweighted to have the same X distribution as the group 1) is used instead of γ_0^ν (the unadjusted coefficient for group 0). The reason for using γ_C^ν instead of γ_0^ν is that the difference $\gamma_1^\nu - \gamma_C^\nu$ solely reflects differences between the returns structures while the difference $\gamma_1^\nu - \gamma_0^\nu$ may be contaminated by differences in the distribution of X between the two groups. If the linear approximation is correct then $\gamma_C^\nu = \gamma_0^\nu$, and we obtain the usual Oaxaca-Blinder decomposition.

Estimation

We need to estimate the different decomposition elements that we have just introduced: ν_1 , ν_0 , ν_C , γ_1 , γ_0 and γ_C . For ν_1 , ν_0 , γ_1 and γ_0 , the estimation is very standard because the distributions F_1 and F_0 are directly identified from data on (Y, T, X) . The distributional statistics ν can be estimated using their sample analog in the data, while the γ 's are estimated using ordinary least square methods. However, to estimate ν_C and γ_C , we need to estimate the weighting function $\psi(X)$. The estimation strategy proceeds in three steps : first we estimate the weights, then the

distributional statistics, and finally the estimates of the RIF-regressions. We apply the setting to the quantile distributions. To estimate the weight, we estimate the reweighting factor as

$$\hat{\psi}(X) = \left(\frac{1 - \hat{p}}{\hat{p}} \right) \left(\frac{\hat{p}(X)}{1 - \hat{p}(X)} \right),$$

where $\hat{p}(\cdot)$ is an estimator of the true probability of being in group 1 given X and is estimated with a probit model. We multiply the sample weights by the reweighting factor to obtain the final weights and run the regressions. We estimate ν_0 , ν_1 and ν_C using their sample analog. We compute the distributional statistics $\hat{\nu}_T = \nu(\hat{F}_T)$ and $\hat{\nu}_C = \nu(\hat{F}_C)$ directly from the observations (with appropriated weights). Then, we can compute the return and the composition effects as $\hat{\Delta}_R^\nu = \hat{\nu}_1 - \hat{\nu}_C$ and $\hat{\Delta}_X^\nu = \hat{\nu}_C - \hat{\nu}_0$.

Table 13: Model A with reweighting

	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90	
Prediction 2009	-394.34	2.80	-438.60	1.92	-470.81	2.06	-496.77	1.84	-518.78	1.41
Prediction 2000	-369.44	3.43	-419.87	2.38	-456.43	2.04	-486.45	2.17	-510.40	2.07
Difference	-24.89	4.28	-18.73	3.10	-14.38	3.03	-10.31	2.97	-8.38	2.50
Explained										
Month of birth	0.44	4.43	4.97	5.25	2.00	4.58	4.30	3.29	0.44	3.56
Girl	-2.58	2.62	-1.90	3.09	-3.65	2.80	-4.24	2.29	-4.53	2.38
Immigration background	2.50	1.99	4.07	1.76	3.33	1.75	2.70	1.69	3.14	1.70
Socioeconomic background	-18.95	12.25	-9.82	9.60	-9.57	10.11	-14.68	9.34	-12.84	9.00
School degree and track	2.67	2.94	3.71	2.78	2.08	3.38	5.44	2.93	7.72	4.19
Specialized Education	-1.39	1.14	-1.43	1.08	-1.91	0.83	-1.61	0.54	-0.93	0.43
Private sector	3.37	1.30	3.44	1.45	3.75	1.64	3.29	1.26	2.96	1.31
Total	1.04	3.35	-1.44	2.84	0.28	3.22	0.66	2.82	3.01	2.28
Unexplained										
Month of birth	6.06	5.75	-0.57	7.13	4.71	5.16	4.11	4.11	6.29	4.50
Girl	10.29	4.37	6.73	3.88	8.28	3.49	7.35	2.68	9.82	3.00
Immigration background	1.02	2.66	-4.25	2.22	-1.85	2.07	-1.54	2.10	-2.47	1.91
Socioeconomic background	7.17	18.32	-3.13	13.37	1.22	12.42	7.09	11.28	12.87	9.09
School degree and track	-9.41	3.93	-15.13	3.94	-7.97	4.23	-10.24	4.02	-10.48	4.75
Specialized Education	-6.25	2.18	-3.25	1.63	-0.57	1.26	0.22	0.80	0.46	0.62
Private sector	-4.45	2.11	-4.33	1.83	-4.32	1.74	-2.64	1.50	-3.38	1.65
Intercept	-30.37	22.14	6.63	16.73	-14.16	14.68	-15.32	12.16	-24.50	10.57
Total	-25.94	4.07	-17.30	3.38	-14.66	3.24	-10.97	2.79	-11.38	2.51

	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90	
Prediction 2009	-394.34	2.80	-438.60	1.92	-470.81	2.06	-496.77	1.84	-518.78	1.41
Prediction 2000	-369.44	3.43	-419.87	2.38	-456.43	2.04	-486.45	2.17	-510.40	2.07
Difference	-24.89	4.28	-18.73	3.10	-14.38	3.03	-10.31	2.97	-8.38	2.50
Explained										
Month of birth	0.44	4.43	4.97	5.25	2.00	4.58	4.30	3.29	0.44	3.56
Girl	-2.58	2.62	-1.90	3.09	-3.65	2.80	-4.24	2.29	-4.53	2.38
Immigration background	2.50	1.99	4.07	1.76	3.33	1.75	2.70	1.69	3.14	1.70
Socioeconomic background	-18.95	12.25	-9.82	9.60	-9.57	10.11	-14.68	9.34	-12.84	9.00
School degree and track	2.67	2.94	3.71	2.78	2.08	3.38	5.44	2.93	7.72	4.19
Specialized Education	-1.39	1.14	-1.43	1.08	-1.91	0.83	-1.61	0.54	-0.93	0.43
Private sector	3.37	1.30	3.44	1.45	3.75	1.64	3.29	1.26	2.96	1.31
Total	1.04	3.35	-1.44	2.84	0.28	3.22	0.66	2.82	3.01	2.28
Unexplained										
Month of birth	6.06	5.75	-0.57	7.13	4.71	5.16	4.11	4.11	6.29	4.50
Girl	10.29	4.37	6.73	3.88	8.28	3.49	7.35	2.68	9.82	3.00
Immigration background	1.02	2.66	-4.25	2.22	-1.85	2.07	-1.54	2.10	-2.47	1.91
Socioeconomic background	7.17	18.32	-3.13	13.37	1.22	12.42	7.09	11.28	12.87	9.09
School degree and track	-9.41	3.93	-15.13	3.94	-7.97	4.23	-10.24	4.02	-10.48	4.75
Specialized Education	-6.25	2.18	-3.25	1.63	-0.57	1.26	0.22	0.80	0.46	0.62
Private sector	-4.45	2.11	-4.33	1.83	-4.32	1.74	-2.64	1.50	-3.38	1.65
Intercept	-30.37	22.14	6.63	16.73	-14.16	14.68	-15.32	12.16	-24.50	10.57
Total	-25.94	4.07	-17.30	3.38	-14.66	3.24	-10.97	2.79	-11.38	2.51

Chapter 4

Grade Retention and Unobserved Heterogeneity

Abstract

We study the treatment effect of grade retention, using a panel of French junior high-school students, taking unobserved heterogeneity and the endogeneity of grade repetitions into account. We specify a multi-stage model of human-capital accumulation with a finite number of types representing unobserved individual characteristics. Class-size and latent student-performance indices are assumed to follow finite mixtures of normal distributions. Grade retention may increase or decrease the student's knowledge capital in a type-dependent way. Our estimation results show that the Average Treatment effect on the Treated (ATT) of grade retention on test scores is small but positive at the end of grade 9. The ATT of grade retention is higher for the weakest students. We also show that class size is endogenous and tends to increase with unobserved student ability. The Average Treatment Effect (ATE) of grade retention is negative, again with the exception of the weakest group of students. Grade repetitions reduce the probability of access to grade 9 of all student types.

1 Introduction

Grade retention practices are common in the schools of some countries but absent from others. Some educational systems have been designed to play the role of public certification agencies. If this is the case, a student is promoted to the next grade only if her (his) test scores are sufficiently high, and the students who can't pass are tracked or retained. France and Germany are good instances of such systems, in which grade retention is familiar. In contrast, social promotion, that is, the practice of passing students to the next grade, regardless of school performance, seems to prevail in more egalitarian societies, or in countries promoting mass education. Scandinavian countries and the UK are good instances of the latter system. At the same time, grade retention is a form of second-best remedial education; in some countries it is the main if not the only form of remedial education, but it entails substantial costs. Grade repetitions consume resources, since they permanently increase the stock of enrolled students. There are opportunity costs, since grade repeaters could become productive sooner or have a longer productive life. There also exists substantial costs in the long run, since grade repeaters tend to obtain lower wages on the labor market, conditional on their highest credential¹. Grade retention may also entail some benefits. The mere presence of grade repetitions acts as an incentive device and may increase study effort.² Finally, the distribution of skills in a given cohort of outgoing students may be improved if grade repeaters benefit from a longer period of schooling. Yet, many important aspects of a cost-benefit analysis are imperfectly known. As a consequence, in spite of its widespread use, it is hard to tell if grade retention dominates social promotion, or which of the two systems has the highest value as a social policy. As is well known, the question is hotly debated and international comparisons show trends in both directions. For instance, in the recent years, France has relied less often on grade repetitions, while in the US, grade retention has made a certain comeback, as an ingredient of school accountability policies.

The consequences of grade retention are not easy to estimate. This is essentially due to the endogenous character of the decision to hold a student back and to unobservable heterogeneity. Many studies in the past may have found a negative impact of grade retention on various outcomes because grade repeaters are a selected population with abilities below the average. In the sequel, we propose a way of evaluating the treatment effects of grade repetition in French junior high schools (grades 6 till 9), using a rich set of micro-data, and taking the endogeneity of retention decisions and class size into account. We do not observe the students' wages and focus on educational outcomes.³

In a preliminary study of the data, we find that the local average treatment effect (*i.e.*, the LATE⁴) of grade retention on value-added, defined here as the difference between grade-9 and grade-6 scores, is significant and positive, using the quarter of birth as an instrument for retention. But the result doesn't seem to be very robust. We know that when treatment effects are heterogeneous, the linear Instrumental Variable (IV) estimator is a weighted average of marginal treatment effects (see the work of Heckman and Vytlacil (2005); see also Heckman (2010)). It follows that the IV estimates obtained with a particular instrument may not correctly

¹On this question, see Brodaty et al. (2012)

²On study effort, see De Fraja et al. (2010).

³For a study of the impact of grade retention on wages, using French data, see Brodaty et al. (2012)

⁴On this concept, see Imbens and Angrist (1994).

identify the relevant effects. Indeed, in the following, we show that the treatment effect of grade repetition varies with unobserved characteristics of students, being positive for some individuals and negative for others.

Taking our inspiration from the work of Heckman and his co-authors, we propose a tractable model in which treatment effects are heterogeneous (see, [Carneiro et al. \(2003\)](#)). We assume the existence of a finite number of *latent student types* and that the effects of retention may vary from one type of individual to the next. Our approach is parametric: the observed outcomes and the latent variables, such as unobserved test scores, are modeled as finite mixtures of normal distributions. The model can then be used to compute counterfactuals and treatment effects.

