Essais d'économie appliquée sur l'éducation, l'information et les salaires

Thibault Brodaty

ECOLE d'ÉCONOMIE DE PARIS
UNIVERSITÉ PARIS-I PANTHÉON-SORBONNE
UFR de Sciences-Economiques

Thèse pour le Doctorat ès Sciences Economiques
Présentée et soutenue publiquement par

Thibault Brodaty

le 8 décembre 2008

Essais d'Economie Appliquée sur l'Education, l'Information et les Salaires

---

Directeur de Thèse

M. Jean-Marc Robin, Professeur à l'Ecole d'Economie de Paris

Jury

Rapporteurs

Mme. Brigitte Dormont	Professeur, Université Paris-Dauphine
M. Thierry Magnac	Professeur, Toulouse School of Economics

Suffragants

M. Denis Fougère	Directeur de recherche, CNRS
M. Robert Gary-Bobo	Professeur, Ecole d'Economie de Paris

Résumé

Cette thèse présente trois essais en microéconométrie de l'éducation. Dans le premier chapitre nous estimons les effets de pairs et les effets maîtres dans le cadre particulier d'une université française. L'hypothèse identifiante d'affectation aléatoire des étudiants dans les groupes est testée et n'est pas rejetée par les données. Les effets estimés sont significatifs, d'amplitude comparable et robustes. Les étudiants les moins bons bénéficient de la présence dans leur groupe des étudiants les meilleurs, sans que cela nuise à ces derniers. Par conséquent, il est optimal de mixer des étudiants de différente qualité dans les groupes, et nous montrons qu'instituer des groupes de niveau ferait chuter de manière substantielle la note moyenne des étudiants les moins bons.

Dans le deuxième chapitre nous travaillons sur l'impact de l'imperfection d'information sur les choix d'éducation. Nous nous plaçons dans le cadre des modèles structurels dynamiques de choix d'éducation. Nous remettons en cause l'idée selon laquelle au moment de faire leur choix les individus connaissent parfaitement leurs propres caractéristiques sur le marché du travail. Nous supposons que les individus forment leurs croyances en fonction de leur aptitude et leur goût à l'école, qu'ils connaissent. Ces croyances sont révisées de manière bayésienne à chaque observation d'un salaire. Nous montrons qu'environ 20% des individus feraient des choix différents en information parfaite. Les différences de niveau d'éducation ne seraient cependant en général que d'une année d'études en plus ou en moins.

Dans le dernier chapitre, nous introduisons de l'hétérogénéité inobservée dans la mobilité des salaires en modélisant la dynamique des quintiles de salaires à l'aide d'un logit multinomial dynamique avec hétérogénéité inobservée. L'estimation semi-paramétrique du modèle met à jour une segmentation de la distribution des quintiles, fortement corrélée avec le niveau d'éducation des individus. Le caractère asymétrique des matrices de transition conditionnelles implique le rejet des modélisations ARMA standard gaussiennes de la dynamique des salaires. Nous confirmons que la mobilité salariale ne réduit quasiment pas les inégalités salariales de long terme entre groupes, mais qu'au sein de chaque groupe elle les réduit plus ce qui avait été précédemment estimé avec de l'hétérogénéité observée.

Mots clés : Microéconométrie, Education, Effets de Pairs, Effets Maîtres, Modèles Structurels Dynamiques de Choix Discret, Mobilité Salariale, Hétérogénéité Inobservée.

Abstract

This work contains three essays in Microeconometrics of Education. In the first chapter we estimate peer effects and teacher effects in a French university. Random assignment is tested and is not rejected. Teacher effects are significative and the weakest students benefit from the presence of the best students in their group, which is not detrimental for the latter. These results are robust. Consequently, the optimal policy is to mix students, and we show that segmenting students by level would imply a substantial fall of the weakest students average results.

In the second chapter we study the impact of imperfect information on educational choices using an educational choice dynamic structural model. We discard the idea that individuals know perfectly their own characteristics on the labor market when they make their decision. We suppose that they build their initial a priori in accordance with their observed ability in school. Then these beliefs are updated in a bayesian way each time the individual observes a wage. We show that around 20% of individuals would educate differently with perfect information, but generally only one year more or one year less.

In the third chapter, we introduce unobserved heterogeneity in earnings mobility by modelling earnings quintiles dynamics with a dynamic multinomial logit with unobserved heterogeneity. The semi-parametric estimation shows a segmentation of earnings quintiles distribution, very correlated with individuals education. Conditional transition matrices asymmetry leads to reject earnings dynamics modelisations that use standard gaussian ARMA processes. Moreover we confirm that the reduction in long term earnings inequality due to earnings mobility is almost totally driven by a reduction in within-group inequality. Also, we find a more important reduction of within-group inequality than previously estimated with observed heterogeneity.

Keywords : Microeconometrics, Education, Peer effects, Teacher effects, Educational Choice Dynamic Structural Models, Earnings Mobility, Unobserved Heterogeneity.

A mes parents,

A mon frère,

A ma grand-mère.

Remerciements

Mes premiers remerciements sont adressés à Jean-Marc Robin. Tout d'abord pour les conseils décisifs qu'ils m'a donnés en tant que directeur de thèse. Ensuite pour tout ce que j'ai appris à son contact en tant que co-auteur du travail dont est issu le deuxième chapitre de cette thèse. Je lui témoigne toute ma reconnaissance pour la confiance qu'il m'a accordée et pour m'avoir toujours incité à aller plus loin. Je remercie également son tableau Véléda sans lequel rien n'aurait été possible.

Je dois également beaucoup à Marc Gurgand, co-auteur du travail dont est issu le premier chapitre de cette thèse. Je le remercie sincèrement de m'avoir fait bénéficier de sa grande connaissance de l'économie de l'éducation, et de ses immenses qualités à la fois de rigueur et de compréhension profonde, intuitive, de l'économétrie. Les effets de pairs ont sans aucun doute joué en ma faveur.

Je tiens à exprimer toute ma gratitude aux membres de mon jury pour leur lecture attentive. Brigitte Dormont et Thierry Magnac ont accepté d'être rapporteurs de ma thèse, je leur en suis très reconnaissant. Les commentaires et les suggestions dont j'ai bénéficié lors de la pré-soutenance m'auront permis d'améliorer significativement la qualité de ce travail.

L'Université Paris-Dauphine m'a donné accès aux notes et à certaines caractéristiques de ses étudiants. Je remercie Laurent Batsch et Bernard de Montmorillon, nouveau et ancien présidents de l'Université pour avoir donné leur accord ; Martine Bellec pour m'avoir aidé à convaincre de l'intérêt de ce projet ; Bernard Guillochon pour avoir accepté de fournir les notes et les emplois du temps des étudiants du département dont il est responsable ; Jean-Marie Janod pour m'avoir transmis les données, pour les heures qu'il a passées à les mettre en forme, et à m'expliquer les codes des variables et l'évolution administrative de l'Université ; enfin Daphnée Martin pour nous avoir reçu et nous avoir expliqué concrètement comment se déroulait l'affectation des étudiants dans les groupes.

Je remercie l'Université Paris-1 Panthéon-Sorbonne de m'avoir accordé une allocation de recherche durant les premières années de ma thèse. Je remercie à nouveau l'Université Paris-Dauphine de m'avoir accordé le statut de moniteur puis d'ATER. J'ai eu la chance d'enseigner à Dauphine au sein d'équipes pédagogiques dynamiques et stimulantes. Je tiens à remercier leur responsable, Geneviève Pons pour l'analyse et l'algèbre, Dominique Pujal pour l'algèbre, Claudine Dhuin pour les statistiques, Pierre Bezbakh et Pierre Levy pour l'histoire de la pensée économique.

La réalisation de ma thèse est intimement liée au Centre de Recherche en Economie et Statistique. Je remercie Stéphane Grégoir, ancien directeur du CREST, Francis Kramarz, nouveau directeur, et Thierry Kamionka, directeur du Laboratoire de Microéconométrie, de m'avoir accueilli pendant toute la durée de ma thèse. J'y ai incontestablement bénéficié des effets de pairs d'un laboratoire de recherche de niveau international et de conditions matérielles exceptionnelles. Je remercie les membres de l'équipe administrative du laboratoire, en particulier Nadine Guedj, Dominique Idir et Danièle Porez pour leur efficacité et leur gentillesse. Je remercie enfin tous ceux qui ont rendu le quotidien intéressant, studieux, agréable et plein d'humour. Je remercie en particulier Abla Safir, Christelle Dumas, Elise Coudin, Luc Behagel, Nicolas Jacquemet, Marc Ferracci, Marion Leturcq, Raùl Sampognaro, Grégory Jolivet, Jérôme Le, Frédérique Savignac, Rana Hendy, Sylvie Blasco, Arnaud Maurel, Laurent Bach, Julien Guitard, Nicolas Dromel, Pierre-Yves Cabannes, Adam Booij, Edwin Leuven et Roland Rathelot.

Mes dernières pensées vont vers mes proches.

Table des matières

Introduction générale **21**

1 Effets Maître et Effets de Pairs dans l'Enseignement Supérieur : le Cas d'une Université Française **29**

1.1 Introduction 29
1.2 Contexte institutionnel et données 32
1.3 Cadre économétrique 34
1.3.1 Le modèle 34
1.3.2 Utiliser l'Analyse des Correspondances Multiples pour contruire l'indice de qualité 38
1.3.3 Identification 39
1.4 Résultats empiriques 42
1.5 Conclusion 47
1.6 Annexe 63
1.6.1 Interpréter les paramètres d'effets de pairs 63
1.6.2 Redistribution des groupes de L1 en L2 65

2 L'impact de l'Imperfection d'Information sur les Choix Educatifs : une Approche Structurelle **67**

2.1 Introduction 67

2.2 Un modèle d'investissement en capital humain avec information imparfaite et apprentissage bayésien 69
2.2.1 Structure générale 69
2.2.2 Spécifications des utilités instantanées 74
2.2.3 Déroulement du temps et ensemble d'information 76
2.2.4 Apprentissage bayésien 80
2.2.5 Equations de Bellman 82
2.2.6 Résolution du modèle 84
2.3 Identification 90
2.4 Estimation 93
2.5 Données 95
2.5.1 La base de données et la définition des variables 96
2.5.2 Statistiques descriptives 97
2.6 Résultats 98
2.6.1 Qualité de l'ajustement du modèle 98
2.6.2 Les paramètres et leurs interprétations 100
2.6.3 Analyse contrefactuelle 103
2.7 Conclusion 105

3 Introduire de l'Hétérogénéité Inobservée Dans la Mobilité des Salaires : une Approche Semi-Paramétrique 121

3.1 Introduction 121
3.2 Description des données utilisées 124
3.2.1 La base de données 124
3.2.2 Statistiques descriptives 127
3.3 Cadre Econométrique 129
3.3.1 Environnement théorique 129

3.3.2 Modéliser les niveaux ou les rangs ? 130
3.3.3 Le modèle 133
3.3.4 Identification et Estimation 135
3.4 Résultats 139
3.4.1 Nombre de types et ajustement du modèle 139
3.4.2 Dépendance d'état 140
3.4.3 Hétérogénéité inobservée 145
3.4.4 Exploitation du modèle 151
3.5 Conclusion 157

Conclusion générale **159**

Bibliographie **162**

Liste des tableaux

1.1 Qualité des étudiants, par cohorte 49

1.2 Sélection de l'échantillon 50

1.3 Interpéter les paramètres d'effets de pairs dans le cas discret 51

1.4 Tests de Kolmogorov-Smirnov d'affectation aléatoire, L1 52

1.5 Tests de Kolmogorov-Smirnov d'affectation aléatoire, L2 53

1.6 Hétérogénéité de la qualité des groupes, cas continu 54

1.7 Hétérogénéité de la qualité des groupes, cas discret 55

1.8 Effets de pairs, cas continu 56

1.9 Effets de pairs, cas discret 57

1.10 Effets maîtres, cas discret 58

1.11 Effets maîtres, cas discret, suite 59

1.12 Effets maîtres, cas continu 60

1.13 Effets maîtres, cas continu, suite 61

1.14 Note moyenne des étudiants de faible qualité en 2002/2003, analyse contre-factuelle 62

1.15 Redistribution des groupes de L1 en L2, 2003 65

1.16 Redistribution des groupes de L1 en L2, 2004 66

2.1 Distribution du niveau d'éducation final 107

2.2 Distribution des choix à chaque âge 108

2.3 Les retours à l'école et leur nombre 109

2.4 Distribution du niveau d'éducation au moment du retour à l'école 110

2.5 Distribution du nombre d'années travaillées avant de retourner à l'école . . 111

2.6 Distribution du nombre d'années d'études pendant le retour à l'école 112

2.7 Estimations des paramètres et écarts-types 113

2.8 Distribution des TypeSH conditionnellement au niveau d'éducation initial . 114

2.9 Distribution du type sur le marché du travail conditionnellement au TypeSH 114

2.10 Distribution des niveaux d'éducation finaux conditionnellement aux types . 115

2.11 Distribution des retours conditionnellement aux types 116

2.12 Distribution des différences de niveau d'étude en information parfaite et imparfaite 116

2.13 Distribution des niveaux d'éducation en information imparfaite conditionnellement aux différences d'éducation information parfaite/inparfaite 117