We take dynamics into account, exploiting the data's panel structure. Our approach is similar in spirit to that of [Cunha and Heckman \(2007, 2008\)](#) and [Cunha et al. \(2010\)](#), but different (and somewhat simpler) in a number of technical details. The educational outcomes of the same individuals are observed recursively through time, either completely (quantitative test scores) or partially (qualitative promotion decisions). The successive observations are used to identify the model parameters and the latent student types. In particular, the coefficients of student types, that is, their impact on the different outcomes, are identified under a limited set of reasonable assumptions.

To be more precise, we specify a structural model of knowledge-capital accumulation in junior high school. The model explains grade retention, class size, promotion decisions and test scores. It is estimated using panel data, on scores in grades 6 and 9, information on class sizes and on student transitions (promotion to next grade, retention and redirection towards vocational education). The panel provides a rich set of control variables describing family background and the environment of students. Repeated grades contribute to the accumulation (or destruction) of human capital (or skills) in a specific and type-dependent way. We present estimation results for a variant of our model with four unobserved student types or *groups*. Groups are clearly distinct and a clear hierarchy appears in terms of student ability. Groups are ranked in the same way if we use test scores in Math, in French, at the beginning of grade 6 or at the end of grade 9. The ranking of groups explains a similar ranking in the students' probabilities of grade retention (or promotion to the next grade). In a parallel fashion, the weaker the group, the smaller the class-size, in every grade. This result shows the endogeneity of class-size, which is used as a remediation instrument. Finally, to assess the impact of grade repetition on test scores at the end of grade 9, we compute the *ATT* and the *ATE* of the grade-repetition treatment. To this end, with the help of the model, we compute the counterfactual class-size and test scores of grade repeaters (resp. non-repeaters) that would be observed if they had not repeated a grade (resp. if they had repeated a grade), averaging over students and all possible types of each student, using their posterior probabilities of belonging to a group. We find that the *ATE* is negative, while the *ATT* is positive, but small and barely significant. The *ATE* and *ATT* are also computed within each of the four groups separately. This confirms that treatment effects are heterogeneous: grade retention is detrimental to able students but has some positive effects on the weakest students' final test scores. It is also shown that grade repetition has a negative impact on the student's probabilities of access to grade 9. We conclude that grade retention should be replaced by some other form of remediation.

There is a substantial literature on grade retention, but many early contributions did not address endogeneity or selection problems in a convincing way (Holmes and Matthews, 1984; Holmes, 1989). Few contributions have managed to propose a causal econometric evaluation of grade retention. An early attempt, providing IV estimates on US High-School data is due to Eide and Showalter (2001). Also in the US, Jacob and Lefgren (2004, 2009) use regression discontinuity methods to evaluate grade repetitions in the Chicago Public-Sector Schools. Jacob and Lefgren (2004) find some positive short-term effects of grade retention on test scores for primary school children. Neal and Schanzenbach-Whitmore (2010) also propose an evaluation of the 1996 reforms that ended social promotion in Chicago Public Schools. Dong (2010) studies grade retention in Kindergarten and finds positive effects. The same data is used by Cooley et al. (2011) to estimate a multi-period structural model in which the treatment effect of retention depends on the year of application. They also find positive effects. Recently, Baert et al. (2013) used a structural dynamic choice model, estimated with Belgian data, and found that grade retention has a positive impact on the next evaluation, and persistent effects. On Latin American countries see, Gomes-Neto and Hanushek (1994). Manacorda (2012) applies a regression discontinuity approach to Uruguayan junior high-school data and finds negative effects on the dropout rate. In France, contributions on this topic (with a causal approach) are due to Mahjoub (2007); Alet (2010); Brodaty et al. (2012); Alet et al. (2012); d'Haultfoeuille (2010). Among these authors, d'Haultfoeuille (2010) applies a new non-parametric method for the estimation of treatment effects to French primary education data and also finds positive effects. Finally, Brodaty et al. (2012) find negative signaling effects of grade retention on wages. None of the quoted papers use the methods and the data employed in the present article.

In the following, Section 2 describes the data. Section 3 presents a preliminary analysis of grade retention using linear IV methods. Section 4 presents our multi-stage skill accumulation model. The estimation strategy is exposed in Section 5. Section 6 and 7 present the estimation results and the average treatment effects. Concluding remarks are in Section 8.

2 Data

The data set used in this study is the 1995 secondary education panel of the French Ministry of Education (DEPP⁵ Panel 1995), which follows 17,830 students in junior high-school (*i.e.*, *collège*) from grade 6 to grade 9 (grade 6 is the equivalent of the French *classe de sixième*) during the years 1995-2001. The principals of a sample of junior high-schools were asked to collect data on all pupils born on the 17th day of each month, with the exception of March, July, and October, and entering grade 6 in September 1995 — about 1/40th of the whole cohort. A recruitment survey was conducted at the beginning of the first school year (1995-96). Then, a number of follow-up questionnaires were filled by the principals in every subsequent year until 2001, and a questionnaire was filled by the families in 1998 (with a response rate of 80%). Each student's junior high-school history was recorded without interruption, even when the student moved to another school. For each pupil and each year, we know the attended grade (6 to 9), the size of the class, and the promotion decision made by the teachers at the end of the year. In fact there are

⁵Département de l'Evaluation, de la Prospective et de la Performance

Table 1: Individual Grade Histories

Grade History	Count	
1234	9403	71,58%
12334	732	
12234	910	
11234	684	
Subtotal	2326	17,71%
1233V	33	
1223V	114	
1123V	154	
123V	147	
122V	146	
112V	246	
11V	7	
12V	560	
Subtotal	1407	10,71%
Total	13136	

three possible decisions: promotion to next grade, grade retention or redirection to vocational education (*i.e.* “steering”). These transition decisions are made during the last staff meeting (*i.e.*, the *conseil de classe*), at the end of every school year, on the basis of test scores and other more or less objective assessments of the pupil’s ability and potential in the next grade. Test scores in Mathematics and French are available at the beginning of grade 6 and at the end of grade 9. Grade 9 test scores are missing for the individuals who dropped out of general education for apprenticeship or vocational training, and therefore never reached grade 9 in the general (non vocational) middle schools.

Table 2: Students Promoted, Retained or Redirected in Each Grade and Year

Year t	Grade	Initial Stock	Promoted P	Retained R	Redirected V
$t = 1$	Grade 6	13136	12045	1091	0
$t = 2$	Grade 6	1091	1084	0	7
	Grade 7	12045	10315	1170	560
$t = 3$	Grade 7	2254	1862	0	392
	Grade 8	10315	9403	765	147
$t = 4$	Grade 8	2627	2326	0	301
	Grade 9	9403			

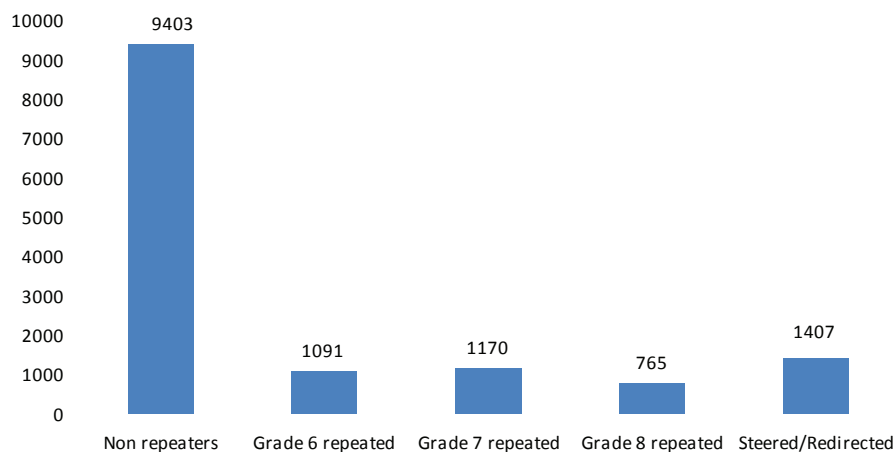
In addition, matching these data with another source from the Ministry of Education, the *Base Scolarité*, we obtain further information on school characteristics. In particular, total school enrollment and total grade enrollment (in each grade) for each year during the 1995-2001 period. These data will allow us to compute instruments for class-size, based on local variations of enrollment. There are some missing data, but the quality of the panel is very good. For example, initial test scores are known for 95% of the sampled individuals. Discarding observations with obvious coding errors and missing data, and slightly more than 450 histories of pupils registered

in special education programs (for mentally retarded children), we finally ended up with a sample of more than 13,000 individuals: 9,403 of them are in grade 9 in 1999, 2,594 are in grade 8 and 250 in grade 7. The last subset contains the few individuals who repeated a grade twice. We chose to discard these observations to reduce the number of cases. The final sample has 13,136 Students, which amounts to almost 75% of the individuals in the initial survey.

In the following, grades are denoted by g , and $g \in \{1, 2, 3, 4\}$, where $g = 1$ corresponds to grade 6, and so on. The year is denoted t with $t \in \{1, 2, 3, 4, 5\}$, where $t = 1$ corresponds to year 1995, etc. Individuals are indexed by i . Let g_{it} denote the grade of individual i in year t . With this notation system, a student i who doesn't repeat any grade is such that $g_{it} = t$. A grade repeater is such that $g_{it} = t - 1$. Table 1 gives the observed distribution of grade histories (in junior high school). Each row corresponds to a different type of trajectory. Letter V stands for vocational education. For example, the sequence 11234 means that grade 6 was repeated and therefore, that the student is observed in grade $g = 4$ in year $t = 5$. The sequence 123V indicates that the student was steered towards vocational education after grade 8. In total, about 30 % of the pupils do not complete junior high-school in four years: 18% are retained in one grade, 11 % are redirected.

Individual histories are described by Table 2 and on Fig. 1. Table 2 presents two rows per year, except in year $t = 1$. During the first year, all students are in grade 6. Out of the 13136 students initially enrolled in grade $g = 1$, 12045 are promoted, and 1091 are retained. In year $t = 2$, we see that 1084 repeaters in grade $g = 1$ are promoted and only 7 students have been redirected. In year $t = 3$ there are $2254 = 1170 + 1084$ students in grade 7; 1170 students repeating grade 7 and 1084 students that were in grade 6 the year before, etc. Figure 1 shows that the 9403 non-repeaters constitute a majority of more than 70% of the students. Repeaters amount to less than 9% of the latter cohort each year.

Figure 1: Number of Repeaters in each Grade



3 Preliminary Analysis : IV Estimates

We start our study of the causal effect of grade retention on educational achievement, using the student's quarter of birth as an instrument for grade retention, in a linear model. The quarter

or the month of birth has been used by various authors as an instrument ([Angrist and Krueger, 1991](#)). Recent work has shown that the month of birth can have long-lasting effects (see, [Bedard and Dhuey \(2006\)](#); [Grenet \(2009\)](#)). In his dissertation, and a recent paper, [Mahjoub \(2007, 2009\)](#), used the quarter of birth as an instrument for grade retention. This approach yields a positive impact of grade retention on value-added scores, defined as the difference between standardized grade-6 and grade-9 scores, in Mathematics and in French. We follow the same approach here, as a preliminary step.

Value added is higher for repeaters than for non-repeaters. This is true both in French and Mathematics. There exists a strong link between the age of a child, as measured by the month of birth, or quarter of birth, and the probability of grade repetition (for details, see the Appendix). The probability of grade retention is clearly higher for children born later in the year. In principle, children must be 6 years old on September 1st of year t to be admitted in primary school, grade 1, year t . First-quarter students tend to be relatively older in their class, with an age difference that can reach 11 months, and relatively older children tend to perform better. At the same time, teachers are reluctant to retain older children in a grade, as retention may change a difference —being older— into a stigma —being too old.

It follows that the month, quarter or season of birth is a candidate instrument for the grade-retention treatment, because it has good chances of being independent of the error term in an outcome equation with many controls. Note, in addition, as emphasized by [Mahjoub \(2007\)](#), that the value-added outcome being the difference of two test scores, possible specific and persistent effects of the birth quarter are “differenced out”.

We now estimate the effect of grade retention on value added by 2SLS, using the quarter of birth as an instrument for grade retention. Some descriptive statistics on value-added, as well as further details on this IV approach are relegated in the Appendix. Scores are standardized to have a mean of 50 and a standard deviation of 10 in grade 6 and in the whole sample (including all redirected pupils). Scores in grade 9 are standardized in the same way, using the sub-sample of individuals who reached grade 9. The first-stage is a linear regression of the grade-retention dummy on birth quarter dummies and controls (the linear probability model). Results are displayed in Table 3. The fourth quarter being the reference in the regressions, we see that relatively older students have a significantly lower probability of being held back.

Table 3: Grade Retention Probability

Variables	Grade retention
First quarter	-0.0513*** (0.0110)
Second quarter	-0.0459*** (0.00991)
Third quarter	-0.0133 (0.0109)
R^2	0.054
F statistic for instruments	31.74

Estimated by OLS. The dependent variable is the grade-retention dummy here. The following list of control variables were included in the regressions: gender; number of siblings; birth order (rank among siblings); parental occupation; parental education; indicator of grade repetition in primary school; total school enrollment. Standard errors are in parentheses; ***, ** and * indicate significance at the levels of 1, 5, and 10%, resp.

Table 4 presents OLS and 2SLS estimates of the effect of grade retention on value-added scores using the same set of controls. Instrumenting grade retention by the quarter of birth has dramatic effects on the sign and the size of the effect. Grade retention increases the score by about twice the standard deviation of value-added. These results confirm that the retention decision is endogenous.

Table 4: OLS and IV Estimates of Grade-Retention Effects

Dependent Variable	OLS		2SLS	
	Math VA	French VA	Math VA	French VA
Grade repetition	1.757*** (0.200)	1.899*** (0.196)	21.94*** (5.391)	14.79*** (4.510)
R^2	0.035	0.043		

The table reports the estimated coefficient of the retention dummy in different regressions. Gender is included as a control in all regressions in addition to number of siblings; birth order (rank among siblings); parental occupation; parental education; indicator of grade repetition in primary school; total school enrollment. Standard errors are in parentheses; ***, ** and * indicate significance at the levels of 1, 5, and 10%, resp.

Now, trying to estimate the impact of grade repetition in variants of this model, we found that the 2SLS results of Table 4 were not very robust, being very sensitive to the set of controls introduced in the equation of interest. But it is well known that IV estimates can be difficult to interpret when treatment effects vary with unobservable characteristics of individuals. To see this, let R denote the grade retention indicator. The outcome variable Y is value added. Let Y_1, Y_0 denote counterfactual outcomes for grade repeaters and non-repeaters. Let Z denote a dummy variable indicating whether the student was born in the first half of the year or in the second half. $Z = 1$ thus points at relatively older children. Let R_1, R_0 denote the counterfactual grade retention dummies, conditional on the instrument Z being 1 or 0. Under monotonicity,

i.e., $\mathbb{P}(R_1 \leq R_0) = 1$, the IV estimator converges to the Local Average Treatment Effect :

$$LATE = \mathbb{E}(Y_1 - Y_0 | R_1 - R_0 = -1)$$

This measures the average value-added score for the individuals whose retention in a grade would have been avoided, had they been born at the beginning of the year instead of at the end. The question is whether these marginal individuals are representative of the whole sample or not. To help answering this question, suppose that counterfactual scores follow a generalized Roy model⁶:

$$\begin{aligned} Y &= Y_1 = m_1 + U_1 & \text{if } Y_1 - Y_0 > c(Z) + V \\ &= Y_0 = m_0 + U_0 & \text{if } Y_1 - Y_0 \leq c(Z) + V, \end{aligned}$$

where $c(Z)$ is an increasing function of Z , interpreted as a cost. Assume that U_1, U_0 and V are independent given Z and that they are normally distributed. It is easy to show that

$$LATE = m_1 - m_0 + \frac{Var(U_1 - U_0)}{Var(U_1 - U_0) + Var(V)} \left(\frac{\phi(d_0) - \phi(d_1)}{\Phi(d_1) - \Phi(d_0)} \right)$$

with

$$d_z = c(z) - (m_1 - m_0), \quad z = 1, 0,$$

where ϕ is the normal density and Φ is the normal c.d.f. The cost of grade retention is higher for older individuals, so $c(1) > c(0)$, hence, $d_1 > d_0$. Let us assume, for the sake of the argument, that $c(1) > m_1 - m_0 > c(0)$, so $d_1 > 0 > d_0$. It is clear in this case that the $LATE$ may be positive or negative, without this telling us anything certain about the sign of $m_1 - m_0$. The $LATE$ being a marginal effect, it may predominantly reflect cost parameters and may not be informative about treatment effects. This is why, in the next section, we design a structural model to uncover the mechanisms of grade repetition and their impact on educational attainment.

⁶see [Heckman and Vytlačil \(2005\)](#)

4 A Model of Knowledge-Capital Accumulation

We construct a model of knowledge capital accumulation with unobserved heterogeneity. We found a source of inspiration in a series of influential papers by James Heckman and his coauthors, in which heterogeneity is captured by means of dynamic factor models (Cunha and Heckman, 2008; Cunha et al., 2010). Although close in spirit, the present approach relies on a somewhat simpler model. We use a multi-period setting. We rely on the idea that, in the educational process, inputs are imperfectly observed and outputs are imperfectly measured, by means of test scores and teacher's decisions. Unobserved heterogeneity is modeled by means of a discrete set of unobserved individual types, generating finite mixtures of normal distributions.

The model is designed to match the following data features. We observe test scores, in French and Mathematics, but only at the beginning of grade 6 and at the end of grade 9. Promotion decisions (promotion to the next grade, grade retention or redirection to vocational training) are observed in all years. In addition to these test scores and transitions, we also observe class size and total school enrollment. The students who do not drop off into vocational education at some point reach the terminal grade after four or five years, depending on retention, during the period 1995-2000. For children who never repeat a grade, we have observations in years $t = 1, 2, 3, 4$. For those who repeat a grade once and are not redirected to a vocational track, t can take all five values 1, 2, 3, 4, 5. Redirected children are the cause of attrition. Pupils are indexed by $i = 1, \dots, N$. Let $g_{it} \in \{1, 2, 3, 4\}$ denote the grade of student i in year t , and let $S_{it} \in \{P, R, V\}$ denote the promotion decision (*i.e.*, promotion, retention and redirection) at the last staff meeting of year t . $g_{i,t+1}$ is missing if $S_{it} = V$. All students start in grade 6 in year 1 ($g_{i1} = 1$), so we set $S_{i0} = P$ for all i . There is no redirection of children towards vocational education in grade 6, so $S_{i1} \in \{P, R\}$.

4.1 Initial conditions

Initial scores in Mathematics and French measure initial knowledge-capital in Mathematics and in French, denoted h_{m0} and h_{f0} respectively. We assume that individuals have four possible unobservable types, or equivalently, belong to one of four possible *groups*. Let G_{ik} denote the dummy which is equal to 1 if i belongs to group k and equal to 0 otherwise. Let p_k denote the unconditional probability of belonging to group k and, of course, $p_1 + p_2 + p_3 + p_4 = 1$. Knowledge-capital levels, at the beginning of grade 6, *i.e.*, h_{m0} and h_{f0} , have the following form:

$$h_{mi0} = c_{m01} + c_{m02}G_{i2} + c_{m03}G_{i3} + c_{m04}G_{i4}, \quad (1)$$

$$h_{fi0} = c_{f01} + c_{f02}G_{i2} + c_{f03}G_{i3} + c_{f04}G_{i4}. \quad (2)$$

In this formulation, Group 1 is the reference group. It follows that c_{m01} and c_{f01} are the average initial levels of knowledge-capital in Mathematics, and French, respectively, for Group 1 individuals. Subscript m , resp. f , indicates a coefficient related to the initial Mathematics capital, resp., the French language-capital equation. The average initial Mathematics-capital of Group k is thus $c_{m01} + c_{m0k}$, for $k = 2, 3, 4$, etc.

Human capital is therefore discrete, but this should not be taken literally. We could add a random term with a continuous distribution, representing other unobserved inputs to the

expressions of h_{mi0} and h_{fi0} , but the distribution of this term would not be identifiable, because it could not be distinguished from the teachers' "grading error", defined below. At this stage, we could also have added a list of controls, including indicators of family-background characteristics, but we omitted them, mainly to limit the number of parameters to be estimated. It follows that the groups may capture some of the effects of family background. Family-background variables and other controls will later be used to explain the probability of belonging to a given group, in separate regressions. We suppose that the test scores in French, denoted y_f , and in Math, denoted y_m , at the beginning of grade 6, are two different measures of the same knowledge-capital, that is,

$$y_{mi} = h_{mi0} + \varepsilon_{mi0} \quad (3)$$

$$y_{fi} = h_{fi0} + \varepsilon_{fi0}. \quad (4)$$

where ε_{m0} , ε_{f0} are random variables with a normal distribution and a zero mean, representing "grading" errors. The latter regression functions will identify the variance of ε_{m0} and ε_{f0} .

During the schooling of each student, we observe different variables that we regroup in different categories. There are time-invariant characteristics of the individual, such as family background observations, denoted X_0 ; time-varying characteristics of the individual denoted X_t , $t = 1, \dots, 5$ and time-varying characteristics of the school, used as instruments for class size, denoted Z_t . The variables used in regressions are listed in Table 5.

The instrument for class size exploits discontinuities induced by the application of a class-opening threshold, as in Angrist and Lavy (1999) and Hoxby (2000). Let N_{it} denote total grade enrollment in i 's school in year t . The *theoretical class size* in year t , denoted Z_{it} , is the class size that would obtain if the headmaster's rule was to open a new class, as soon as total grade enrollment in grade g_{it} became greater than τq and to minimize class-size differences, where τ is the class-opening threshold and q is an integer. Given these definitions, the theoretical *number* of classes in grade g_{it} , denoted κ_{it} , is by definition,

$$\kappa_{it} = \text{int}\left[\frac{N_{it} - 1}{\tau}\right] + 1,$$

where $\text{int}[x]$ is the largest integer q such that $q \leq x$. The theoretical number of students per class in grade g_{it} is simply

$$Z_{it} = \frac{N_{it}}{\kappa_{it}}.$$

Piketty and Valdenaire (2006) and Gary-Bobo and Mahjoub (2013) show how this function of total grade enrollment fits the observed data in the French Educational system. We set the threshold value $\tau = 25$ because it seems to provide the best fit with Panel 1995. We will see below that Z_{it} has a strong effect in class-size regressions.