2.14 Distribution des modifications d'études en information parfaite/imparfaite conditionnellement aux erreurs d'anticipations 118

2.15 Distribution des modifications d'études en information parfaite/imparfaite conditionnellement aux types d'individus 119

3.1 Matrices de transition inter-quintiles, 1970-1998 128

3.2 Qualité de l'ajustement du modèle, années 70 140

3.3 Qualité de l'ajustement du modèle, années 80 141

3.4 Qualité de l'ajustement du modèle, années 90 141

3.5 Paramètres de dépendance d'état, années 70 143

3.6 Paramètres de dépendance d'état, années 80 143

3.7 Paramètres de dépendance d'état, années 90 144

3.8 Tests de Wald de présence de dépendance d'état 144

3.9 Matrices de transitions par type, années 70 147
3.10 Matrices de transitions par type, années 80 148
3.11 Matrices de transitions par type, années 90 149
3.12 Distributions stationnaires des quintiles, par type 150
3.13 Corrélation entre les types inobservés et les caractéristiques observables . . . 152
3.14 Part de la dépendance d'état dans le logarithme du ratio stabilité/baisse . . 154
3.15 Effet égalisateur de la mobilité salariale 157

Introduction générale

Cette thèse s'insère dans le champs de recherche de la microéconomie de l'éducation. L'étude de la demande d'éducation est centrée sur la théorie de l'investissement en capital humain, initialement développée par Mincer (1958), Becker (1964) et Ben Porath (1967). L'idée générale de cette théorie est qu'un individu, pour décider de s'éduquer ou non, compare le coût instantané d'une année d'étude supplémentaire avec les gains salariaux futurs qu'engendrerait cette dernière, grâce à une augmentation de productivité[1]. Le coût instantané se divise en coûts directs comme les frais d'inscription ou les achats de fourniture, et le coût indirect, ou coût d'opportunité, qui représente le salaire dont on se prive en étudiant une année supplémentaire. Cette théorie a donné lieu à un nombre gigantesque de travaux qui ont tenté de mesurer les rendements de l'investissement en capital humain en termes salariaux[2], mais aussi à des évaluations des effets externes de l'éducation en termes de santé ou de délinquance par exemple. Par ailleurs, l'imperfection des marchés financiers est un sujet crucial dans la mesure où l'existence de contraintes de crédit impliquerait des investissements sous-optimaux. Une autre branche de la microéconomie de l'éducation s'intéresse à l'offre d'éducation, et plus précisément travaille sur l'efficacité du système éducatif. Y

[1]Ce lien entre capital humain et productivité est au coeur de cette théorie, mais il s'agit d'une hypothèse qui fait encore débat aujourd'hui, les tenants des modèles de signalement (Spence (1973), Arrow (1973), Stiglitz (1975), Riley (1976)) insistant plutôt sur l'idée que comme les employeurs connaissent mal la productivité des salariés, ils utilisent l'éducation comme un signal. Il est ainsi possible selon eux qu'à l'extrême, l'éducation n'ait aucun effet sur la productivité des salariés, mais que l'équilibre soit tel que les plus éduqués, plus productifs par nature, soient mieux rémunérés que les moins éduqués.

[2]Travaux qui ont d'ailleurs donné lieu à d'importantes avancées économétriques.

sont traitées des questions relatives à la taille des classes, aux effets de pairs, à la sélection des étudiants, aux étudiants salariés, à la formation et aux modes de rémunération des enseignants, aux frais d'inscription ou encore l'impact des ressources des établissements scolaires. Nous nous proposons dans cette thèse de travailler à la fois sur l'offre et la demande d'éducation. Dans un premier chapitre, nous étudierons certaines propriétés de la fonction de production d'éducation à l'université. Nous travaillerons ensuite sur les problèmes d'information que peuvent rencontrer les étudiants au moment de faire le choix d'arrêter ou de continuer leurs études. Nous analyserons enfin les effets des investissements en capital humain en termes de mobilité salariale.

Le premier chapitre de cette thèse est dédié à l'étude conjointe des effets de pairs et des effets maîtres dans l'enseignement supérieur. Alors que cette question est traitée de manière extensive pour les enseignements primaires et secondaires, les études concernant l'université sont très rares. En vue d'étudier l'efficacité de la fonction de production universitaire, Il est pourtant essentiel de savoir s'il existe des intéractions entre les étudiants et s'il existe de l'hétérogénéité dans la manière dont les enseignants préparent leurs étudiants aux examens. La nature des intéractions entre les étudiants, s'il y en a, auront des conséquences sur la manière de répartir les étudiants dans les groupes, alors qu'une éventuelle hétérogénéité dans l'enseignement posera la question des moyens nécessaires à son harmonisation vers le haut.

Arcidiaccono et al. (2007), De Giorgi et al. (2006) et De Paola et Scoppa (2007) détectent la présence d'effets de pairs, alors que Martins et Walker (2006) n'en trouvent pas. Cependant, ces travaux ne documentent pas comment ces effets de pairs pourraient affecter différemment des étudiants d'aptitude différente. Au contraire, cette question est investiguée en détails par Levy (2003), Sacerdote (2001), Winston et Zimmerman (2004) et Zimmerman (2003) qui trouvent des effets non linéaires, mais pas d'une façon systématiquement robuste. De plus, cette série d'études est attentive aux intéractions sociales entre étudiants qui partagent la même chambre sur le campus universitaire, pas le même groupe de Travaux Dirigés, ce qui pourrait être moins pertinent pour l'organisation de l'éducation. Enfin, aucun de ces travaux ne contrôle d'un éventuel effet enseignant, ce qui pourrait être de nature à biaiser les estimations.

Dans ce chapitre nous utilisons la structure particulière d'enseignement en petits groupes de l'Université Paris-Dauphine pour identifier conjointement les effets de pairs et les effets maîtres. Les intéractions entre étudiants sont ainsi observées au sein du groupe de Travaux Dirigés dans lequel l'étudiant reste toute l'année, et nous autorisons les effets de pairs à être non linéaires. Nous observons les deux premières années universitaires (L1, L2). Nous disposons d'informations sur les caractéristiques des étudiants au lycée, comme la mention au bac ou le nombre de fois où ils ont redoublé. Ces informations permettent de créer un indice de qualité des étudiants en L1. A la fin de la première année, une note globale permet de définir la qualité des étudiants en L2. Ces qualités définies a priori permettent d'appréhender la qualité de chaque groupe et ainsi d'étudier la présence d'effets de pairs. L'hypothèse cruciale qui nous donne l'identification des effets de pairs est celle d'affectation aléatoire dans les groupes. Cette hypothèse est testée, n'est pas rejetée, et permet donc d'être confiant sur la stratégie d'identification.

Nous détectons des effets de pairs et des effets maîtres de même amplitude. Les effets de pairs sont non linéaires. Plus précisément, il apparaît que, de manière robuste entre le L1 et L2, les moins bons étudiants bénéficient de la présence des meilleurs étudiants dans leur groupe, alors que les meilleurs étudiants ne sont pas affectés par la présence des moins bons. La stratégie optimale de répartition des groupes est donc de mixer dans des proportions égales dans chaque groupe les bons et les moins bons étudiants. Des simulations montrent qu'en affectant de manière optimale plutôt qu'aléatoire les étudiants aux groupes, la note moyenne des moins bons étudiants de la cohorte augmenterait de 10%, alors que si des groupes de niveaux étaient institués, cette note moyenne pourrait chuter jusqu'à 40%.

Dans le deuxième chapitre nous nous intéressons à la demande d'éducation, et plus précisément à la décision individuelle de s'éduquer. Depuis le travail fondateur de Keane et Wolpin (1997), la théorie de l'investissement en capital humain s'est traduite en microéconométrie structurelle par l'utilisation de modèles structurels dynamiques de choix d'éducation. Le grand avantage d'estimer des modèles structurels est de disposer de paramètres clairement interprétables économiquement et qui permettent d'effectuer des simulations de situations contrefactuelles. D'un autre côté, les coûts élevés en temps de programmation et en temps de calcul rendent l'exercice délicat et expliquent le nombre très faible de travaux réalisés dans ce domaine, comparé à la très populaire méthode des variables instrumentales par exemple.

De manière conforme avec la théorie du capital humain, les individus prennent en compte leurs salaires futurs espérés pour faire leur choix d'éducation. Ils utilisent pour cela une équation de salaire Mincerienne avec la plupart du temps une constante spécifique à chaque individu et parfois des rendements de l'éducation hétérogènes (Belzil et Hansen, 2007). Dans l'immense majorité des articles de cette littérature, une des hypothèses cruciales est que les individus connaissent parfaitement tous les paramètres de cette équation. Ils connaissent en particulier leur propre constante spécifique, interprétée comme l'aptitude sur le marché du travail ionbservée par l'économètre. Cette hypothèse est très forte. Comment en effet un individu qui n'a jamais fait l'expérience du marché du travail pourrait-il connaître sa propre aptitude sur ce marché ? Deux séries de travaux, Cunha et Heckman (2007) d'une part et Belzil (2007) d'autre part, ont commencé à travailler sur cette question. Leur apport est de proposer pour la première fois une modélisation de cette information imparfaite, les individus étant alors supposés avoir plus modestement des a priori qui leur permettent de prévoir la valeur des paramètres qu'ils ne connaissent pas. Mais d'une part, la manière dont ces a priori sont formés n'est pas expliquée par leur modèle ; d'autre part, reconnaître l'existence de problèmes d'imperfections d'informations impose de prendre en compte la façon dont cette information est apprise, ce qui n'est pas fait chez ces auteurs.

Notre travail propose d'apporter des solutions à ces deux questions. D'une part nous supposons que l'individu connaît parfaitement son aptitude à l'école et sa corrélation avec son aptitude sur le marché du travail, ce qui lui permet de former son a priori initial. D'autre part nous introduisons de l'apprentissage : à chaque passage sur le marché du travail, l'individu utilise son salaire, reflet de sa productivité, pour actualiser de manière bayésienne ses croyances sur ses propres caractéristiques. Les données utilisées sont celles du National Longitudinal Survey of Youth (1979) qui permet de suivre une cohorte d'individus entre 16 et 28 ans à partir de 1979. L'estimation est réalisée par maximum de vraisemblance. Le premier enseignement de ce travail est que l'information parfaite est rejetée par les données. Il apparaît donc nécessaire de prendre en compte cette dimension dans la modélisation des choix d'éducation. Le deuxième enseignement est que ces problèmes d'information sont tels que 20% des individus font des choix différents de ceux qu'ils feraient en information parfaite. Ces choix ne diffèrent cependant que d'une ou deux années d'études.

Partant de l'idée que les investissements en capital humain sont réalisés en vue d'atteindre un certain niveau de salaire, et donc un certain rang dans l'échelle des salaires, il est naturel de se demander à quel point les individus conservent ce rang au cours du temps. Cette question est celle de la mobilité des salaires, question d'autant plus importante qu'elle a des conséquences sur les inégalités salariales de long terme (Atkinson et al, 1992). Pour illustrer cette idée considérons un exemple simple. Considérons deux individus 1 et 2, deux périodes 1 et 2, supposons qu'il existe deux salaires $\underline{w}$ and $\overline{w}$ avec $\underline{w} < \overline{w}$, et que ces deux salaires restent constants entre les deux périodes. Supposons que lors de la première période les individus 1 et 2 reçoivent respectivement $\underline{w}$ et $\overline{w}$. Deux cas sont alors possibles : soit il n'y a aucune mobilité salariale entre les deux périodes, auquel cas les individus conservent leur rang et leur salaire en période 2 ; soit il existe de la mobilité salariale, auquel cas les individus échangent leur salaire entre les deux périodes. Dans le premier cas, sans mobilité salariale, le salaire moyen de l'individu 1 sur les deux périodes vaut $\underline{w}$ et celui de l'individu 2 vaut $\overline{w}$. Dans le deuxième cas, avec mobilité salariale, le salaire moyen sur les deux périodes des deux individus est égal et vaut $\frac{\underline{w}+\overline{w}}{2}$, de sorte que la dispersion des salaires moyens sur les deux périodes est nulle. Il est donc clair que la distribution des salaires moyens sur les deux périodes est moins dispersée s'il y a de la mobilité. C'est le sens de l'idée selon laquelle la mobilité salariale réduit les inégalités de long terme.