Table 5: Sets of Variables

Time-invariant characteristics X_0	Time-varying characteristics X_1, X_2, X_3, X_4, X_5	Time-varying instruments Z_1, Z_2, Z_3, Z_4, Z_5
Gender.	Foreign language studied.	Theoretical class size
Father's occupation.	Special education zone.	(<i>i.e.</i> , Maimonides'
Mother's education.	Number of foreigners in school.	Rule)
Number of Siblings.	Class size.	
Grade retention in primary school.	Total school enrollment.	
Private sector in primary school.	Size of the urban area.	
	Private sector.	

4.2 Knowledge-capital accumulation

Knowledge, or human capital, accumulates according to the following equation:

$$h_{i1} = a_1 n_{i1} + b_1 X_{i1} + c_{11} + c_{12} G_{i2} + c_{13} G_{i3} + c_{14} G_{i4}, \quad (5)$$

where n_{i1} denotes class size in individual i 's class, grade $g_{i1} = 1$. Again, in equation (5), Group 1 is the reference, so that c_{11} is the impact of Group 1 on h_{i1} , and the impact of group k is $c_{11} + c_{1k}$ for all $k > 1$.

Many studies have established that class size is an endogenous variable. In particular, available evidence for France shows that class size is positively correlated with student performance because smaller classes are typically used to redistribute resources in favor of weaker students, or in favor of schools located in areas targeted for special help in education (see [Piketty and Valdenaire \(2006\)](#); [Gary-Bobo and Mahjoub \(2013\)](#)). We therefore model class size n_{i1} separately, as follows. Using Group 1 as the reference, we have,

$$n_{i1} = \alpha_{11} X_{i1} + \alpha_{12} Z_{i1} + \beta_{11} + \beta_{12} G_{i2} + \beta_{13} G_{i3} + \beta_{14} G_{i4} + \zeta_{i1} \quad (6)$$

The random term ζ_{i1} is an independent, normally distributed error.

Since we do not have any quantitative measure of performance at the end of grades $g \in \{1, 2, 3\}$, repeated or not, we define a single, latent education score for those years. In grade 6, *i.e.*, if $g_{it} = 1$, we define the latent variable,

$$y_{i1} = h_{i1} + \varepsilon_{i1} \quad (7)$$

where ε_1 is an independent normal error with a zero mean.

An individual is promoted to grade 7, *i.e.*, $g_{i,2} = 2$, if his(her) human capital is high enough, and repeats a grade otherwise. The promotion decision is modeled as a simple Probit. Let C_{11} be a human-capital threshold above which students are promoted. We have,

$$S_{i1} = \begin{cases} P & \text{if } y_{1i} \geq C_{11} \\ R & \text{if } y_{1i} < C_{11}. \end{cases} \quad (8)$$

The distribution of ε_1 is assumed to be standard normal, as usual in such a case, to identify the coefficients of the latent index. Given our specification of h_1 given by (5) above, we see that the model will only identify the constant

$$\delta_{11} = C_{11} - c_{11}.$$

This is of course technically equivalent to normalizing C_{11} , but, in principle, C_{11} is the human-capital level above which students pass, while c_{11} is the specific mean level reached by Group 1 students in the hypothetical situation $n_1 = X_1 = 0$. In essence, our model identifies differences between groups, not the absolute mean level of a group.

4.3 From second to fifth year

Similarly, still using Group 1 as the reference, in the second and third years, the human capital has the following representation.

If $g_{it} = t$ (non-repeaters),

$$h_{it} = a_t n_{it} + b_t X_{it} + c_{t1} + c_{t2} G_{i2} + c_{t3} G_{i3} + c_{t4} G_{i4}. \quad (9)$$

If $g_{it} < t$ (repeaters), we have

$$h_{it} = a_{tr} n_{it} + b_{tr} X_{it} + c_{t1r} + c_{t2r} G_{i2} + c_{t3r} G_{i3} + c_{t4r} G_{i4}. \quad (10)$$

The class-size equations are specified as follows.

If $g_{it} = t$ (non-repeaters), we have,

$$n_{it} = \alpha_{t1} X_{it} + \alpha_{t2} Z_{it} + \beta_{t1} + \beta_{t2} G_{i2} + \beta_{t3} G_{i3} + \beta_{t4} G_{i4} + \zeta_{it}, \quad (11)$$

where ζ_{it} is an independent normal random variable.

If $g_{it} < t$ (repeaters), we have,

$$n_{it} = \alpha_{t1r} X_{it} + \alpha_{t2r} Z_{it} + \beta_{t1r} + \beta_{t2r} G_{i2} + \beta_{t3r} G_{i3} + \beta_{t4r} G_{i4} + \zeta_{itr}. \quad (12)$$

where ζ_{itr} is an independent normal random variable.

At the end of the second and third years, if the student has not repeated a grade before, he or she can either pass to the next grade (P), repeat the year (R) or be redirected towards a vocational track (V). We model these three different transitions with an Ordered Probit. Promotion or retention decisions are made by the teachers' staff meetings (*i.e.*, the *conseils de classe*), at the end of every school year. In essence, these staff meetings base decisions on the student's grade-point average (hereafter GPA) at the end of the year, and decide whether to promote, to hold back, or to "steer" the student towards vocational education. Students with a GPA above a certain threshold are promoted; students with a low record are "steered"; students with a mediocre, below-the-average record repeat the grade, if the teachers' committee thinks that they can benefit from the repetition. It seems reasonable to assume that the promotion decision is based on some average of the teachers' assessments of the student's cognitive capital,

plus an unobserved individual effect, reflecting other unobservable factors that the members of the teaching staff take into consideration. We have in mind that the student's unobservable GPA in year t is highly correlated with the latent capital h_{it} , or to fix ideas, that h_{it} is the GPA in year t plus some random factor. We then model the unobservable capital h_{it} as an educational output, which is the result of some educational inputs: class-size, time-varying variables, and individual ability, as captured by the group indicator G_{ik} . Given this, and given the clear hierarchy of the three possible decisions, it seems reasonable to use an Ordered Probit structure.

Define first the latent variable

$$y_{it} = h_{it} + \varepsilon_{it},$$

where ε_t is an independent normal error. The decision S_{it} is then specified as follows,

$$S_{it} = \begin{cases} V & \text{if } y_{it} < C_t \\ R & \text{if } C_t \leq y_{it} < D_t \\ P & \text{if } y_{it} \geq D_t, \end{cases} \quad (13)$$

where C_t and D_t are the Probit cuts. We assume that ε_t has a standard normal distribution. As above, the model in fact identifies only the differences,

$$\delta_{t1} = C_t - c_{t1}, \quad \text{and} \quad \delta_{t2} = D_t - c_{t1}.$$

In the sample, a student never repeats a grade twice. Thus, the model embodies the fact that, if the student has already repeated a grade, he or she cannot repeat a second time. For repeaters, the possible decisions are: promotion to the next grade or redirection. We model the two different transitions with a simple Probit. We first define the latent variable,

$$y_{itr} = h_{it} + \varepsilon_{itr},$$

where ε_{tr} is an independent normal error. The decision S_{itr} is then specified as follows,

$$S_{itr} = \begin{cases} P & \text{if } y_{itr} \geq C_{tr} \\ V & \text{if } y_{itr} < C_{tr}, \end{cases} \quad (14)$$

where C_{tr} is a threshold, and we assume that ε_{tr} has a standard normal distribution. The model identifies only the difference, $\delta_{tr} = C_{tr} - c_{t1r}$.

It follows from these assumptions that the latent human capital h_{it} is affected by the promotion and retention decisions, because all the coefficients are free to vary in expressions (9) and (10), as well as in the auxiliary class-size equations (11)-(12), to describe a different productivity of inputs for students who repeated a grade.

The test scores in French, denoted y_{f4} , and in Math, denoted y_{m4} are two different measures of the final human capital. For non-repeaters, with obvious notations for the random error terms,

we have,

$$y_{mi4} = h_{mi4} + \varepsilon_{mi4}, \quad (15)$$

$$y_{fi4} = h_{fi4} + \varepsilon_{fi4}, \quad (16)$$

where ε_{m4} and ε_{f4} are independent normal random variables. For repeaters, at the end of grade 9, test scores in French are observed in year $t = 5$ and denoted y_{f5} . Similarly, test scores in Mathematics are denoted y_{m5} . We have two different measures of the repeaters' final human capital, with obvious notations for the independent random error terms,

$$y_{mi5} = h_{mi5} + \varepsilon_{mi5}, \quad (17)$$

$$y_{fi5} = h_{fi5} + \varepsilon_{fi5}. \quad (18)$$

The functions h_{mit} and h_{fit} , with $t = 4, 5$ have the same specification as h_{it} (as given by (9) above), with coefficients a_{mt} , b_{mt} , c_{mt} and a_{ft} , b_{ft} , c_{ft} , etc., that may be different for Mathematics and French. Our model is now fully specified.

5 Estimation Method

The estimation method is a variation on the EM algorithm. Let Y_i be the set of outcomes observed for individual i : $Y_i = (y_{mi0}, y_{fi0}, S_{i1}, \dots, S_{i4}, y_{mi4}, y_{fi4})$. Let $X = (X_1, X_2, X_3, X_4, X_5)$ and $Z = (Z_1, Z_2, Z_3, Z_4, Z_5)$. Then, we denote θ the vector of all model parameters, namely, $\theta = (p_1, p_2, p_3, p_4, a_i, b_i, c_{ij}, \alpha_i, \dots)$. We replicate each individual i in the sample to create 4 different artificial observations of i . Student i 's replicas differ by the unobserved type, or group k only, but the values of X_i , Y_i and Z_i are the same for each replica. We arbitrarily choose initial values for the unconditional prior probabilities of the groups p_k , $k = 1, \dots, 4$, and for the posterior probabilities of belonging to a certain group knowing the observed characteristics of i , that is, $p_{ik} = \mathbb{P}(G_{ik} = 1|Y, X, Z)$. They will be updated after each iteration.

The estimation algorithm can be described as follows.

1. We first run 20 *weighted* regressions and Ordered Probits.
 - (a) Two regressions for the initial test scores in Math and French.
 - (b) Two regressions of class size by grade: one for the repeaters and one for the non-repeaters (except for the first year, because there are only non-repeaters in year $t = 1$ and for year $t = 5$, because there are only repeaters). This amounts to 8 regressions.
 - (c) One simple Probit to model the transition at the end of grade 6 in year $t = 1$. Two Ordered Probits to model the decision at the end of grades 7 and 8 for non-repeaters. Three simple Probits to model steering decisions relative to repeaters in grades 6, 7 and 8. There are 4 Probits and 2 Ordered Probits in total.
 - (d) Two final test-score regressions in Math and French, for repeaters and non-repeaters (4 regressions).