L'augmentation brutale des inégalités salariales dans les années 80 dans les pays anglo-saxons en particulier (Katz et Autor, 1999) a ainsi engendré une littérature importante sur la mobilité salariale, dont l'objectif était de savoir si une éventuelle hausse du degré de mobilité n'avait pas limité les effets de l'augmentation des inégalités. Nous savons aujourd'hui que la réponse est négative (Levy et Murnane, 1992). Cette littérature s'est également posée la question de l'existence de différents niveaux de mobilité selon les caractéristiques des individus. Gittleman and Joyce (1995) ont trouvé que les jeunes, les moins éduqués et les afro-américains sont plus mobiles que ceux qui sont plus vieux, plus éduqués ou blancs. Schiller (1994) a montré que dans les années 80 les jeunes femmes ont plus de mobilité descendante et moins de mobilité ascendante que les jeunes hommes. Kopczuk et al. (2007) ont confirmé le résultat que les femmes montent moins fréquemment que les hommes dans l'échelle des salaires. Ainsi nous savons aujourd'hui que la structure de la mobilité des salaires n'est pas la même selon les différents groupes démographiques. Buchinsky et Hunt (1999) ont de plus montré que la mobilité salariale réduit beaucoup plus les inégalités au sein de chaque groupe que les inégalités entre groupes.

Nous nous proposons dans ce dernier chapitre d'étudier la structure de la mobilité salariale, non pas uniquement en fonction de caractéristiques observables comme le sexe, la race, l'éducation ou l'expérience, mais plus généralement sur la base de l'hétérogénéité ionbservée. Plutôt que de postuler a priori une certaine partition de la population, nous laissons notre procédure d'estimation s'en charger. Nous utilisons pour cela les données du Panel Study of Income Dynamics entre 1970 et 2000. Nous introduisons de l'hétérogénéité inobservée dans les matrices de transition inter-quintiles en modélisant la dynamique des quintiles de salaire par un logit multinomial dynamique avec hétérogénéité inobservée. Nous estimons dans une première étape les paramètres de dépendance d'état grâce à la méthode semi-paramétrique du maximum de vraisemblance conditionnel (Magnac, 2000). Nous utilisons ensuite ces estimations pour estimer la loi de l'hétérogénéité inobservée, supposée discrète. En estimant 6 types d'individus chaque décennie, nous montrons que le processus de mobilité est fortement hétérogène. De plus, ce qui caractérise cette hétérogénéité inobservée n'est pas tant un degré de mobilité plus ou moins grand selon les groupes d'individus, mais le fait que chaque type d'individu gravite autour d'un quintile particulier, ce qui a pour effet de segmenter la distribution des quintiles. Nous montrons également que cette segmentation est très corrélée avec des caractéristiques observables comme le sexe, la race ou l'éducation. Ces résultats ont deux implications importantes. Tout d'abord, les matrices de transition conditionnelles estimées n'étant pas symétriques, nous sommes amenés à rejeter les modélisations ARMA gaussiennes standard de la dynamique des salaires. D'autre part nous confirmons que, sur la base de cette partition de la population, la mobilité ne réduit pratiquement pas les inégalités entre groupes mais qu'elle réduit plus les inégalités intra-groupes que ce qui avait été estimé précédemment.

Chapitre 1

Effets Maître et Effets de Pairs dans l'Enseignement Supérieur : le Cas d'une Université Française

1.1 Introduction

Dans la période récente l'étude des effets de pairs dans l'éducation s'est beaucoup développée. Ceci a plusieurs raisons. D'abord, si les effets de pairs sont présents, les politiques publiques d'éducation sont renforcées par un effet dit de *multiplicateur social*. Mais cette question est également décisive dans le débat sur la ségrégation à l'école. D'un point de vue positif, les effets de pairs sont importants pour les modèles qui expliquent la stratification et la sélection à l'école. Par exemple, les effets de pairs sont la source d'un équilibre stratifié dans le modèle d'Epple et Romano (1998), à condition que les bons étudiants valorisent plus d'être entourés de bons étudiants que ne le font les moins bons d'entre eux, une structure très spécifique. D'un point de vue normatif, mixer les classes a un impact qui dépend de manière cruciale de la forme des effets de pairs, car réorganiser les étudiants engendrerait des gagants et des perdants. Mixer les classes peut également être désirable pour des raisons d'efficacité : il est généralement efficace de générer des groupes hétérogènes si les effets de pairs sont plus forts pour les moins bons étudiants. Comme le montrent Arnott et Rouse (1987), les décisions optimales sont très sensibles à la fonction de production d'éducation si le planificateur social doit décider simultanément l'allocation des étudiants et des ressources. Par conséquent, l'existence, l'importance et les détails de la structure des effets de pairs sont des questions décisives pour l'organisation de l'éducation.

Dans ce contexte général, l'enseignement supérieur pose des questions spécifiques. L'université est le plus souvent beaucoup plus sélective que l'éducation obligatoire : les effets de pairs expliquent-ils cette situation, et est-il efficace que les universités soient fortement stratifiées ? D'autre part, beaucoup de pays élargissent l'accès à l'enseignement supérieur : les résultats en sont-ils modifiés en raison de l'environnement des pairs ? Bien qu'il y ait une importante littérature empirique sur les effets de pairs dans l'éducation, dont une certaine partie est attentive à la forme et aux éventuels effets non-linéaires, il y a peu de travaux réalisés sur l'enseignement supérieur. Arcidiaccono et al. (2007), De Giorgi et al. (2006) et De Paola et Scoppa (2007) détectent la présence d'effets de pairs, alors que Martins et Walker (2006) n'en trouvent pas. Cependant, ces travaux ne documentent pas comment ces effets de pairs pourraient affecter différemment des étudiants d'aptitude différente. Au contraire, cette question est investiguée en détails par Levy (2003), Sacerdote (2001), Winston et Zimmerman (2004) et Zimmerman (2003) qui trouvent des effets non linéaires, mais pas d'une façon systématiquement robuste. De plus, cette série d'études est attentive aux intéractions sociales entre étudiants qui partagent la même chambre sur le campus universitaire, pas le même groupe de Travaux Dirigés, ce qui pourrait être moins pertinent pour l'organisation de l'éducation. Enfin, aucun de ces travaux ne contrôle d'un éventuel effet enseignant, ce qui pourrait être de nature à biaiser les estimations.

Dans ce chapitre nous estimons les effets de pairs au niveau de la classe dans une université française, en autorisant les effets à être non-linéaires, c'est-à-dire que tous les types d'étudiants sont susceptibles d'intéragir de manière spécifique. Plus spécifiquement, cette université a vu sa population évoluer rapidement en termes de qualité sur une courte période de temps. Les meilleurs étudiants ont-ils souffert d'une plus faible sélection ? Les moins bons étudiants (restants) ont-ils bénéficié de ce meilleur environnement ? Naturellement, la validité externe des effets de pairs estimés est sujette à caution. Cette université est fortement sélective par rapport aux standards français, et tout l'enseignement est effectué en petit groupe, ce qui est une situation exceptionnelle. Par conséquent, ni la technologie ni la population ne sont représentatives. Cependant, ce contexte nous fournit un laboratoire intéressant pour en apprendre plus sur la structure profonde des effets de pairs. Enfin, contrairement au reste de la littérature existante, nous estimons conjointement les éventuels effets maîtres qui peuvent biaiser l'estimation des effets de pairs, et qui fournissent une référence pour juger de l'importance des pairs.

Nous savons depuis l'article fondateur de Manski (1993) que l'identification des effets de pairs est difficile. Nous ne cherchons pas à estimer les *effets endogènes*, c'est-à-dire les effets du comportement courant, nous ne considérons qu'une forme réduite et nous estimons l'impact de mesures d'aptitude prédéterminées. La difficulté réside malgré tout dans la séparation des effets des caractéristiques des pairs des effets de groupes inobservés. Nous argumenterons que, dans cette université, la formation des groupes peut-être considérée comme aléatoire, et nous le testerons. Ceci fournit l'identification de manière directe.

Dans la section 1.2, nous présentons le contexte institutionnel et les données, dans la section 1.3 nous introduisons le modèle et nous discutons la stratégie d'identification, nous fournissons les résultats empiriques dans la section 1.4 et nous concluons dans la section 1.5.

1.2 Contexte institutionnel et données

Nous considérons les deux premières années (L1 et L2) d'une université française publique qui compte 700 à 800 étudiants par cohorte. Les étudiants sont affectés à des groupes d'environ 30, et la totalité de l'enseignement est founie au groupe. Il n'y a pas de cours dispensés à l'ensemble de la cohorte et les groupes sont fixés pour la totalité de l'année. C'est une situation tout à fait exceptionnelle dans le système universitaire français, et c'est ce qui rend possible d'observer conjointement les effets de pairs et les effets enseignants dans l'enseignement supérieur.

Chaque année, pour le L1 et le L2, nous utilisons en tant que résultats les notes d'examen final dans 7 matières : Maths, Microéconomie 1 et Microéconomie 2, Macroéconomie 1 et Macroéconomie 2, Statistiques et Informatique. Beaucoup d'enseignants enseignent la même matière dans plusieurs groupes, et occasionnellement enseignent différentes matières. Les notes dans chacune des 7 matières sont basées sur un examen général commun à tous les étudiants. Ces notes sont donc comparables entre groupes. Cependant les examens sont différents d'une année sur l'autre, de sorte que la comparaison entre les cohortes ne participe pas à l'identification. Parce que l'on accorde une seconde chance en septembre aux étudiants qui échouent, nous distinguerons parfois entre les résultats de septembre et ceux de juin.

Nous considérons les années 2002 à 2006. La pratique et le statut légal de cette université est cependant spécifique et il a évolué durant cette période. En 2002, elle sélectionnait les étudiants, mais un nombre non négligeable d'entre eux réussissait à détourner le processus. En 2003, à l'inverse, les autorités administratives ont forcé l'université à recruter de manière sectorielle, ce qui a conduit cette année-là à une cohorte composée de 50% d'étudiants qui auraient été sélectionnés et de 50% d'étudiants qui ne l'auraient pas été car détenteurs d'un dossier insuffisant. Enfin, à partir de 2004, l'université sélectionne de nouveau et il est devenu extrêmement rare de pouvoir contourner la sélection. La conséquence de l'évolution de ce recrutement est impressionnante. La Table 1.1 montre les performances au lycée de chaque cohorte. Les trois premières lignes de la table donnent la distribution de la mention au Baccalauréat. La part des mentions Passable a décru de 34%à 5%, alors que la part des mentions Bien ou Très Bien a augmenté de 18% à 67%. L'année 2003 constitue clairement une cassure dans cette tendance, et ceci est dû cette année-là à la part importante des étudiants qui n'auraient pas été sélectionnés si le processus de sélection avait été en vigueur. La part des individus qui ont redoublé au moins une fois au lycée suit un mouvement similaire, et la part des étudiants qui redoublent leur L1 suit la tendance de manière retardée, bien que de façon atténuée. On ne distingue cependant pas de telle tendance concernant la série du baccalauréat (Economie, Scientifique option Maths ou autre option), ce qui reflète la volonté de l'université d'équilibrer ces populations.

Dans un tel contexte, il est intéressant de réfléchir à l'impact social d'une telle politique, au moins en relation avec la composition des pairs. Existe-t-il des gains visibles pour les meilleurs ou les moins bons étudiants à avoir évolué vers une cohorte d'étudiants plus homogène et de meilleure qualité ? Il est bien connu que s'il existe des effets de pairs non-linéaires, la composition des pairs a un impact en terme d'efficacité. Bien sûr, nous ne pouvons fournir qu'une évaluation incomplète du caractère désirable d'un point de vue social d'un tel changement dans la mesure où nous n'avons aucune information sur les résultats des étudiants hors de cette université.

Nous avons collecté les données de L1 sur la période 2002-2006 ; nous ne suivons en L2 que les deux cohortes entrées en 2002 et 2003 (donc observées en 2003 et 2004). Certains étudiants ont rempli un dossier d'admission lors du processus de sélection : nous avons pour ceux-là leurs caractéristiques relatives au lycée. Nous n'en disposons pas pour ceux qui ont intégré l'université sans même avoir déposé de dossier d'admission. Le premier bloc de la Table 1.2 décrit l'échantillon brut. En 2003, la moitié des étudiants ont intégré l'université alors qu'ils n'avaient pas été sélectionnés, alors que cette proportion n'est que d'environ 2% en 2005 et 2006. Une partie des étudiants inscrits n'a pas participé aux examens en L1 (7% à 8% et presque 13% en 2003). Cette situation est beaucoup plus rare en L2. Nous excluons de l'échantillon ces non-participants, bien qu'il n'est pas évident de savoir à quel point ils n'ont pas participé aux cours et donc aux effets de pairs.

Le second bloc présente l'échantillon dit initial, c'est-à-dire l'échantillon brut duquel on a éliminé les non-participants. En L1, il y a des données manquantes sur les performances académiques au lycée pour les étudiants qui n'ont pas soumis de dossier. Il y en a 4% à 5% en début de période et seulement 1% à 2% à la fin. En L2, la performance passée est basée sur la note totale obtenue en L1 : parce qu'un certain nombre d'étudiants intègrent l'université directement en L2, cette information est manquante pour eux. Ceci représente une part importante de la cohorte : 10% à 20%. Nous sommes contraints de supprimer ces observations de l'échantillon, ce qui est de nature à générer de l'erreur dans la mesure de la qualité des groupes.

L'échantillon final, composé des participants et des étudiants dont on a les caractéristiques passées, est décrit dans le troisième bloc de la Table 1.2. Nous avons 22 à 27 groupes chaque année avec 25-30 étudiants par groupe, ce qui représente 3457 étudiants en tout en L1 et 1136 en L2. L'évolution de la qualité générale de l'échantillon est illustrée par la part d'étudiants non-sélectionnés, les notes et le pourcentage d'individus qui n'ont pas leur année.