2. We obtain an estimation of θ by means of our system of weighted regressions and weighted Probits.
3. The residuals of regressions and the probabilities of passing to the next grade are collected to compute the individual contributions to likelihood, that is, by definition,

$$l_i(X, Z, Y, \theta) = \sum_{k=1}^K p_k l_i(Y, X, Z, \theta \mid G_{ik} = 1). \quad (19)$$

4. Individual posterior probabilities p_{ik} of belonging to a group are then updated, using Bayes' rule and the likelihood as follows,

$$p_{ik} = \mathbb{P}(G_{ik} = 1 \mid Y, X, Z, \theta) = \frac{p_k l_i(Y, X, Z, \theta \mid G_{ik} = 1)}{\sum_{j=1}^K p_j l_i(Y, X, Z, \theta \mid G_{jk} = 1)}. \quad (20)$$

These individual probabilities are then averaged to update the prior probabilities p_k , as follows,

$$p_k = \mathbb{P}(G_k = 1) = \frac{1}{N} \sum_{i=1}^N p_{ik}. \quad (21)$$

5. A new iteration begins until convergence of the estimated unconditional probabilities.

All standard deviations have been bootstrapped, using 50 drawings with replacement in the sample.

The estimation method used here has been advocated and justified by various authors ([Arcidiacono and Jones, 2003](#); [Bonhomme and Robin, 2009](#)).

6 Estimation Results

6.1 Distribution of groups

Table 6: Estimated Group Probabilities

	Group 1	Group 2	Group 3	Group 4
Probabilities	15.54 %	31.16 %	33.56 %	19.74 %
	(0.69)	(0.64)	(0.58)	(0.82)

The results of the algorithm, using $K = 4$ groups, are given by Table 6. We chose to use only 4 groups because of weak identifiability and computational problems when $K > 4$. In Table 7, we compare the most likely groups of individuals, estimated with the full model, called Classification 1, with the results of a limited sub-model, based on grade 6 entry scores only, called Classification 2. Both models have 4 unobserved types or groups. This has been done to try to assess the impact of initial test scores on the individual's posterior probabilities of belonging to a group. In other words, are students fully predetermined by their initial stock of knowledge? We observe that, according to Classification 2, 75% of Group 1 individuals are also most likely to become

Table 7: Comparison of Two Classifications

Classification 1	Classification 2				Total
	Group 1	Group 2	Group 3	Group 4	
Group 1	74 %	1 %	0 %	3 %	2021
Group 2	24 %	59 %	2 %	61 %	4076
Group 3	0 %	38 %	48 %	34 %	4383
Group 4	2 %	2 %	50 %	2 %	2656
	100 %	100 %	100 %	100 %	
Total	2547	2883	4967	2739	13136

members of Group 1, according to Classification 1 (the full model). Observing the grade 6 scores in Math and French only allows us to assign the student to the first group, to a large extent. But Group 4 students are not predetermined by their entry test scores, since less than 2% of the students assigned to Group 4 on the basis of the latter scores end up being members of Group 4 in the full model. The corresponding percentages are 59% and 48% for Groups 2 and 3, respectively. We conclude that, with the exception of Group 1, unobserved types are far from being perfectly predicted in year $t = 1$ (*i.e.*, in grade 6). It seems that the weakest students are easily detected from the beginning, but the brightest students are not. We will come back to this point in the general discussion of estimation results below.

Table 8 presents the parameters obtained when we regress the individual posterior probabilities of belonging to a certain group k , defined as p_{ik} above, on the socio-demographic and family-background variables X_0 . We find that the probabilities of belonging to the two extreme groups, Group 1 and Group 4, are quite well predicted by the social background, with an R^2 superior to 14%. These results also show, among other things, that when the mother is educated and the father is an executive, the probability of belonging to Group 4 is significantly increased. Group 2 and Group 3 are not so easy to distinguish on the basis of observed student characteristics.

Table 8: Individual Group Probabilities and Family Background

	Group 1	Group 2	Group 3	Group 4
Female	0.0493*** (0.00565)	0.0388*** (0.00766)	-0.00773 (0.00779)	-0.0804*** (0.00638)
Mother education: Junior High School	-0.0126 (0.00873)	-0.0167 (0.0118)	0.0234** (0.0120)	0.00594 (0.00985)
Mother education: Vocational Certificate	-0.0521*** (0.00937)	-0.0175 (0.0127)	0.0531*** (0.0129)	0.0165 (0.0106)
Mother education: High-School Graduate	-0.0901*** (0.0109)	-0.103*** (0.0147)	0.0550*** (0.0150)	0.138*** (0.0123)
Mother education: 2 years of college	-0.0864*** (0.0118)	-0.154*** (0.0160)	0.0832*** (0.0162)	0.157*** (0.0133)
Mother education: 4 years of college and more	-0.103*** (0.0142)	-0.174*** (0.0192)	0.0240 (0.0195)	0.253*** (0.0160)
Father occupation: Executives and educated professionals	-0.0514*** (0.0181)	-0.0373 (0.0245)	0.0100 (0.0250)	0.0786*** (0.0204)
Father occupation: White collars	-0.00141 (0.0184)	0.0674*** (0.0249)	-0.0314 (0.0253)	-0.0346* (0.0207)
Father occupation: Blue collars	0.0557*** (0.0169)	0.0777*** (0.0229)	-0.0696*** (0.0233)	-0.0638*** (0.0191)
More than three children in family	0.0845*** (0.00840)	0.0004 (0.0114)	-0.0432*** (0.0116)	-0.0418*** (0.00949)
Retention in primary school	0.206*** (0.00731)	0.0412*** (0.00990)	-0.169*** (0.0101)	-0.0781*** (0.00825)
Quarter of birth				
Q2	0.0000 (0.0077)	0.0144 (0.0105)	-0.0193* (0.0107)	0.0049 (0.0087)
Q3	0.0198** (0.0085)	0.0256** (0.0115)	-0.0331*** (0.0117)	-0.0123 (0.0096)
Q4	0.0145* (0.0087)	0.0463*** (0.0117)	-0.0260** (0.0119)	-0.0349*** (0.0098)
R^2	0.187	0.059	0.060	0.143

Linear regressions of probabilities p_{ik} on controls X_0 . Standard errors are in parentheses; ***, ** and * indicate significance at the levels of 1, 5, and 10%, resp. There are 12,937 observations.

6.2 Group effects on test scores

We present here the estimated parameters of group effects and class size. Table 9 shows the estimated coefficients for the initial test scores (at the beginning of grade 6) and the final test scores (at the end of grade 9). Group 1 is the reference. We see how well the four groups are defined. Scores in French and Math increase with group k and the estimated coefficients yield the same ranking of ability groups in all columns, except the rightmost column of Table 9. More precisely, Group 4 has everywhere the highest scores, with the exception of Group 4 repeaters, in French, but the latter coefficient is estimated with less precision than the others. Intuitively, this is because Group 4 students have a low probability of repeating a grade. Apart from this exception, Group 4 is above Group 3, which in turn dominates Group 2, and Group 1 is unambiguously the lowest ability group.

Table 9: Estimated Impact of Groups and Class Size on Test Scores

	Score in Math			Score in French		
	Initial	Final		Initial	Final	
		Non-repeaters	Repeaters		Non-repeaters	Repeaters
Class size $t = 4$		-0.25*** (0.03)			-0.25*** (0.04)	
Class size $t = 5$			-0.19*** (0.07)			-0.25*** (0.05)
Group 2	10.44*** (0.27)	8.14*** (0.57)	5.32*** (0.67)	10.82*** (0.23)	9.10*** (0.65)	5.80*** (0.69)
Group 3	19.17*** (0.22)	15.80*** (0.62)	9.07*** (0.91)	19.16*** (0.30)	16.65*** (0.61)	10.13*** (0.73)
Group 4	25.42*** (0.25)	26.18*** (0.62)	16.05*** (5.24)	25.60*** (0.28)	27.50*** (0.68)	9.22** (5.18)
Constant	35.34*** (0.24)	41.88*** (0.92)	43.31*** (1.71)	35.20*** (0.26)	40.87*** (0.94)	44.29*** (1.20)
R^2	0.68	0.60	0.18	0.68	0.63	0.21

Standard errors are in parentheses; ***, ** and * indicate significance at the levels of 1, 5, and 10%, resp.

If we now focus on final scores, it is easy to see that Group 1 gets higher scores on average when a grade was repeated (*i.e.*, this is because the constant is higher). In contrast with Group 1, individuals in Groups 3 and 4 who didn't repeat a grade obtain higher scores than the repeaters of these two groups. Take Group 3 for instance. To obtain the final score in Math of the average Group 3 student who repeated a grade, we add the constant in the column, *i.e.*, 43.31 to the differential impact of Group 3, *i.e.*, 9.07. The total is 52.38. But if we compute the corresponding term for Group 3 non-repeaters, in Math, we obtain, $15.80 + 41.88 = 57.68$. Grade repetition seems detrimental to Group 3. The same is true with Group 4. For the latter group, the corresponding additions yield 68.06 in the non-repeaters' column and 59.36 in the repeaters' column. However, individuals in Group 2 get approximately the same increase in their score, whether they repeat or not.

6.3 Promotion decision model and effects of class size

If we now look at the top rows in Table 9, we find that increasing class-size has a negative impact in grade 9 for all students. The standard deviation of class size is around 3.⁷ It follows that the estimated impact of a standard deviation of class size is around three quarters of a normalized test-score point for non-repeaters, or 7.5% of the standard deviation of test scores. The significant negative coefficient on class-size appears because we control for unobserved heterogeneity, and therefore, for the endogeneity of this variable. Otherwise, the coefficient on class size would be positive (we return to this question below, when we discuss the class-size regressions). This being said, we do not find a very strong class-size effect on final scores (a quarter of a point, or 1/40th

⁷To be precise, the standard deviation of class size in year t , denoted σ_{nt} has the following values $\sigma_{n1} = 3.02$, $\sigma_{n2} = 2.90$, $\sigma_{n3} = 3.32$, $\sigma_{n4} = 3.38$.

of the standard deviation of test scores, for a one-student reduction in class-size).

Table 10: Estimated Impact of Groups and Class Size on Promotion Decisions

Dependent variable ↓	Class size	Group 2	Group 3	Group 4	Cut 1 δ_{t1}	Cut 2 δ_{t2}	Cut R δ_{tr}
S_1	−0.025*** (0.007)	0.67*** (0.04)	2.24*** (0.12)	2.45*** (0.72)	−1.13*** (0.16)		
S_2 repeaters	0.010** (0.04)	4.29*** (0.42)	4.22*** (0.61)	3.17*** (1.3)			−1.80* (1.12)
S_2	−0.004 (0.006)	0.63*** (0.044)	1.62*** (0.057)	2.72*** (0.64)	−0.85*** (0.14)	−0.08 (0.14)	
S_3 repeaters	−0.016* (0.012)	0.38*** (0.017)	0.92*** (0.24)	4.43*** (1.37)			−0.93*** (0.29)
S_3	0.045*** (0.006)	0.33*** (0.05)	0.92*** (0.06)	1.67*** (0.17)	−0.64*** (0.18)	0.34** (0.16)	
S_4 repeaters	0.035*** (0.01)	0.33*** (0.07)	0.65*** (0.12)	0.55 (1.91)			−0.002 (0.24)

The promotion decisions S_t are modeled with the help of an Ordered Probit. They take the value 0 for redirection, 1 for retention and 2 for pass. Standard errors are in parentheses; ***, ** and * indicate significance at the levels of 1, 5, and 10%, resp.

Table 10 shows the main parameters of the promotion decision model. Dependent variables determine rows, while the coefficients of a given explanatory variable in equations are displayed in the same column. A higher group label means a higher average knowledge-capital. As a consequence, the greater the group label, the greater the probability of passing to the next grade, for non-repeaters as well as for repeaters, in each grade. The estimated coefficients reflect this ranking of groups very clearly, again, with the exception of the impact of Group 4 in the Probit concerning grade 8 repeaters (*i.e.*, S_4 repeaters). The latter coefficient is not estimated with precision because Group 4 students have a small probability of repeating a grade. Apart from this exception, all other coefficients are estimated with good precision. The first column of Table 10 shows that increasing class size decreases the probability of promotion to grade 7, but has a non-significant (or even a positive impact) on pass rates in later grades.

Table 11: Estimates of Class-Size Equation Parameters

Dependent variable ↓	Maimonides' Rule	Constant	Group 2	Group 3	Group 4	R^2
Class size $t = 1$	0.32*** (0.02)	16.09*** (0.36)	1.12*** (0.18)	1.75*** (0.15)	1.78*** (0.16)	0.20
Class size $t = 2$ (repeaters)	0.49*** (0.05)	14.81*** (1.09)	0.68*** (0.27)	-5.75*** (1.51)	3.15* (2.35)	0.25
Class size $t = 2$	0.37*** (0.02)	15.17*** (0.17)	1.07*** (0.30)	1.85*** (0.14)	1.96*** (0.16)	0.21
Class size $t = 3$ (repeaters)	0.36*** (0.05)	16.06*** (0.91)	0.53*** (0.17)	1.13*** (0.32)	-1.96* (1.27)	0.18
Class size $t = 3$	0.35*** (0.05)	13.66*** (0.40)	1.87*** (0.22)	2.90*** (0.20)	3.10*** (0.26)	0.24
Class size $t = 4$ (repeaters)	0.33*** (0.05)	15.61*** (0.90)	0.95*** (0.22)	2.00*** (0.29)	1.85 (2.16)	0.19
Class size $t = 4$	0.35*** (0.02)	14.05*** (0.46)	1.62*** (0.34)	2.62*** (0.26)	2.94*** (0.29)	0.26
Class size $t = 5$ (repeaters)	0.26*** (0.04)	16.34*** (0.73)	0.91*** (0.32)	2.67*** (0.31)	0.32 (2.92)	0.22

Standard errors are in parentheses; ***, ** and * indicate significance at the levels of 1, 5, and 10%, resp.

6.4 Endogeneity of class size

Table 11 finally gives the coefficients of group dummies and of instruments in class-size equations. Each row in the table corresponds to a dependent variable. One of the class-size instruments is theoretical class size (*i.e.*, Maimonides' rule), that is, the class size that would be experienced by the student if a class-opening threshold of 25 was applied, given total grade enrollment. The coefficient of this variable is significant and positive, as expected. We also find that class size increases with the ability (*i.e.*, the group) of students. The only exceptions are the coefficients on Group 4 dummies, that cannot be estimated with precision among grade repeaters. These results prove that class-size is strongly endogenous, and that it is used as a remediation instrument by school principals.

Our estimates are robust if the group dummies are exogenous variables in each year. To check this, we regressed the posterior probabilities of belonging to a group over a set of permanent individual characteristics X_0 and the time-varying characteristics X_1, X_2, X_3, X_4, X_5 . The results of these latter regressions are not presented here, but they show that, if the coefficients on X_0 are strongly significant, in contrast, time-varying characteristics are not significant. Thus, our model seems well specified (and we found a confirmation of well-known results). A better social background (that is, richer, more educated and more qualified parents) significantly increases the initial capital and therefore, the probability of belonging to high-ability groups.

7 The Treatment Effects of Grade Retention

We now turn to the key question of the present paper: the treatment effects of grade repetition. The model will be used to compute counterfactuals.

7.1 Effect of grade retention on grade 9 scores

Each individual i has a posterior conditional probability p_{ik} of belonging to each of the four groups $k = 1, \dots, 4$. For each individual and each of his (her) possible types, we compute a counterfactual class size and a counterfactual final test score. Each individual has four counterfactual final scores and four counterfactual final class sizes. Using the posterior probabilities, we can then compute expected counterfactual grades.

For each group, and for each student who hasn't repeated a grade,

1. we compute the class size he or she would have experienced in grade 9, if he or she had repeated a grade.

To do this, we assume that the student doesn't move to a different school and that his class environment has the same characteristics (same number of foreigners, same foreign language chosen, same size of the urban area, same sector (private or public), same classification as priority education zone). However, we use the information that we have on total school enrollment and total grade enrollment in the same school one year later.

2. we compute the grade predicted in grade 9 if the student had repeated a grade (this counterfactual is denoted Y_r^c).

For each grade repeater and each group,

1. we compute the class size predicted in grade 9 if the student had not repeated a grade;
2. we compute the student's predicted grade in grade 9 if he or she had not repeated a grade (this counterfactual is denoted Y^c).

Let N_r denote the number of individuals who repeated a grade and let N_p denote the number of individuals who didn't repeat a grade. Of course, we have, $N = N_p + N_r$. Let y_{ri} be the observed final grade of i , if i is a repeater. Let y_i be the observed final grade of i , if i never repeated a grade. We can now compute the following treatment effects.

The average treatment effect (*i.e.*, ATE) is defined as follows.

$$ATE = \frac{1}{N} \left(\sum_{i \in \mathbf{N}_p} \sum_{k=1}^4 (\mathbb{E}(Y_{ri}^c | G_{ik} = 1) - y_i) p_{ik} + \sum_{i \in \mathbf{N}_r} \sum_{k=1}^4 (y_{ri} - \mathbb{E}(Y_i^c | G_{ik} = 1)) p_{ik} \right), \quad (22)$$

where $p_{ik} = \mathbb{P}(G_{ki} = 1 | X, Z, Y)$ is i 's posterior probability of belonging to Group k . In the above expression, $\mathbb{E}(Y_{ri}^c | G_{ik} = 1)$ and $\mathbb{E}(Y_i^c | G_{ik} = 1)$ are the predictions of i 's final grades, in the counterfactual situations of grade repetition and not repeating, respectively, using the estimated regression functions, and conditional on belonging to Group k .

The average treatment effect on the treated (*i.e.*, ATT) is then,

$$ATT = \frac{1}{N_r} \sum_{k=1}^4 \sum_{i \in \mathbf{N}_r} (y_{ri} - \mathbb{E}(Y_i^c | G_{ik} = 1)) p_{ik}. \quad (23)$$

We also compute an ATE by group. For Group k , the average treatment effect ATE_k is defined

as,

$$ATE_k = \frac{1}{Np_k} \left(\sum_{i \in \mathbf{N}_p} (\mathbb{E}(Y_{ri}^c \mid G_{ik} = 1) - y_i) p_{ik} + \sum_{i \in \mathbf{N}_r} (y_{ri} - \mathbb{E}(Y_i^c \mid G_{ik} = 1)) p_{ik} \right), \quad (24)$$

where $p_k = (1/N) \sum_i p_{ik}$. The, ATT within group k , denoted ATT_k , can be defined in a similar way,

$$ATT_k = \frac{1}{\sum_{i \in \mathbf{N}_r} p_{ik}} \sum_{i \in \mathbf{N}_r} (y_{ri} - \mathbb{E}(Y_i^c \mid G_{ik} = 1)) p_{ik}. \quad (25)$$

Table 12: Counterfactuals required to compute the probabilities of accessing grade 9

History	Grade 7		Grade 6R		Grade 8		Grade 7R		Grade 8R	
	Pr(S_2)	n_2	Pr(S_{2r})	n_{2r}	Pr(S_3)	n_3	Pr(S_{3r})	n_{3r}	Pr(S_{4r})	n_{4r}
1234			C	C			C	C	C	C
12334			C	C			C	C		
12234			C	C	C	C				
11234	C	C			C	C				
1233V			C	C			C	C		
1223V			C	C	C	C				
1123V	C	C			C	C				
123V			C	C			C	C	C	C
122V			C	C	C	C			C	C
112V	C	C			C	C			C	C
12V			C	C	C	C	C	C	C	C
11V	C	C			C	C	C	C	C	C

Letter C indicates that a counterfactual value has been computed. Letter R indicates that a grade-repeater model is used. $\text{Pr}(S_t)$ means the probability distribution of decision $S_t \in \{P, V, R\}$. n_t denotes class-size in year t . Subscript r indicates the specific model for grade repeaters, $S_{tr} \in \{P, V\}$.

7.2 Effect of grade retention on the probability of access to grade 9

Individual i 's estimated probability of access to grade 9, knowing Group k , is denoted P_{9ik} and can be decomposed in the following way:

$$\begin{aligned} P_{9ik} = & \text{Pr}(S_{i1} = P \mid k) \text{Pr}(S_{i2} = P \mid k) \text{Pr}(S_{i3} = P \mid k) \text{ (does not repeat)} \\ & + \text{Pr}(S_{i1} = P \mid k) \text{Pr}(S_{i2} = P \mid k) \text{Pr}(S_{i3} = R \mid k) \text{Pr}(S_{i4r} = P \mid k) \text{ (repeats grade 8)} \\ & + \text{Pr}(S_{i1} = P \mid k) \text{Pr}(S_{i2} = R \mid k) \text{Pr}(S_{i3r} = P \mid k) \text{Pr}(S_{i4r} = P \mid k) \text{ (repeats grade 7)} \\ & + \text{Pr}(S_{i1} = R \mid k) \text{Pr}(S_{i2r} = P \mid k) \text{Pr}(S_{i3r} = P \mid k) \text{Pr}(S_{i4r} = P \mid k), \text{ (repeats grade 6)} \end{aligned}$$

where, to simplify notation, we denote $\text{Pr}(S_{it} = X \mid k) = \text{Pr}(S_{it} = X \mid G_{ik} = 1)$, for all $X = P, R, V$. If the government decides to abolish grade retention (but keeps the possibility of steering students towards the vocational track) then, the only way of reaching grade 9 is to pass the three grades directly. Let P_{9ik}^c be the counterfactual probability of accessing grade 9 when grade retention is abolished. Given that no student is redirected to the vocational track at the

end of grade 6, this probability can be expressed as follows,

$$P_{9ik}^c = \Pr(S_{i2} = P \mid k) \Pr(S_{i3} = P \mid k).$$

To find the average treatment effect of grade retention, we need to compute the individual probabilities P_{9ik} and P_{9ik}^c for all the students in the sample, including those who have actually been redirected. This requires the computation of many counterfactuals.