1.3 Cadre économétrique

1.3.1 Le modèle

Dans cette section nous présentons comment nous modélisons les notes des étudiants. Considérons un individu i de la cohorte c, dans le groupe g. Notons y_{igmkc} la note de l'étudiant i, avec l'enseignant k dans la matière m. Les notes dépendent essentiellement de la qualité de i, de la qualité de ses pairs dans son groupe et de son enseignant dans la matière m. Supposons que l'on observe un indice de qualité continu I_i pour chaque individu

i. Nous utiliserons cet indice d'une part dans un premier modèle sous sa forme continue, et d'autre part dans un deuxième modèle sous une forme discrétisée, basée sur ses quantiles.

Modèle avec l'indice de qualité continu

Le modèle est

$$y_{imgkc} = a + \alpha I_i + \beta \overline{I_g^{-i}} + \gamma I_i * \overline{I_g^{-i}} + \delta_k + \gamma_{mc} + \nu_i + \epsilon_{imgkc} \tag{1.1}$$

où I_i est l'indice de qualité de l'étudiant i. Pour faciliter l'interprétation, l'indice est normalisé avec un minimum de 0 et une variance égale à 1. $\overline{I_g^{-i}}$ est la qualité moyenne du groupe g, en excluant l'individu i. δ_k est l'effet de l'enseignant k ; γ_{mc} est l'effet de la matière m pour la cohorte c, ceci prend en compte le fait que la difficulté des examens et les standards de notation sont potentiellement différents entre matières et cohortes ; ν_i est un effet individuel aléatoire reflétant la part inobservée de la qualité de l'étudiant i et ϵ_{imgkc} est un choc iid. Les paramètres du modèle sont facilement interprétables. L'effet d'une augmentation d'un écart-type de la qualité moyenne du groupe $\overline{I_g^{-i}}$ sur la note vaut, conditionnellement à la qualité de l'étudiant i, $\beta + \gamma I_i$. Ainsi, β est l'effet sur l'étudiant le plus faible (avec $I_i = 0$) et γ nous dit comment cet effet change quand la qualité de l'étudiant i augmente.

Tout d'abord remarquons que la qualité de l'étudiant ne dépend pas de la matière : nous supposons que les étudiants ont les mêmes capacités en Maths, Microéconomie, Macroéconomie, Statistique et Informatique. L'objectif de cette hypothèse raisonnable est d'augmenter la longueur du panel.

Dans cette spécification les effets de pairs sont entièrement pris en compte par les variables de qualité moyenne. Comme Arcidiacono et al. (2007) les notes des autres étudiants du groupe ne sont pas incluses, ce qui veut dire que nous ne considérons pas d'effets endogènes (Manski, 1993), mais plutôt que nous estimons une forme réduite qui mixe les effets endogènes et exogènes.

Les effets maîtres ne dépendent pas de la cohorte c, ce qui veut dire que la qualité des enseignants est constante dans le temps. Par conséquent nous ne prenons pas en compte de potentiels effets d'expérience dans la qualité des enseignants, et donc ces effets doivent être interprétés comme une qualité moyenne sur la période observée. De plus, ils ne dépendent pas non plus de la matière m. En pratique, ceci implique que des enseignants qui sont présents dans plusieurs matières sont considérés comme des enseignants différents dans chaque matière. Cette hypothèse est faite car nous n'observons pas suffisamment de combinaisons enseignants/matières. Par conséquent nous ne pouvons pas comparer les enseignants dans des matières différentes et nous pouvons uniquement tester l'égalité de la qualité des enseignants au sein de chaque matière. Ainsi, dans ce travail, un enseignant sera dit de meilleure qualité qu'un autre si toute chose égale par ailleurs[1], la note moyenne de son groupe est significativement supérieure à celle de l'autre enseignant. Il doit donc être clair qu'ici, la qualité des enseignants est définie, de manière restrictive, par leur aptitude à préparer leurs étudiants aux examens courants.

Enfin, nous supposons que toutes les variables explicatives sont indépendantes de ν_i et ϵ_{imgkc}. L'hypothèse d'indépendance entre la qualité moyenne du groupe et l'effet individuel aléatoire, cruciale pour l'identification, sera justifiée dans la section suivante sur la base de l'allocation aléatoire des étudiants dans les groupes.

Modèle avec les quantiles de l'indice de qualité

L'indice de qualité I peut être discrétisé par rapport à certains pourcentiles pour créer un indice de qualité discret avec trois catégories, faible (L), moyen (M) et fort (H). Nous utiliserons les indices basés sur les découpages 25/75 et 33/66. Ainsi, les étudiants dont l'indice de qualité se situe en dessous du 25ième pourcentile seront classés comme faibles, ceux dont l'indice se situe entre les 25ième et 75ième pourcentiles seront classés comme moyens, et comme forts au dessus du 75ième pourcentile. Il en est de même pour l'indice

[1] En particulier la qualité du groupe.

basé sur les pourcentiles 33 et 66.

Le modèle est

$$
\begin{aligned}
y_{imgkc} = & a + \alpha_L \mathbb{1}_{\{I_i=L\}} + \alpha_H \mathbb{1}_{\{I_i=H\}} + \beta_L P_{gL}^{-i} + \beta_H P_{gH}^{-i} \\
& + \beta_{LL} \mathbb{1}_{\{I_i=L\}} * P_{gL}^{-i} + \beta_{LH} \mathbb{1}_{\{I_i=L\}} * P_{gH}^{-i} + \beta_{HL} \mathbb{1}_{\{I_i=H\}} * P_{gL}^{-i} + \beta_{HH} \mathbb{1}_{\{I_i=H\}} * P_{gH}^{-i} \\
& + \delta_k + \gamma_{mc} + \nu_i + \epsilon_{imgkc}
\end{aligned}
\tag{1.2}
$$

où $\mathbb{1}_{\{I_i=L\}}$ et $\mathbb{1}_{\{I_i=H\}}$ sont des variables indicatrices indiquant si la qualité de i est respectivement faible ou forte ; P_{gL}^{-i} et P_{gH}^{-i} sont respectivement les proportions de faibles et de forts dans son groupe, en l'excluant ; nous utilisons les mêmes trois quantiles et les mêmes intéractions que Sacerdote (2001).

Les paramètres correspondant aux effets de pairs (β_L, β_H, β_{LL}, β_{LH}, β_{HL}, β_{HH}) ne sont pas facilement interprétables sous cette forme. Mais il est clair que les intéractions entre les niveaux de qualité et les proportions de chaque niveau dans le groupe permettent d'étudier des effets de pairs non-linéaires, c'est-à-dire l'effet d'un niveau de qualité sur tous les autres niveaux. Les effets de pairs ne sont pas supposés a priori être les mêmes pour tous les étudiants. Pour faciliter l'interprétation des paramètres, nous montrons dans l'annexe 1.6.1 que l'effet de l'échange d'un point de pourcentage d'étudiants d'un certain niveau par un point de pourcentage d'un autre niveau sur un étudiant d'un certain niveau s'exprime en fonction des paramètres du modèle. Ces expressions sont présentées dans la table 1.3.

Les deux régressions considérées dans ce chapitre sont des modèles de panel linéaires à effets aléatoires. Nous les estimons par Moindres Carrés Généralisés avec des écarts-types robustes.

1.3.2 Utiliser l'Analyse des Correspondances Multiples pour contruire l'indice de qualité

Pour les étudiants de L2 le choix d'une mesure de qualité est évident. A la fin du L1, l'administration calcule une note sur 200 basée sur les notes d'examen et les notes de contrôle continu. Les étudiants qui ont au moins 100 passent en L2. Ceux qui en juin ont échoué sont autorisés à passer en septembre une session de rattrapage pour améliorer leur score. Ces deux scores reflètent l'aptitude des étudiants à la fin du L1 et peuvent donc être utilisés pour approximer leur qualité en L2.

Pour les étudiants de L1, nous disposons d'informations sur leur parcours dans l'enseignement secondaire (série du bac, mention au bac, redoublements éventuels), sur la manière dont ils ont intégré l'université (sélectionné ou non) et s'ils sont en train de redoubler leur L1. Notre objectif est d'utiliser toutes ces informations. La première idée est d'inclure de manière brute toutes ces variables : insérer toutes les indicatrices pour la qualité de l'étudiant, et les proportions de chacune de ces caractéristiques dans le groupe pour juger de la qualité de celui-ci. Ceci permettrait d'estimer l'effet de chacune de ces variables. Mais comme la taille des groupes tourne autour de trente étudiants, il n'y aurait qu'un ou deux étudiants (0 parfois) dans chaque ''case", et l'estimation serait soit impossible soit complètement imprécise. La faible taille des groupes nous empêche donc d'adopter cette solution. L'idée est donc de créer un indice qui résume l'ensemble de cette information pour n'avoir à inclure qu'une seule variable dans la régression. On peut penser à la solution suivante. Régresser dans un premier temps les notes des étudiants sur toutes les caractéristiques, calculer la valeur prédite ''$X_i\widehat{\beta}$" et l'utiliser comme indice de qualité. Bien que séduisante, cette procédure aurait le mauvais goût d'inclure une variable explicative fortement endogène, puisque calculée préalablement avec les notes. Il faut donc trouver un indice qui ne soit pas corrélé de manière préalable avec les résultats des étudiants, c'est-à-dire un indice qui résume les caractéristiques des étudiants indépendamment de leurs notes. Ceci peut être réalisé grâce à l'Analyse des Correspondances Multiples de l'ensemble des caractéristiques, en utilisant les coordonnées des individus sur son premier axe. C'est la solution que nous adoptons.

1.3.3 Identification

Affectation aléatoire dans les groupes

Le point principal concernant l'identification des effets de pairs est la question de l'affectation aléatoire dans les groupes. Imaginons par exemple que les étudiants s'auto-sélectionnent dans les groupes et que cette sélection soit basée sur leur aptitude. Dans cet exemple extrême, les bons étudiants se retrouvent avec les bons étudiants et les moins bons également entre eux. Il y a donc une corrélation positive entre la qualité individuelle des étudiants et la qualité moyenne de leurs pairs. ν_i représentant dans les deux modèles (1.1) et (1.2) la part inobservée de l'aptitude, il y a dans ce cas une corrélation positive entre ν_i et la qualité moyenne observée du groupe, ce qui biaise à la hausse l'estimation des effets de pairs. Si l'affectation dans les groupes est aléatoire vis-à-vis de la qualité des étudiants, de tels effets ne peuvent pas exister et les estimations sont sans biais.

Voici comment est réalisée concrètement l'affectation dans les groupes. Chaque année, certaines matières sont obligatoires (celles que nous analysons) et d'autres (que nous ne prenons pas en compte) sont optionnelles. Les étudiants doivent choisir un certain nombre de matières optionnelles, et juste avant le commencement de l'année, ils remettent à l'administration leur liste de choix, dans l'ordre de leur préférence car tous n'obtiendront pas leurs premiers choix. Ensuite, l'administration constitue les groupes en essayant de satisfaire au mieux les choix des étudiants tout en étant sous la contrainte que chaque groupe n'est compatible qu'avec 4 ou 5 options. Les groupes sont remplis séquentiellement dans l'ordre d'arrivée des demandes, et chaque groupe est ''fermé'' à chaque fois qu'un ensemble de 5 étudiants y a été affecté. Cette procédure administrative séquentielle interdit toute auto-sélection dans les groupes et nous amène à penser que l'affectation est aléatoire.

Pour vérifier statistiquement si c'est réellement le cas, nous proposons un moyen de le tester. L'idée est d'utiliser un générateur de nombres aléatoires pour allouer nous-mêmes les étudiants dans les groupes et de tester si les distributions observées et simulées de la qualité des groupes peuvent être considérées comme étant issues de la même distribution. Si la réponse est positive, cela signifiera que le processus d'allocation réel ne peut pas être distingué statistiquement d'une allocation réalisée avec un générateur de nombres aléatoires. Considérons par exemple la variable indicatrice qui caractérise si les étudiants ont redoublé au moins une fois au lycée. Nous pouvons calculer la proportion q de redoublants dans chaque groupe et obtenir ainsi un échantillon de proportions observées $S = \{q_g\}_{g=1...G}$, où G est le nombre total de groupes. Lorsque nous simulons une première fois l'allocation des étudiants dans les groupes, nous obtenons un premier échantillon de proportions simulées $S_1^* = \left\{q_g^{*1}\right\}_{g=1...G}$. En répétant R fois cette opération nous obtenons un large échantillon de proportions simulées $S^* = \{S_1^*, ..., S_R^*\}$. Le test consiste à comparer les distributions des échantillons S and S^*. Si nous ne pouvons pas rejeter que ces deux échantillons sont issus de la même distribution, nous concluerons que le processus réel d'affectation est aléatoire. Nous utilisons le test bien connu de Kolmogorov-Smirnov appliqué à deux échantillons. Les résultats de ces tests sont présentés dans la table 1.4 pour les L1 et 1.5 pour les L2. Concernant les L1 les tests sont réalisés séparément chaque année pour toutes les caractéristiques des étudiants. Pour les L2, les tests sont basés sur les indices de qualité discrétisés. Les nombres entre parenthèses sont les p-values. Les résultats ne peuvent pas être plus clairs : les tests ne rejettent jamais l'affectation aléatoire.