Table 13: Average Treatment Effects of Grade Retention

	Mathematics		French		Probability of access to grade 9	
	ATE	ATT	ATE	ATT	ATE	ATT
Group 1	2.43 (0.76)	2.45 (0.76)	3.09 (0.81)	3.20 (0.80)	-0.11 (0.014)	-0.11 (0.014)
Group 2	0.12 (0.42)	0.36 (0.42)	0.18 (0.41)	0.47 (0.42)	-0.12 (0.012)	-0.12 (0.012)
Group 3	-3.79 (0.76)	-2.92 (0.75)	-3.66 (0.60)	-2.77 (0.59)	-0.09 (0.022)	-0.10 (0.023)
Group 4	-6.68 (4.61)	-14.08 (4.53)	-6.86 (4.54)	-14.22 (4.52)	-0.06 (0.07)	-0.06 (0.07)
All	-2.56 (0.85)	0.27 (0.31)	-3.73 (0.94)	0.71 (0.33)	-0.09 (0.017)	-0.11 (0.008)

Standard deviations are in parentheses.

For those who repeated grade 6 and then passed or were redirected, we need counterfactual class sizes and counterfactual school-environment characteristics for year 2 and 3, that they would have experienced, had they not repeated a grade. For those who repeated grade 5 or have been redirected at the end of grade 5, we need their counterfactual class size and counterfactual characteristics for year 3, as if they hadn't repeated this grade. Finally, for those who were never held back, we need the counterfactual class size and characteristics that they would have experienced, had they repeated a grade. Table 12 summarizes the counterfactual probabilities and the counterfactual class size we computed for each different grade history. Then we can compute the following treatment effects. The average treatment effect is,

$$ATE = \frac{1}{N} \sum_{k=1}^4 \left(\sum_{i \in \mathbf{N}_r} (P_{9ik} - P_{9ik}^c) p_{ik} + \sum_{i \in \mathbf{N}_p} (P_{9ik} - P_{9ik}^c) p_{ik} \right).$$

The average treatment effect on the treated is then,

$$ATT = \frac{1}{N_r} \sum_{k=1}^4 \sum_{i \in \mathbf{N}_r} (P_{9ik} - P_{9ik}^c) p_{ik}.$$

7.3 Results and discussion

Table 13 displays the results of the various computations. The last row in this table shows the overall results. If we consider the final tests scores in Math and French (at the end of grade 9), the ATT is positive, but small. Given that the mean value of the scores is 50 with a standard deviation of 10, the effects are smaller than a tenth of a standard deviation and barely significant. The ATE is clearly negative in Math and in French. As we will see, this is mainly due to the fact that the most able students would suffer from grade repetitions. If we now look at the values of ATE_k and ATT_k , the treatment effects within group k , it is easy to see that only Group 1 students benefit for grade repetitions. The effect of grade repetitions is not significantly different from zero for Group 2 students. In contrast, in the case of Group 3, and Group 4, both the ATE and the ATT are negative, in Math and in French. This shows that grade repetition hurts the students belonging to top groups.⁸ We conclude that grade repetitions have some usefulness for the weakest students, with an effect of the order of a quarter of a standard deviation on the final grades.

⁸Note that ATT_k and ATE_k should be equal for each k , if Group k was the only variable used to predict counterfactual scores. But other control variables are used to predict these scores, such as class size, family background characteristics, etc. This determines differences between ATT_k and ATE_k in Table 13. However, the differences are neither large nor significant.

Figure 2: Histogram of Individual Probabilities of Access to Grade 9

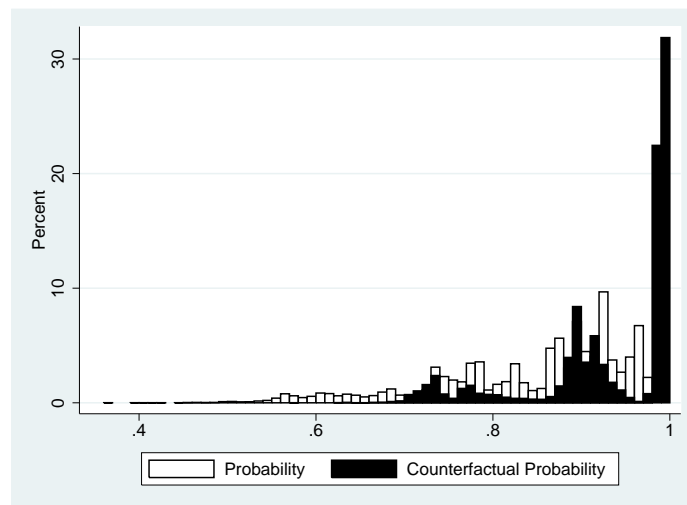


Figure 3: Histograms of Probabilities of Access to Grade 9, by Group

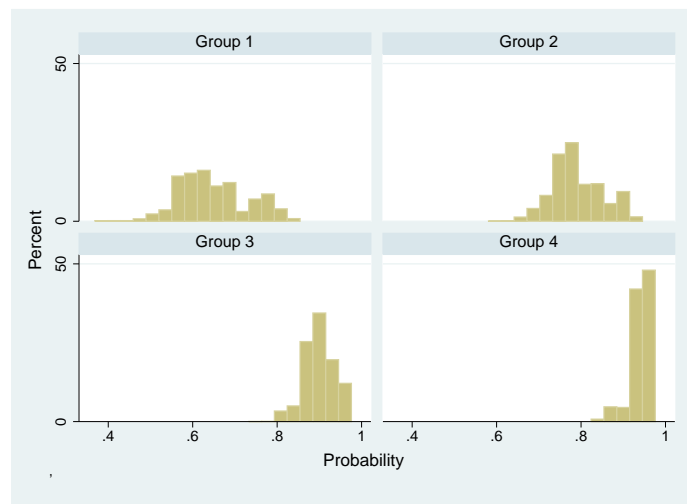
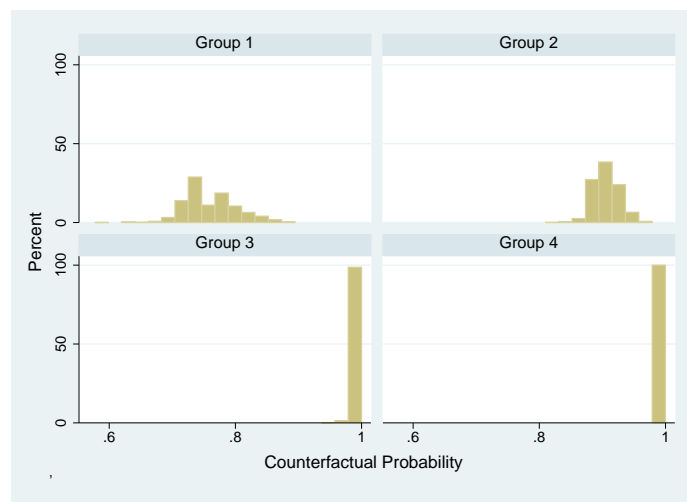


Figure 4: Histograms of Counterfactual Access Probabilities, by Group



We now discuss the effect on the probability of access to grade 9. The treatment effects of grade repetition on final scores rely essentially on the regression equations determining the final

test scores, and the latter equations are estimated with the subset of individuals who reached grade 9. The fact that this population is selected is taken into account by the posterior individual probabilities p_{ik} . But it is reassuring to derive results for an outcome that depends on the entire structure of the model. This is the case of access to grade 9, because the probabilities P_{9ik} , defined above, depend on all the decision and class-size equations.

It is striking to see that in Table 13, the *ATTs* and *ATEs* of grade retention are all negative, even if we consider within-group treatment effects. This means that introducing grade retention, if grade retention does not already exist, will be detrimental to students, on average, and detrimental to students of each group, taken separately. The effects are particularly strong for Groups 1 and 2. To see this, we computed the distribution of the individual probabilities P_{9ik} and individual counterfactual probabilities P_{9ik}^c in the student population. The histograms of these distributions are displayed on Figure 2.

On Fig. 2 it is easy to see that the counterfactual probabilities have a mass near 1, meaning that the abolition of grade repetitions would help many students to reach grade 9. Yet, there are clearly subgroups of individuals that keep a low probability of access: these individuals bear a high risk of being tracked in vocational programs. We will understand the effect of grade repetition on access to grade 9 more fully if we compute the histograms of P_{9ik} and P_{9ik}^c separately for each group. This is done in the following figures. Figure 3 gives the distributions of P_{9ik} , while Figure 4 displays the distributions of the counterfactual P_{9ik}^c .

Comparing the histograms, it immediately comes to mind that when grade repetitions are abolished, access to grade 9 becomes certain for Group 3 and Group 4 students. The effect of abolition is less obvious for the weakest groups, 1 and 2, but in fact, these probabilities increase and become more favorable. To sum up, these effects explain why the treatment effects of grade repetitions on access to grade 9 are unambiguously negative. We see also that these effects are very strong, since a drop of 11 or 12 points of probability, very roughly, amounts to 50% of the best chances of access to grade 9 among Group 1 and Group 2 students.

The treatment effects are positive only for the weakest students, and these effects are weak when they are positive. Given these results, and the results of Table 13 in general, it seems that we can only recommend the abolition of grade retention. The results of Table 7 suggests a path for reform. Coming back to this table, we see that the weakest (*i.e.*, the Group 1) students are more easily detected in grade 6 than other types. In cases of grade retention, forcing weaker students to follow the same teaching twice is only a rough second best. It would be more efficient to track these students from the start of junior high school, with additional remediation resources. One could imagine a slow track and a fast track, with, say, a year of difference in duration to reach the certification exams at the end of grade 9, and with flexible possibilities of track changes in both directions. To avoid the stigma of tracking, the slow track should probably be the norm, and students that seem promising would be steered towards the fast track. A system of that sort would lead to a more efficient use of resources than grade repetitions. It would clearly give weak students better chances of reaching the end of grade 9 with the required stock of knowledge and skills.

8 Conclusion

Grade retention is difficult to evaluate because grade repeaters have been selected on the basis of many characteristics that the econometrician doesn't observe. The difficult problem is to find a reasonable model to compute what would be the counterfactual performance of a student who has repeated a grade, if instead of being held back, he or she had been promoted to the next grade. To this end, we have assumed that the distribution of student test scores can be represented by a finite mixture of normal distributions, conditional on observed covariates, during each year of the observation period. The class size experienced by a student is also assumed to be distributed as a mixture of normals. All such mixtures are relying on the same finite number of latent student classes, called *groups*. In a flexible formulation, we show that class-size, probabilities of grade retention and test scores all depend on the unobserved group in a non-trivial and consistent way. We estimated a model with four groups and found that the four groups are unambiguously ranked. The higher the group index, the larger the student's ability, and the larger his class size. This proves that class size is endogenous, smaller classes being used by school principals to redistribute resources towards weaker students. With the help of our model, we computed counterfactual test scores to evaluate the average treatment effect and the average treatment effect *on the treated* of grade retention. We found that the *ATE* is negative, while the *ATT* is generally positive, but small. We computed treatment effects in each student group separately, and found that the *ATE* is positive for less able students and negative for more able students. Finally we computed the *ATT* and *ATE* of grade retention on the probability of access to grade 9, and found that this effect is significant and negative. Grade retention is a form of remedial education and seems to help the weakest students, insofar as it tends to increase their test scores at the end of grade 9. But these effects are weak. It follows that grade retention could probably be replaced by a form of tracking, or by different forms of remediation. Other studies have shown that grade retention is a stigma, that repeated years are interpreted as a negative signal by employers (on this point, see [Brodaty et al. \(2012\)](#)). The long-run effects of grade retention seem to be detrimental. We can only conclude that grade retention is unlikely to be an efficient public policy, because its impact on student performance — when positive — is weak.