Concernant l'identification en L2 un autre point doit être vérifié. En L2, la qualité des étudiants est basée sur la note totale en L1. Supposons que les groupes de L1 et de L2 soient identiques. Il est clair que dans ce cas de potentiels effets de pairs en L1 engendreraient une corrélation entre la qualité moyenne des groupes de L2 et l'aptitude de l'étudiant, et en particulier avec sa part inobservée ν_i. Les effets de pairs seraient alors biaisés. Si les groupes sont complètement réalloués entre le L1 et le L2 ces effets n'existent pas. Nous vérifions que c'est bien le cas (voir annexe 1.6.2).

Hétérogénéité de la qualité des groupes

La qualité de l'estimation des effets de pairs dépend de manière basique de la présence d'hétérogénéité dans la composition des groupes. La table 1.6 montre la distribution de l'indice de qualité. Nous regroupons les cohortes 2002 et 2003 d'une part et les cohortes 2004 à 2006 d'autre part car chacune correspond à un régime différent vis-à-vis du processus de sélection. La valeur moyenne de l'indice en L1 est normalisée à 0, de sorte que les changements dans la qualité des étudiants lors de ces deux sous-périodes ne sont pas visibles dans cette table. Nous calculons ensuite la valeur moyenne de l'indice dans chaque groupe : la variance inter-groupes vaut 0.042 en 2002/2003 et représente 14% de la variance totale. Ceci implique que l'allocation aléatoire dans des groupes de 25 à 30 maintient des contrastes significatifs, de sorte que nous pouvons espérer estimer l'impact des effets de pairs. L'hypothèse que les moyennes des groupes sont égales est fortement rejetée. Pour les cohortes 2004/2006, la variance inter-groupes est plus faible et vaut 0.026, ce qui est le résultat direct de l'arrivée de cohortes plus homogènes. Comme nous le verrons plus tard, cette situation est moins favorable pour estimer précisément l'impact d'un changement de qualité des pairs entre groupes.

Les observations de L2 appartiennent toutes aux cohortes 2002/2003, et ont été sélectionnées en réussissant les examens de L1. Parce que les étudiants les plus faibles ont amélioré leur score en septembre par rapport à juin, les variances sont généralement plus faibles, au point que les différences entre groupes ne sont pas très significatives pour septembre.

1.4 Résultats empiriques

Les principaux résultats de ce chapitre sont présentés dans les tables 1.8 et 1.9. La table 1.8 présente une spécification simple avec la qualité de l'étudiant (standardisée), la qualité moyenne du groupe et la qualité de l'étudiant intéragit avec la qualité du groupe, de manière à capturer, d'une manière contrainte, de possibles effets non-linéaires. La qualité est soit l'indice de qualité basé sur les performances au lycée pour le L1 ou la note totale de L1 pour les L2. La variable explicative est la note sur 20 dans les 7 matières citées précédemment, et les contrôles sont une variable indicatrice pour chaque matière croisée avec l'année, et des indicatrices d'enseignant. Des estimations séparées sont réalisées pour la période agrégée 2002/2003 d'une part et pour la période 2004/2006 d'autre part en L1 ; en L2, nous présentons les résultats des régressions basées sur l'indice construit avec la note de juin d'une part et de septembre d'autre part.

Concernant le L1 en 2002/2003, nous voyons que l'effet direct d'une augmentation soit de la qualité de l'étudiant soit de la qualité moyenne du groupe est positif et significatif. L'intéraction, cependant, est significativement négative. A la valeur moyenne de la qualité de l'étudiant et de la qualité du groupe (2.8), augmenter sa propre qualité d'un écart-type augmente la note d'environ un point (sur 20). Par contre, augmenter la qualité du groupe de la même valeur (ce qui est énorme car c'est une moyenne de groupe) ne fait pas augmenter le résultat. La non-linéarité est donc très forte, et les effets de pairs sont positifs uniquement pour les étudiants en dessous de la qualité initiale moyenne. Se baser sur cette spécification, cependant, impliquerait des effets de pairs négatifs pour les bons étudiants, ce qui n'a pas beaucoup de sens. Nous considérons donc une spécification plus flexible plus bas.

Pour les L1 des cohortes 2004 à 2006, les effets de pairs ne sont pas significatifs, bien que les estimations soient comparables. Notre interprétation est qu'il y a significativement moins d'hétérogénéité dans ces données, de sorte que la composition des groupes est moins contrastée et les tests ont moins de puissance, comme on le voit avec des écarts-types systématiquement plus grand.

Enfin, la même cohorte 2002/2003 maintenant observée en L2 présente des résultats comparables à ceux trouvés en L1, mais uniquement avec les indices de qualité construits sur la note de septembre en L1. Celle-ci est la note finale, et donc la mesure la plus complète des résultats des étudiants, spécialement des plus faibles. L'indice est maintenant simplement basé sur la note totale à la fin du L1. La moyenne de cette variable vaut 1.6 en septembre : ceci implique qu'une augmentation d'un écart-type de la qualité de l'étudiant à la composition moyenne du groupe augmente la note de 1.6 points sur 20. Une augmentation similaire de la qualité du groupe augmente la note de 1.2 points pour l'étudiant moyen. Bien qu'encore présentes, les non-linéarités sont moins marquées.

Il est important que les résultats de L1 et L2 en 2002/2003 soient qualitativement similaires, bien que d'amplitudes différentes, car les groupes sont complètement réalloués. Par conséquent, nous avons deux allocations des groupes indépendantes qui mènent à des effets de pairs similaires. Ceci indique que nos résultats sont robustes.

La table 1.9 présente les résultats d'une spécification plus flexible. Rappelons que nous avons classifié tous les étudiants en trois catégories basées sur l'indice initial : les meilleurs, ceux qui sont au dessus des 75% dans leur cohorte (H), les plus faibles, ceux qui sont en dessous des 25% (L) et les moyens, les 50% restants (M). Nous définissons également trois catégories en utilisant les seuils 33% et 66%. Dans chaque classe nous avons calculé le pourcentage de faibles, moyens et forts. Comme indiqué précédemment, une spécification flexible, étant donnée la quantité limitée de données, intéragit l'impact de chacun de ces trois pourcentages avec la qualité de l'étudiant. Ceci nous permet d'étudier la distribution des effets de pairs, plutôt que seulement la moyenne, et d'autoriser les effets de pairs à être non-linéaires dans le sens où la qualité des pairs n'affecte pas les individus de la même manière.

Les coefficients d'une telle régression ne sont pas interprétables directement, à cause des normalisations. A la place, la table 1.9 présente les effets qu'impliquent des changements marginaux dans la composition des pairs sur les résultats des autres types d'étudiants, faibles, moyens et forts.

Concernant les L1 en 2002/2003, on observe que les effets de pairs affectent significativement la performance des étudiants uniquement de faible qualité. Réduire d'un point de pourcentage la part des étudiants de qualité faible et le remplacer par un point d'étudiants de qualité élevée, augmentent leur résultat de 0.047 points. En d'autres termes, substituer un étudiant (environ 3% d'un groupe) de faible à fort augmente la performance des faibles de 0.15 points (sur 20). Remplacer un moyen par un fort est un changement moins puissant : les effets sont parfois significatifs, mais d'amplitude plus faible. Par contre, les étudiants de qualités moyenne et forte ne semblent pas être affectés par des changements dans la composition des pairs.

Les effets correspondants en utilisant les cohortes 2004/2006 sont plus petits et ne sont pas significatifs. Rappelons que les catégories, faible-moyen-fort sont basées sur les quantiles de la cohorte. Ainsi, un étudiant faible à 25% en 2004 est meilleur en valeur absolue qu'un tel étudiant en 2003 car la qualité générale de la cohorte était moins bonne en 2003. Les étudiants de faible qualité qui sont typiquement affectés par la présence de bons étudiants en 2002/2003 ont très peu de chances d'être présents parmi les étudiants faibles de 2004/2006. Ceci est une autre indication que les effets de pairs sont réellement effectifs sur le bas de la distribution.

Au niveau du L2, nous trouvons un impact significatif d'un changement marginal de la composition des pairs de faible à fort, pour les étudiants de faible qualité. Ainsi, substituer un étudiant augmenterait la note des plus faibles restants entre 0.1 et 0.15 points (sur 20). Un certain effet est trouvé de manière occasionnelle sur les meilleurs étudiants, ce qui est cohérent avec l'observation de la table 1.8 que les effets non linéaires, bien que présents, sont moins forts dans cet échantillon.

De manière générale, ces résultats mettent en lumière deux idées. Premièrement, avoir des pairs de bonne qualité plutôt que des moyens ou des faibles est une bonne chose , alors qu'être entouré d'étudiants faibles n'est pas pénalisant. Les changements dans le bas de la distribution ne semblent pas avoir d'impact. Deuxièmement, les effets de pairs ont l'air d'affecter principalement les étudiants les plus faibles. Ceci implique que, pour une cohorte d'étudiants donnée, il est désirable d'instituer des groupes hétérogènes : ceci bénéficie aux étudiants les plus faibles et ne pénalise pas les meilleurs.

Certains résultats de la littérature sur les effets de pairs non-linéaires sont compatibles avec ces résultats : Sacerdote (2001) trouve que partager sa chambre avec un bon étudiant bénéficie aux plus faibles, et moins aux bons étudiants eux-mêmes ; Winston et Zimmerman (2004) trouvent également que ce sont les étudiants faibles et moyens qui sont principalement affectés par la composition des pairs. Plusieurs de ces résultats ne sont cependant pas très robustes : un point intéressant de nos résultats est qu'ils sont robustes à travers différents échantillons. En dehors de la littérature sur l'enseignement supérieur, les résultats sont assez différents : Hoxby et Weingarth (2005) insistent sur l'importance d'une hétérogénéité limitée, alors que Lavy et al. (2007) trouvent que les bons étudiants ont une bonne influence sur les étudiants plutôt bons, alors que les étudiants les plus faibles ont une influence négative sur les étudiants moyens. Des processus assez différents pourraient être à l'oeuvre à différents niveaux du système éducatif. De fait, dans le contexte d'une université d'élite, nous ne nous attendons pas à ce que les moins bons étudiants perturbent le cours (Lazear, 2001), contrairement à ce qui peut se passer dans le primaire ou le secondaire.

Evaluer l'impact de sélectionner toujours de meilleurs étudiants dans l'université est une tâche plus compliquée. Les bons étudiants ne semblent pas clairement bénéficier d'une proportion plus grande de pairs de qualité. Alors que pour les étudiants les plus faibles, ceux qui ont suffisamment de chance pour accéder à l'université sous ce régime plus sélectif bénéficient clairement d'être dans de meilleures classes. Pour ceux qui sont exclus, il est difficile d'évaluer leur perte sans information supplémentaire sur leur destination.

Les tables 1.11, 1.12 et 1.13 fournissent un résumé des effets maîtres estimés dans le modèle. Tous les effets sont identifiés intra-matière, de sorte que nous n'essayons pas de comparer les efficacités des enseignants en Maths et en Statistique : la raison est que nous n'avons pas un nombre suffisant d'enseignants qui enseignent en même temps dans des matières différentes pour identifier proprement de tels paramètres. Ces tables présentent d'abord, pour chaque matière et chaque spécification du modèle un test d'égalité des effets maîtres. L'impact des enseignants est partout extrêmement significatif. Les tables reportent ensuite quelques éléments de la distribution de ces effets : l'impact d'avoir le meilleur plutôt que le pire enseignant et l'écart-type de la distribution des effets. Le contraste le plus fort est typiquement de 3 points en L1 et 2 points en L2. Mais l'écart-type vaut plus modestement entre 0.5 et 1 points. Pour comparaison, imaginons un changement de 12% d'une classe (environ 4 étudiants) d'étudiants faibles pour des étudiants meilleurs, ce qui est un contraste standard dans l'échantillon 2002/2003 : d'après nos estimations précédentes, ceci ferait augmenter la note d'environ 0.5 points. De ce point de vue, les effets de pairs et les effets maîtres sont d'amplitude comparable, ce qui implique qu'avoir des pairs de bonne qualité est aussi important que d'avoir de bons enseignants.

En 2002 et 2003, la proportion massive d'étudiants non sélectionnés crée suffisamment d'hétérogénéité pour estimer de manière précise les effets de pairs. Nous montrons ainsi que les étudiants les plus faibles bénéficient de la présence dans leur groupe des étudiants les meilleurs, sans que cela nuise à ces derniers. Pour maximiser la note moyenne des étudiants les plus faibles, il est alors optimal de mixer les étudiants dans les groupes de sorte que les proportions de bons et de mauvais étudiants soient exactement les mêmes dans tous les groupes. On peut alors se demander ce qu'auraient gagné en moyenne les étudiants les plus faibles de la cohorte si la répartition n'avait pas été faite de manière aléatoire mais de manière optimale. On peut également se demander ce qu'il en aurait été si des groupes de niveau avaient été institués, c'est-à-dire si on avait groupé les bons entre eux et les moins bons entre eux. Nous utilisons nos estimations pour simuler ces contrefactuels. Nous montrons dans la table 1.14 que si la répartition avait été optimale, la note moyenne des étudiants les plus faibles aurait augmenté environ de 10% en L1 et de manière plus modeste en L2, alors que si des groupes de niveau avaient été institués, les résultats moyens des plus faibles auraient chuté de manière substantielle.