Bibliography

- E. Alet. Is grade repetition a second chance? *manuscript, Toulouse School of Economics*, 2010.
- El. Alet, L. Bonnal, and P. Favard. Repetition: Medicine for a short-run remission. *Document de travail*, 2012.
- J.D. Angrist and A.B. Krueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991.
- J.D. Angrist and V. Lavy. Using maimonides’ rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2):533–575, 1999.
- P. Arcidiacono and J.B. Jones. Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, 71(3):933–946, 2003.
- S. Baert, B. Cockx, and M. Picchio. On track mobility, grade retention and secondary school completion. 2013.
- K. Bedard and E. Dhuey. The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics*, 121(4):1437–1472, 2006.
- S. Bonhomme and J-M. Robin. Assessing the equalizing force of mobility using short panels: France, 1990–2000. *The Review of Economic Studies*, 76(1):63–92, 2009.
- T. Brodaty, R.J. Gary-Bobo, and A. Prieto. Does speed signal ability: The impact of grade retention on wages. *Document de travail, Crest-Ensaie*, 2012.
- P. Carneiro, K.T. Hansen, and J.J. Heckman. 2001 lawrence r. klein lecture estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review*, 44(2):361–422, 2003.
- J. Cooley, S. Navarro, and Y. Takahashi. How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects. 2011.
- F. Cunha and J. Heckman. The technology of skill formation. Technical report, National Bureau of Economic Research, 2007.
- F. Cunha and J.J. Heckman. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782, 2008.
- F. Cunha, J.J. Heckman, and S.M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.
- G. De Fraja, T. Oliveira, and L. Zanchi. Must try harder: Evaluating the role of effort in educational attainment. *The Review of Economics and Statistics*, 92(3):577–597, 2010.
- X. d’Haultfoeuille. A new instrumental method for dealing with endogenous selection. *Journal of Econometrics*, 154(1):1–15, 2010.

- Y. Dong. Kept back to get ahead? kindergarten retention and academic performance. *European Economic Review*, 54(2):219–236, 2010.
- E.R. Eide and M.H. Showalter. The effect of grade retention on educational and labor market outcomes. *Economics of Education Review*, 20(6):563–576, 2001.
- R.J. Gary-Bobo and M.B. Mahjoub. Estimation of class-size effects, using maimonides’ rule and other instruments: The case of french junior high schools. *Annals of Economics and Statistics*, 111:193–225, 2013.
- J.B. Gomes-Neto and E.A. Hanushek. Causes and consequences of grade repetition: Evidence from brazil. *Economic Development and Cultural Change*, 43(1):117–148, 1994.
- J. Grenet. Academic performance, educational trajectories and the persistence of date of birth effects. evidence from france. *Unpublished manuscript*, 2009.
- J.J. Heckman. Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature*, 48(2):356–398, 2010.
- J.J. Heckman and E. Vytlacil. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–738, 2005.
- T.C. Holmes. Grade level retention effects: A meta-analysis of research studies. *Flunking grades: Research and policies on retention*, 16:33, 1989.
- T.C. Holmes and K.M. Matthews. The effects of nonpromotion on elementary and junior high school pupils: A meta-analysis. *Review of Educational Research*, 54(2):225–236, 1984.
- C.M. Hoxby. The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics*, 115(4):1239–1285, 2000.
- G.W. Imbens and J.D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.
- B.A. Jacob and L. Lefgren. Remedial education and student achievement: A regression-discontinuity analysis. *Review of economics and statistics*, 86(1):226–244, 2004.
- B.A. Jacob and L. Lefgren. The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, pages 33–58, 2009.
- M.B. Mahjoub. The treatment effect of grade repetitions. *Paris School of Economics, manuscript*, 2007.
- M.B. Mahjoub. Essai en microéconométrie de l’éducation. *PhD dissertation, University of Paris 1 Panthéon-Sorbonne, under the supervision of R. Gary-Bobo*, 2009.
- M. Manacorda. The cost of grade retention. *Review of Economics and Statistics*, 94(2):596–606, 2012.
- D. Neal and D.W. Schanzenbach-Whitmore. Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, 92(2):263–283, 2010.

T. Piketty and M. Valdenaire. *L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français: estimations à partir du panel primaire 1997 et du panel secondaire 1995*. Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche, Direction de l'évaluation et de la prospective. France, 2006.

1 Appendix: Details on Quarter of Birth as an Instrument for Grade Retention

Table 14 displays descriptive statistics on value-added. Scores in grade 6, ranging between 0 and 20, as is usual in French schools, are standardized to have a mean of 50 and a standard deviation of 10 in the whole sample in grade 6 (including all redirected pupils). Scores in grade 9 are standardized in the same way in the sample of individuals who reached grade 9. Table 14 shows that value added, the sign of which is irrelevant because scores are measures of performance relative to each grade, is nevertheless higher for repeaters than for non-repeaters. This is true both in French and Mathematics.

Table 14: Descriptive Statistics for Value Added

Standardized score	Math		French	
	Balanced sample ^a	Repeaters	Balanced sample ^a	Repeaters
Grade 6	51.10 (9.55)	43.25 (8.48)	51.21 (9.47)	43.38 (8.44)
Grade 9	50 (10)	43.37 (8.23)	50 (10)	43.46 (7.87)
VA = Grade 9 – Grade 6	–1.10 (8.55)	0.11 (9.63)	–1.21 (8.39)	0.08 (9.18)

Note ^a. Sample of all pupils for whom a test score is available both in grade 6 and in grade 9.

There exists a strong link between the age of a child, as measured by the month of birth, or quarter of birth, and the probability of grade repetition. A look at Figure 5 shows the frequency of grade retention by quarter of birth⁹. The probability of grade retention is clearly higher for children born later in the year. In principle, children must be 6 years old on September 1st of year t to be admitted in primary school, grade 1, year t . In practice, many 5-year-old children born between October and December are admitted, but the 5-year-old children born in the first quarter typically have to wait until the next year. It follows that first-quarter students tend to be relatively older in their class, with an age difference that can reach 11 months. Older children being more mature, they tend to perform better. At the same time, teachers are reluctant to retain older children in one grade as retention may change a difference —being older— into a stigma —being too old.

Figure 6 shows that initial (grade 6 entry) scores decrease with quarter of birth. The decreasing trend also exists for final scores but is less pronounced. Figure 7 shows that value-added scores tend to be higher for relatively younger students, who seem to be catching up during their junior high-school years. In a first attempt to check if this is attributable to grade retention, we plot value-added by quarter of birth separately for repeaters and non-repeaters. Figure 8 clearly shows that value-added age profiles are steeper for repeaters than for non-repeaters.

⁹Due to the survey protocol, there are no observations for students born in March, July and October.

Figure 5: Probability of Grade Retention by Quarter of Birth

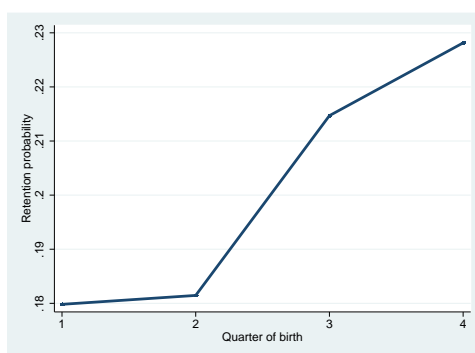


Figure 6: Scores by Quarter of Birth

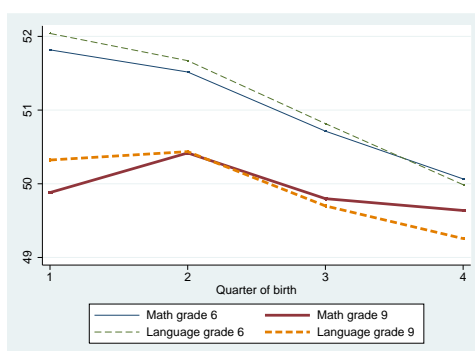


Figure 7: Value Added by Quarter of Birth

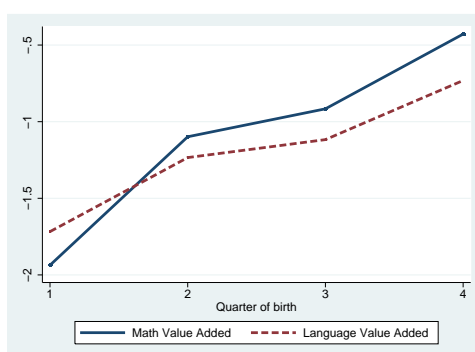
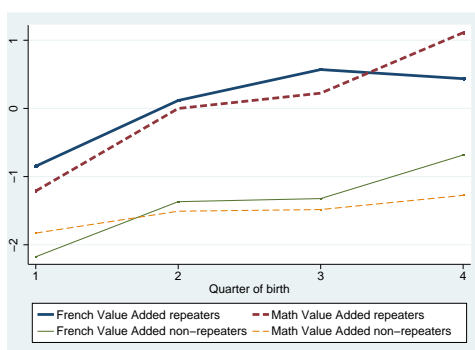


Figure 8: Value Added by Quarter of Birth for Repeaters and Non-Repeaters



Part IV

Conclusion

Conclusion

I wish these four chapters have contributed to convince the reader that the structural economic approach of social behaviors which appear in marriage and education is relevant and allows the derivation of the aggregate effect of social behaviors on economic outcomes.

Several lessons may be learned from this dissertation on the formation and organization of families on one hand, and on the efficiency of education systems to counteract the inequalities formed by the family environment on another hand.

First, structural models of marriage which take into account the preferences of individuals but also the constraints imposed by the marriage market reveal that there exists an attraction of individuals for educated and rich partners. As education is the main determinant for economic success, people look first for a high educated partner. However, the first chapter shows that as people age, they are more interested in their partner's wage than in their education level. People seem to be less patient overtime and more interested by material issues.

Actually, I show in the second chapter that in addition the companionship of their partner, the material gain of living in couples through joint consumption and joint production is an important part of the overall interest for people to live in couple. Income sharing between members and the quantity of time they choose to devote to domestic work or to take care of children, both depend on their relative wages and also on their bargaining power. The second chapter proposes a model whose ultimate goal is to evaluate the impact of a family policy program on labor supplies of men and women and on the resources devoted to children (time and expenditure).

Second, the analysis of schooling performances of children highlights their strong dependence on the family background. Education systems fail in giving the same learning opportunities to all children. Programs which target particularly disadvantaged children seem to be inefficient in improving their results. The third chapter even points out that these programs are less and less efficient overtime in France, which leads to an increase in inequalities of performances. A closer look to the effect of grade retention practice in the fourth chapter shows that its effect is not the same on different types of individuals. The less able individuals may benefit from grade retention but not enough to maintain this practice at the expense of other more efficient education policy reforms.