1.5 Conclusion

Dans ce chapitre, nous estimons les effets de pairs dans l'enseignement supérieur dans une université française, en s'intéressant en particulier aux potentiels effets non-linéaires. De fait, les détails de la structure des effets de pairs, pas uniquement leur existence, sont importants pour l'analyse économique et les recommandations de politiques publiques. Or il n'y a qu'une quantité limitée de résultats empiriques, parfois contradictoires, dans cette littérature. Il est par conséquent urgent de fournir des résultats supplémentaires, en particulier au niveau de l'université, où les résultats trouvés à des niveaux plus bas ne sont sans doute pas généralisables.

L'identification des effets de pairs repose sur le fait que les groupes sont constitués par l'administration en suivant une règle qui ne devrait pas générer de regroupement des étudiants par qualité académique. Ceci est testé, et confirmé, en utilisant la distribution parmi les groupes des caractéristiques prédéterminées observées.

Nous trouvons que les effets de pairs sont présents, mais sont également hautement non-linéaires. Premièrement, la valeur moyenne n'est pas la seule à compter, mais aussi les changements dans leur distribution. La recomposition prenant place dans le bas de la distribution des qualités ne semble pas jouer de rôle. Ce qui compte vraiment est la part des étudiants faisant partie des 25% ou 33% meilleurs de la cohorte tout entière qui est présente dans la classe. Deuxièmement, cette composition de pairs semble être importante presque exclusivement pour les étudiants les plus faibles. Il est alors bien connu qu'il est optimal de mixer les étudiants.

Nos résultats sont robustes, dans le sens où ils sont présents dans différents échantillons de groupes. Ils ne sont cependant pas estimés très précisément en utilisant des cohortes assez homogènes. De plus, à cause de l'organisation pédagogique spécifique de cette université, nos résultats ne peuvent pas être généralisés sans précautions. Cependant, cette organisation nous permet d'estimer des effets de pairs dans un contexte d'intéractions sociales stable dans des petites classes, une situation dans laquelle les effets de pairs devraient être apparents s'ils sont importants.

Tab. 1.1 – Qualité des étudiants, par cohorte

		2002	2003	2004	2005	2006
Note	[10 12]	33.81	44.93	19.55	8.84	4.84
au bac	[12 14]	48.49	40.10	46.37	49.08	28.03
sur 20	[14 20]	17.69	14.98	34.08	42.08	67.13
A redoublé au moins	oui	11.53	16.97	11.10	7.39	4.67
une fois au lycée	non	88.47	83.03	88.90	92.61	95.33
	Economie	32.63	40.77	42.14	34.96	34.60
Série du bac	Science opt non Maths	29.23	25.46	27.87	35.49	33.39
	Science opt Maths	38.14	33.78	29.99	29.55	32.01
Redouble	oui	8.39	14.81	16.78	11.87	10.38
le L1	non	91.61	85.19	83.22	88.13	89.62

Tab. 1.2 – Sélection de l'échantillon

		L1					L2	
		02	03	04	05	06	03	04
Pre éch.	Taille d'éch.	870	738	849	817	635	748	627
	Non Sélec. (%)	14.3	50.7	17.3	1.8	2.1	10.6	29.0
	Non-participants (%)	8.2	12.7	7.7	6.0	6.9	1.5	2.1
Ech. Initial	Taille d'éch.	799	644	784	768	591	737	614
	Non Sélec. (%)	13.9	48.6	16.2	1.8	1.7	10.6	29.0
	Non-participants (%)	0	0	0	0	0	0	0
	Manquant (%)	4.1	5.8	3.2	1.2	2.1	9.5	23.0
	# Groupes	26	24	27	25	22	24	20
	Taille des groupes	30.7	26.8	29.0	30.7	26.8	30.7	30.7
		(2.2)	(2.3)	(1.3)	(1.4)	(1.9)	(1.5)	(1.3)
	Total Juin	105.5	97.9	100.6	117.3	118.6	141.3	138.8
	Total Sept	107	101.2	103.8	119.6	121.2	142.9	137.4
	Redouble (%)	14.3	25.16	19.0	12.0	11.7	4.3	21.7
Ech. Final	Taille des groupes	763	601	757	758	578	666	470
	Non Sélec.(%)	10.0	45.1	13.7	0.7	0	11.7	37.9
	Non-Participants (%)	0	0	0	0	0	0	0
	Manquant (%)	0	0	0	0	0	0	0
	# Groupes	26	24	27	25	22	24	20
	Taille des groupes	29.3	25.0	28.0	30.3	26.2	27.8	23.5
		(2.5)	(2.7)	(1.5)	(2.0)	(2.2)	(1.7)	(1.6)
	Total Juin	105.9	98.2	100.8	117.5	119.2	141.5	139.4
	Total Sept	107.4	101.6	103.9	119.7	121.6	143.2	138.1
	Redouble (%)	13.8	23.8	18.8	11.7	11.3	4.8	20.2

Note : Ecarts-types entre parenthèses.

TAB. 1.3 – Interpéter les paramètres d'effets de pairs dans le cas discret

	Sur	Effet
1pt de faibles en moins	faibles	$(\beta_H - \beta_L) + (\beta_{LH} - \beta_{LL})$
1pt de forts en plus	moyens	$\beta_H - \beta_L$
	forts	$(\beta_H - \beta_L) + (\beta_{HH} - \beta_{HL})$
1pt de moyens en moins	faibles	$\beta_H + \beta_{LH}$
1pt de forts en plus	moyens	β_H
	forts	$\beta_H + \beta_{HH}$
1pt de moyens en moins	faibles	$\beta_L + \beta_{LL}$
1pt faibles en plus	moyens	β_L
	forts	$\beta_L + \beta_{HL}$

Lecture : La première ligne signifie que si un point de pourcentage d'étudiants faibles est remplacé dans un certain groupe par un point d'étudiants de qualité élevée, alors l'effet sur la note des étudiants faibles restants dans ce groupe vaut $(\beta_H - \beta_L) + (\beta_{LH} - \beta_{LL})$, où les β sont définis par l'équation (1.2).

TAB. 1.4 – Tests de Kolmogorov-Smirnov d'affectation aléatoire, L1

		2002		2003		2004		2005		2006	
Sélectionné	oui	0.18	(0.42)	0.13	(0.79)	–	(–)	–	(–)	–	(–)
	non	0.15	(0.63)	0.14	(0.77)	–	(–)	–	(–)	–	(–)
Note au	[10 12]	0.07	(1)	0.14	(0.78)	0.07	(1)	0.08	(1)	0.06	(1)
bac	[12 14]	0.06	(1)	0.15	(0.68)	0.13	(0.79)	0.12	(0.87)	0.24	(0.16)
sur 20	[14 20]	0.14	(0.7)	0.07	(1)	0.12	(0.86)	0.08	(1)	0.18	(0.5)
A redoublé au moins	oui	0.13	(0.82)	0.15	(0.67)	0.14	(0.66)	0.15	(0.61)	0.06	(1)
une fois au lycée	non	0.13	(0.77)	0.13	(0.85)	0.15	(0.58)	0.15	(0.62)	0.06	(1)
	Economie	0.09	(0.99)	0.19	(0.39)	0.17	(0.42)	0.13	(0.83)	0.13	(0.83)
Série du bac	Science opt non Maths	0.12	(0.85)	0.12	(0.88)	0.11	(0.91)	0.07	(1)	0.13	(0.85)
	Science opt Maths	0.13	(0.82)	0.17	(0.55)	0.09	(0.99)	0.14	(0.76)	0.1	(0.98)
Redouble	oui	0.18	(0.37)	0.23	(0.17)	0.25	(0.08)	0.19	(0.37)	0.14	(0.8)
le L1	non	0.19	(0.31)	0.24	(0.15)	0.22	(0.16)	0.17	(0.48)	0.16	(0.64)

Note : Statistiques de Kolmogorov et p-value entre parenthèses.

TAB. 1.5 – Tests de Kolmogorov-Smirnov d'affectation aléatoire, L2

		Juin				Sept			
		2003		2004		2003		2004	
25/75	faibles	0.14	(0.78)	0.19	(0.46)	0.11	(0.95)	0.15	(0.76)
	moyens	0.1	(0.98)	0.16	(0.74)	0.15	(0.71)	0.18	(0.53)
	forts	0.12	(0.91)	0.19	(0.49)	0.08	(1)	0.19	(0.52)
33/66	faibles	0.11	(0.93)	0.14	(0.83)	0.16	(0.57)	0.18	(0.58)
	moyens	0.07	(1)	0.16	(0.74)	0.14	(0.77)	0.16	(0.73)
	forts	0.12	(0.88)	0.1	(0.99)	0.11	(0.93)	0.12	(0.96)
50	faibles	0.13	(0.82)	0.12	(0.94)	0.16	(0.61)	0.08	(1)
	forts	0.15	(0.64)	0.09	(1)	0.15	(0.71)	0.18	(0.58)

Note : Statistiques de Kolmogorov et p-value entre parenthèses.

Tab. 1.6 – Hétérogénéité de la qualité des groupes, cas continu

		L1		L2	
		02/03	04/05/06	Total Juin	Total Sept
Indice de qualité	moyenne	0	0	109.5	112.09
	Std	0.550	0.628	13.8	11.1
	Min	-1.539	-2.430	71	94
	Max	0.634	0.481	163	163
Moyenne des indices de qualité par groupe	Moyenne	-0.015	0.001	109.3	112.0
	Std	0.211	0.165	4.3	2.5
	Min	-0.429	-0.341	91.2	106.8
	Max	0.339	0.403	117.1	117.1
Egalité des moyennes	F-test	4.32	2.02	2.4	1.36
	p-value	< 0.0001	< 0.0001	< 0.0001	0.06
R^2		0.1387	0.0680	0.0862	0.0508
Variance inter groupes		0.042	0.026	16.45	6.27

Lecture : Dans le premier bloc ''Indice de qualité", nous donnons des statistiques qui résument l'indice de qualité pour chaque cohorte. Dans le second bloc, nous considérons la moyenne de l'indice de qualité dans chaque groupe et nous présentons des statistiques qui résument cette distribution. Ensuite nous régressons l'indice de qualité sur des indicatrices de groupe et nous testons l'égalité des paramètres de ces indicatrices. Nous reportons enfin les R^2 de ces régressions et la variance inter-groupes qui en résulte.

Tab. 1.7 – Hétérogénéité de la qualité des groupes, cas discret

		L1				L2			
		25/75		36/66		Total Juin		Total Sept	
		02/03	04/05/06	02/03	04/05/06	25/75	33/66	25/75	33/66
faibles	Moyenne	25.0	20.9	28.1	31.7	23.8	32.6	24.5	29.9
	Std	14.4	11.3	15.6	14.1	15.5	15.5	9.7	10.5
	Min	3.2	0	3.2	0	3.4	10.3	7.1	7.1
	Max	54.1	41.3	61.5	59.3	90.0	95.0	43.4	55.5
moyens	Moyenne	61.5	59.2	47.4	38.7	52.1	35.1	51.1	37.3
	Std	12.8	9.2	12.5	8.2	13.5	12.8	12.9	11.0
	Min	34.6	29.6	26.9	13.0	5.0	0	13.0	8.6
	Max	87.0	81.4	80.6	58.6	77.7	59.2	75.0	57.6
forts	Moyenne	13.3	19.7	24.4	29.5	24.0	32.1	24.3	32.7
	Std	6.4	12.6	11.7	15.4	10.0	10.8	9.6	10.2
	Min	0	3.4	0	6.8	4.5	5.0	4.5	13.6
	Max	29.0	70.3	48.3	74.0	48.2	55.1	48.2	55.1

Lecture : Dans cette table nous calculons les proportions de chaque niveau de qualité dans chaque groupe. Nous reportons des statistiques qui résument cette distribution. ''25/75'' signifie que l'indice de qualité est discrétisé par rapport aux 25ième et 75ième pourcentiles. ''Faibles'' signifie une qualité en dessous du 25ième pourcentile pour cette discrétisation.

Tab. 1.8 – Effets de pairs, cas continu

	L1		L2	
	02/03	04/05/06	Total Juin	Total Sept
Indice de qualité	2.51 ∗ ∗∗	1.84∗	1.72 ∗ ∗∗	2.43 ∗ ∗∗
	(.42)	(.88)	(.45)	(.39)
Qualité du groupe	1.45 ∗ ∗	1.41	.46	1.99 ∗ ∗∗
	(.48)	(1.01)	(.43)	(.45)
Indice de qualité	−.51 ∗ ∗	−.35	−.04	−.54∗
* Qualité du groupe	(.15)	(.23)	(.16)	(.23)

Note : Ecarts-types entre parenthèses. *, **, *** sont les niveaux de significativité à 5%, 1% et 0.1% respectivement.

Tab. 1.9 – Effets de pairs, cas discret

		L1				L2			
		25/75		36/66		Total Juin		Total Sept	
	On	02/03	04/05/06	02/03	04/05/06	25/75	33/66	25/75	33/66
	faibles	.047**	.010	.047***	.003	.025*	.024**	.046**	.038**
1pt de faibles		(.019)	(.015)	(.011)	(.009)	(.011)	(.009)	(.016)	(.012)
en moins	moyens	−.015	−.006	.001	−.014*	.013	.005	.020	.018
1pt de forts		(.013)	(.008)	(.009)	(.007)	(.009)	(.010)	(.011)	(.011)
en plus	forts	−.031	−.008	−.020	−.010	.008	.011	.001	.019
		(.026)	(.011)	(.012)	(.007)	(.012)	(.010)	(.015)	(.011)
	faibles	.014	−.011	.026*	−.009	.026*	.036**	.042***	.031**
1 pt de moyens		(.021)	(.015)	(.014)	(.011)	(.013)	(.011)	(.011)	(.010)
en moins	moyens	−.021	−.006	−.008	−.017	.005	.016	.021*	.021
1 pt de forts		(.013)	(.007)	(.009)	(.010)	(.009)	(.011)	(.009)	(.011)
en plus	forts	−.018	.000	−.014	−.004	.013	.029**	.017	.026*
		(.028)	(.010)	(.013)	(.010)	(.013)	(.011)	(.012)	(.011)
	faibles	−.033**	−.021	−.020	−.013	.001	.012	−.003	−.006
1 pt de moyens		(.010)	(.014)	(.011)	(.010)	(.007)	(.007)	(.011)	(.010)
en moins	moyens	−.006	.000	−.010	−.002	−.007	.011	.001	.002
1 pt de faibles		(.008)	(.008)	(.008)	(.010)	(.008)	(.010)	(.008)	(.010)
en plus	forts	.012	.009	.006	.006	.005	.018	.016	.007
		(.014)	(.015)	(.011)	(.012)	(.011)	(.010)	(.011)	(.010)

Note : Considérons un étudiant de qualité faible (<25ième pourcentile) en L1 en 2002/2003. Si l'on remplace dans son groupe un point de pourcentage d'étudiants faibles par un point d'étudiants forts, alors l'effet sur la note des étudiants faibles restants du groupe est de 0.047 points sur 20. Les écarts-types sont entre parenthèses. *, **, *** sont les niveaux de significativité à 5%, 1% et 0.1% respectivement.

Tab. 1.10 – Effets maîtres, cas discret

		L1				L2			
		25/75		36/66		Total Juin		Total Sept	
		02/03	04/05/06	02/03	04/05/06	25/75	33/66	25/75	33/66
Maths	Chi2	68.11	93.56	63.32	90.81	62.19	54.22	59.67	53.19
	P-value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0001	0.0006	0.0001	0.0008
	Max-Min	3.44	3.84	3.18	3.61	2.38	2.45	2.49	2.43
	Std	0.81	0.78	0.78	0.77	0.68	0.65	0.67	0.64
Micro1	Chi2	95.82	133.08	95.98	135.62	24.81	27.44	25.73	28.44
	P-value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0733	0.0369	0.0580	0.0280
	Max-Min	3.59	2.66	3.55	2.7	1.26	1.4	1.21	1.45
	Std	1	0.87	1	0.87	0.36	0.38	0.37	0.4
Macro1	Chi2	42.37	56.67	42.70	55.56	68.55	70.47	73.05	68.19
	P-value	0.0006	< 0.0001	0.0005	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	Max-Min	1.95	1.94	1.98	1.95	2.48	2.64	2.57	2.61
	Std	0.53	0.52	0.54	0.53	0.64	0.67	0.66	0.67
Micro2	Chi2	69.66	106.83	71.49	108.72	21.68	24.82	23.29	25.70
	P-value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.1538	0.0731	0.1061	0.0584
	Max-Min	2.73	3.55	2.76	3.47	1.92	1.83	1.81	1.89
	Std	0.82	1.05	0.84	1.06	0.44	0.48	0.45	0.49

Note : Nous testons l'égalité des indicatrices enseignants au sein de chaque matière. La ligne ''Chi2'' donne la valeur de la statistique du Chi2, la p-value correspondante est juste en-dessous. Nous reportons ensuite quelques statistiques descriptives (max-min et écart-type) de la distribution de ces indicatrices.

TAB. 1.11 – Effets maîtres, cas discret, suite

		L1				L2			
		25/75		36/66		Total Juin		Total Sept	
		02/03	04/05/06	02/03	04/05/06	25/75	33/66	25/75	33/66
Macro2	Chi2	93.46	183.99	91.83	182.63	49.32	52.12	46.24	51.57
	P-value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0011	0.0005	0.0028	0.0006
	Max-Min	3.75	2.82	3.77	2.8	3.53	3.61	3.54	3.53
	Std	0.93	0.91	0.93	0.9	0.77	0.79	0.73	0.77
Stat	Chi2	76.62	94.42	74.06	94.40	44.37	40.09	43.37	40.37
	P-value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0047	0.0150	0.0063	0.0140
	Max-Min	2.46	3.44	3.07	3.41	2.22	2.17	2.22	2.16
	Std	0.75	0.85	0.82	0.85	0.62	0.59	0.61	0.59
Info	Chi2	54.41	173.34	55.73	173.64	80.81	84.32	82.47	84.47
	P-value	0.0004	< 0.0001	0.0002	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	Max-Min	2.57	5.6	2.48	5.63	3.92	4.05	4.53	4.26
	Std	0.71	1.22	0.7	1.22	1.08	1.1	1.13	1.12

Note : Nous testons l'égalité des indicatrices enseignants au sein de chaque matière. La ligne ''Chi2" donne la valeur de la statistique du Chi2, la p-value correspondante est juste en-dessous. Nous reportons ensuite quelques statistiques descriptives (max-min et écart-type) de la distribution de ces indicatrices.

Tab. 1.12 – Effets maîtres, cas continu

		L1		L2	
		02/03	04/05/06	Total Juin	Total Sept
Maths	Chi2	68.27	93.65	63.77	107.44
	P-value	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	Max-Min	3.52	3.86	2.34	2.92
	Std	0.83	0.79	0.68	0.87
Micro1	Chi2	93.42	132.33	26.76	32.43
	P-value	< 0.0001	< 0.0001	0.0442	0.0088
	Max-Min	3.5	2.66	1.39	1.49
	Std	0.99	0.86	0.37	0.45
Macro1	Chi2	41.37	56.74	66.67	73.10
	P-value	0.0008	< 0.0001	< 0.0001	< 0.0001
	Max-Min	1.83	1.86	2.41	2.79
	Std	0.52	0.52	0.63	0.73
Micro2	Chi2	69.91	107.48	23.91	28.49
	P-value	< 0.0001	< 0.0001	0.0914	0.0276
	Max-Min	2.73	3.56	1.94	2.06
	Std	0.81	1.05	0.47	0.52

Lecture : Nous testons l'égalité des indicatrices enseignants au sein de chaque matière. La ligne ''Chi2" donne la valeur de la statistique du Chi2, la p-value correspondante est juste en-dessous. Nous reportons ensuite quelques statistiques descriptives (max-min et écart-type) de la distribution de ces indicatrices.

TAB. 1.13 – Effets maîtres, cas continu, suite

		L1		L2	
		02/03	04/05/06	Total Juin	Total Sept
Macro2	Chi2	94.24	186.11	59.08	61.56
	P-value	< 0.0001	< 0.0001	0.0001	< 0.0001
	Max-Min	3.63	2.84	3.62	3.26
	Std	0.92	0.91	0.83	0.82
Stat	Chi2	74.25	93.15	44.04	86.27
	P-value	< 0.0001	< 0.0001	0.0052	< 0.0001
	Max-Min	2.89	3.44	2.31	3.4
	Std	0.81	0.85	0.63	0.94
Info	Chi2	54.35	172.98	85.39	89.82
	P-value	0.0004	< 0.0001	< 0.0001	< 0.0001
	Max-Min	2.68	5.64	4.02	3.94
	Std	0.71	1.23	1.07	1.1

Lecture : Nous testons l'égalité des indicatrices enseignants au sein de chaque matière. La ligne ''Chi2" donne la valeur de la statistique du Chi2, la p-value correspondante est juste en-dessous. Nous reportons ensuite quelques statistiques descriptives (max-min et écart-type) de la distribution de ces indicatrices.

Tab. 1.14 – Note moyenne des étudiants de faible qualité en 2002/2003, analyse contrefactuelle

	L1		L2			
			Total Juin		Total Sept	
	25/75	33/66	25/75	33/66	25/75	33/66
Observée	7.68	7.99	8.29	8.55	8.56	8.69
Répartition optimale	8.13 (+9.7%)	8.38 (+8.4%)	8.38 (+1.7%)	8.59 (+0.9%)	8.62 (+1.5%)	8.75 (+1.1%)
Répartition par niveau	5.16 (-44.9%)	6.0 (−33.2%)	7.83 (−6.7%)	8.22 (−4.5%)	7.20 (−18.9%)	7.22 (−20.5%)

Note : La répartition optimale consiste à affecter dans chaque groupe exactement la même proportion d'étudiants faibles, moyens et forts. La répartition par niveau, la moins optimale compte tenu des effets de pairs non-linéaires, consiste à ce que chaque groupe ne soit constitué que d'un seul niveau d'étudiant. Entre parenthèses la variation par rapport à la note moyenne observée, qui correspond à l'affectation aléatoire des étudiants dans les groupes.

1.6 Annexe

1.6.1 Interpréter les paramètres d'effets de pairs

Le modèle initial complet est

$$\begin{aligned}
y =& \alpha_L^* \mathbb{1}_L + \alpha_M^* \mathbb{1}_M + \alpha_H^* \mathbb{1}_H + \beta_L^* P_L + \beta_M^* P_M + \beta_H^* P_H \\
&+ \beta_{LL}^* \mathbb{1}_L P_L + \beta_{LM}^* \mathbb{1}_L P_M + \beta_{LH}^* \mathbb{1}_L P_H \\
&+ \beta_{ML}^* \mathbb{1}_M P_L + \beta_{MM}^* \mathbb{1}_M P_M + \beta_{MH}^* \mathbb{1}_M P_H \\
&+ \beta_{HL}^* \mathbb{1}_H P_L + \beta_{HM}^* \mathbb{1}_H P_M + \beta_{HH}^* \mathbb{1}_H P_H
\end{aligned}$$

Mais bien sûr, $\mathbb{1}_M = 1 - \mathbb{1}_L - \mathbb{1}_H$ et $P_M = 100 - P_L - P_H$. Par substitution,

$$\begin{aligned}
y =& \alpha_M^* + 100\beta_M^* + 100\beta_{MM}^* \\
&+ \left(\alpha_L^* - \alpha_M^* + 100\beta_{LM}^* - 100\beta_{MM}^*\right) \mathbb{1}_L + \left(\alpha_H^* - \alpha_M^* + 100\beta_{HM}^* - 100\beta_{MM}^*\right) \mathbb{1}_H \\
&+ \left(\beta_L^* - \beta_M^* + \beta_{ML}^* - \beta_{MM}^*\right) P_L + \left(\beta_H^* - \beta_M^* + \beta_{MH}^* - \beta_{MM}^*\right) P_H \\
&+ \left(\beta_{LL}^* - \beta_{LM}^* + \beta_{MM}^* - \beta_{ML}^*\right) \mathbb{1}_L P_L + \left(\beta_{LH}^* - \beta_{MH}^* + \beta_{MM}^* - \beta_{LM}^*\right) \mathbb{1}_L P_H \\
&+ \left(\beta_{HL}^* - \beta_{HM}^* + \beta_{MM}^* - \beta_{ML}^*\right) \mathbb{1}_H P_L + \left(\beta_{HH}^* - \beta_{HM}^* + \beta_{MM}^* - \beta_{MH}^*\right) \mathbb{1}_H P_H
\end{aligned}$$

Donc,

$$\begin{aligned}
\alpha_L &= \alpha_L^* - \alpha_M^* + 100\beta_{LM}^* - 100\beta_{MM}^* \\
\alpha_H &= \alpha_H^* - \alpha_M^* + 100\beta_{HM}^* - 100\beta_{MM}^* \\
\beta_L &= \beta_L^* - \beta_M^* + \beta_{ML}^* - \beta_{MM}^* \\
\beta_H &= \beta_H^* - \beta_M^* + \beta_{MH}^* - \beta_{MM}^* \\
\beta_{LL} &= \beta_{LL}^* - \beta_{LM}^* + \beta_{MM}^* - \beta_{ML}^* \\
\beta_{LH} &= \beta_{LH}^* - \beta_{MH}^* + \beta_{MM}^* - \beta_{LM}^* \\
\beta_{HL} &= \beta_{HL}^* - \beta_{HM}^* + \beta_{MM}^* - \beta_{ML}^* \\
\beta_{HH} &= \beta_{HH}^* - \beta_{HM}^* + \beta_{MM}^* - \beta_{MH}^*
\end{aligned}$$

Maintenant considérons un étudiant de qualité faible dans un certain groupe. Sa note vaut

$$\alpha_L^* + (\beta_L^* + \beta_{LL}^*)\, P_L + (\beta_M^* + \beta_{LM}^*)\, P_M + (\beta_H^* + \beta_{LH}^*)\, P_H$$

Si un point de pourcentage d'étudiants faibles est remplacé par un point de forts dans ce groupe ($P_L \to P_L - 1$ et $P_H \to P_H + 1$) alors la différence entre la nouvelle note et l'ancienne note de cet étudiant faible vaut

$$\beta_H^* + \beta_{LH}^* - \beta_L^* - \beta_{LL}^* = \beta_H - \beta_L + \beta_{LH} - \beta_{LL}$$

Nous calculons de la même manière tous les effets de la table 1.1.

1.6.2 Redistribution des groupes de L1 en L2

TAB. 1.15 – Redistribution des groupes de L1 en L2, 2003

Gr L1	Gr L2	freq	Gr L1	Gr L2	freq	Gr L1	Gr L2	freq
1	7	1	2	15	3	4	10	1
1	8	7	2	16	3	4	12	1
1	9	1	2	17	1	4	15	1
1	11	1	2	19	1	4	16	1
1	12	3	2	22	1	4	18	1
1	15	4	2	23	1	4	20	3
1	16	3	2	25	1	4	21	1
1	18	1	3	2	2	4	23	1
1	21	2	3	3	3	4	24	2
1	22	3	3	4	3	4	25	1
1	23	1	3	11	2	4	26	2
1	24	3	3	14	1	5	2	2
1	26	1	3	15	2	5	3	3
2	2	2	3	17	1	5	6	2
2	3	1	3	18	1	5	8	1
2	4	2	3	19	1	5	9	2
2	5	2	3	20	5	5	10	1
2	6	2	3	21	2	5	11	2
2	7	1	3	23	4	5	14	1
2	8	1	4	1	2	5	17	2
2	9	1	4	2	2	5	19	2
2	11	1	4	3	4	5	21	2
2	13	1	4	5	2	5	23	1
2	14	2	4	8	1	5	24	2

Lecture : Sur la première ligne du premier bloc colonne, nous lisons que seulement un étudiant était dans le groupe 1 en L1 en 2002 et dans le groupe 7 en L2 en 2003. Pour gagner de la place, nous ne présentons pas l'ensemble de la table de contingence.

Tab. 1.16 – Redistribution des groupes de L1 en L2, 2004

Gr L1	Gr L2	freq	Gr L1	Gr L2	freq	Gr L1	Gr L2	freq
1	2	1	2	19	1	4	16	4
1	3	1	2	20	3	4	17	2
1	4	1	2	22	2	4	21	1
1	5	1	2	23	2	5	1	1
1	6	3	3	1	1	5	5	1
1	8	1	3	2	2	5	6	5
1	10	1	3	3	3	5	7	2
1	15	3	3	4	1	5	8	1
1	18	3	3	5	3	5	15	3
1	19	3	3	6	2	5	18	1
1	20	1	3	8	2	5	19	4
1	22	1	3	12	2	5	20	1
1	23	3	3	16	1	5	22	4
1	24	1	3	18	1	5	23	1
2	2	1	3	22	1	6	2	2
2	7	4	3	23	1	6	4	3
2	8	1	3	24	3	6	6	3
2	9	2	4	1	5	6	8	1
2	10	1	4	4	2	6	10	2
2	12	2	4	5	2	6	12	2
2	14	1	4	9	1	6	14	1
2	15	2	4	11	1	6	15	1
2	17	2	4	13	6	6	16	3
2	18	1	4	15	1	6	17	1

Lecture : Sur la première ligne du premier bloc colonne, nous lisons que seulement un étudiant était dans le groupe 1 en L1 en 2003 et dans le groupe 2 en L2 en 2004. Pour gagner de la place, nous ne présentons pas l'ensemble de la table de contingence.

Chapitre 2

L'impact de l'Imperfection d'Information sur les Choix Educatifs : une Approche Structurelle

2.1 Introduction

L'objectif de ce chapitre est d'estimer un modèle d'investissement en capital humain avec information imparfaite et apprentissage bayésien. L'idée générale des modèles structurels dynamiques de choix d'éducation est qu'un individu décide de faire une année d'étude supplémentaire si le coût qu'elle engendre (frais d'inscription et renoncement à un salaire) est compensé par des salaires futurs plus élevés. Depuis une dizaine d'années ces modèles sont de plus en plus mobilisés par les économistes, principalement grâce aux capacités de stockage croissantes des ordinateurs, pour mieux comprendre comment les individus font leur choix d'éducation. Les chercheurs ont ainsi traité, en équilibre partiel, des questions de politiques économiques relatives aux frais d'inscription (Keane et Wolpin, 1997), au travail étudiant (Eckstein et Wolpin, 1999), aux contraintes de crédit (Keane et Wolpin, 2001 et

Cameron et Taber, 2004), à l'augmentation de la scolarisation (Magnac et Thesmar, 2002), ou encore aux différentiels d'éducation entre races (Keane et Wolpin, 2000). Toujours en équilibre partiel, d'autres travaux plus économétriques se sont intéressés par exemple à l'estimation des rendements de l'éducation (Belzil et Hansen, 2002) ou à l'étude du modèle de régression de salaire à coefficients aléatoires corrélés (Belzil et Hansen, 2007). D'autres chercheurs, comme Heckman, Lochner et Taber (1998) ou Lee et Wolpin (2006) ont généralisé l'approche à un environnement en équilibre général.

Au-delà de l'approche par un modèle d'investissement en capital humain, ces travaux partagent tous une caractéristique commune fondamentale : les individus sont supposés connaître parfaitement leurs caractéristiques sur le marché du travail, même s'ils n'ont pas encore fait le choix de quitter l'école. Ceci revient donc à supposer qu'un individu qui n'a jamais travaillé connaît en particulier ses rendements de l'éducation, de l'expérience et son aptitude sur le marché du travail. Cette hypothèse nous semble trop forte : comment en effet savoir si je suis doué dans mon travail si je ne l'ai jamais exercé ? Deux séries de travaux ont déjà travaillé sur cette question. Partant de cette idée, Cunha, Heckman et Navarro, dans une série d'articles synthétisés dans Cunha et Heckman (2007), utilisent des modèles à facteurs pour modéliser de manière conjointe les salaires et les coûts de l'éducation, et proposent de tester l'ensemble d'information des individus au moment de faire leur choix en recherchant l'ensemble d'information qui permettra le mieux de reproduire les choix d'éducation observés. Cette méthode, qui a été un grand pas dans la prise en compte des problèmes d'information, a pour inconvénient de proposer une vision binaire de l'information (connaître ou ne pas connaître) et ne traite pas la question qui découle du manque d'information, à savoir l'apprentissage. Dans le cadre d'un modèle structurel dynamique de choix d'éducation, Belzil (2007) s'écarte de la vision binaire de la connaissance en supposant que les individus forment des anticipations subjectives concernant leurs caractéristiques sur le marché du travail, mais cet article n'explique pas comment ces anticipations sont formées et n'aborde pas non plus la question de l'apprentissage.

Les deux innovations que nous apportons sont d'une part de supposer que les individus utilisent une information qu'ils observent, leur goût et leur aptitude à l'école, pour former des anticipations de leurs caractéristiques sur le marché du travail ; à l'inverse de Belzil (2007), ces anticipations restent donc objectives ; d'autre part nous supposons qu'une fois arrivés sur le marché du travail, les individus révisent leur croyance en observant leur productivité grâce au salaire. Le premier enseignement de ce travail est que l'information parfaite est rejetée par les données. Il apparaît donc nécessaire de prendre en compte cette dimension dans la modélisation des choix d'éducation. Le deuxième enseignement est que ces problèmes d'information sont tels que 20% des individus font des choix différents de ceux qu'ils feraient en information parfaite. Ces choix ne diffèrent cependant que d'une ou deux années d'études.

Dans la section suivante, nous présentons le modèle, nous discutons son identification dans la section 2.3, nous détaillons la méthode d'estimation dans la section 2.4, présentons les données dans la section 2.5, les résultats dans la section 2.6 et nous concluons dans la section 2.7.

2.2 Un modèle d'investissement en capital humain avec information imparfaite et apprentissage bayésien

2.2.1 Structure générale

L'éventail des choix

Le modèle que nous utilisons est un modèle structurel dynamique de choix d'éducation dans la lignée de celui développé par Keane et Wolpin (1997). Nous supposons que chaque année à partir de l'âge de 16 ans les individus doivent choisir entre trois alternatives disjointes[1]. Ces trois choix exclusifs sont de faire une année d'étude supplémentaire, aller sur le marché du travail, ou s'investir dans la production domestique, c'est-à-dire rester à la maison. Aucun état n'est absorbant, c'est-à-dire que toutes les options sont possibles à chaque choix. En particulier, l'éducation n'est pas un état absorbant, comme dans les modèles d'arrêt optimal tels que Belzil et Hansen (2002) ou Magnac et Thesmar (2002). On peut noter que notre éventail de choix est plus large que les deux papiers précédents qui ne prennent en compte ni la production domestique ni la possibilité du retour à l'école, mais moins riche que Keane et Wolpin (1997), qui autorisent les individus à choisir leur occupation (col bleu, col blanc, secteur militaire). On peut enfin noter que, à l'inverse de Eckstein et Wolpin (1999), nous ne prenons pas en compte le travail étudiant. Techniquement, restreindre le nombre de choix à 3 permet de n'avoir à intégrer que des lois normales bivariées, ce qui évite d'utiliser des techniques de simulations coûteuses en temps et en précision.

La modélisation d'une prise de décision risquée

Une des hypothèses principales du modèle est que les agents prennent en compte l'avenir dans leur décision présente, ils sont prévisionnels. Les individus ne connaissant pas avec certitude certaines variables futures, ils sont supposés prendre des décisions risquées. Pour modéliser ce risque, la première théorie à laquelle on peut penser est celle de l'Utilité Espérée (von Neumann et Morgenstern (1947)), dans laquelle les agents maximisent l'espérance des utilités associés à chaque résultat. Cette théorie, par son pouvoir explicatif et sa simplicité, s'est imposée pendant plusieurs décennies. Mais à partir du début des années 80, des travaux sont venus pointer les insuffisances de ce modèle (Kahneman et Tversky (1979)), en travaillant par exemple sur le paradoxe d'Allais ou sur la déformation subjective des lois

[1] L'âge de 16 ans est choisi car il correspond à l'âge minimum légal à partir duquel les individus peuvent quitter le système scolaire aux Etats-Unis.

de probabilité des résultats. D'autres travaux encore plus généraux ont supposé que non seulement les résultats n'étaient pas connus avec certitude mais que les lois de probabilités elles-même non plus, ce qui a donné lieu à des modèles de décision dans l'incertain. Dans ce chapitre, nous suivrons l'ensemble de la littérature sur les modéles structurels dynamiques de choix d'éducation et nous nous placerons dans le cadre de la théorie de l'Utilité Espérée. Ce choix est essentiellement dicté par les simplifications calculatoires permises par la linéarité imposée par cette théorie. Nous sommes cependant conscients des limites de ce modèle et considérons ce point comme une piste de recherche future fondamentale.

La modélisation des anticipations

La question de savoir ce que connaissent les agents au moment de faire leur choix et la manière dont ils anticipent ce qu'ils ne connaissent pas est une question cruciale. Dans la quasi totalité de la littérature sur les modèles structurels dynamiques de choix d'éducation, l'hypothèse d'anticipation rationnelle est faite. Cette hypothèse revient à dire que les agents forment des anticipations objectivement correctes compte tenu de leur ensemble d'information. Pour Keane et Wolpin (1997) par exemple l'hypothèse d'anticipation rationnelle se traduit par le fait que les agents connaissent tous les paramètres de l'équation de salaire, y compris les rendements de l'éducation, leur aptitude sur le marché du travail et la distribution des chocs de productivité. Pour Manski (2004) cette hypothèse n'est pas raisonnable. Comment en effet des individus non-spécialistes en microéconométrie appliquée pourraient savoir ce que les spécialistes eux-mêmes ont toutes les peines du monde à trancher ? Alors pourquoi cette hypothèse est-elle si largement utilisée dans les modèles de décision ? C'est, pour Manski, parce qu'elle permet, de manière élégante, de boucler le modèle. Il s'agit donc pour lui de remettre en cause cette hypothèse et de tenter de s'en écarter. La solution qu'il préconise est tout simplement d'enquêter les individus sur la manière dont ils forment leurs anticipations, par exemple en leur faisant choisir des probabilités subjectives d'occurence de certains événements. Ces données peuvent ensuite être utilisées pour modéliser les anticipations des individus. Dans la littérature structurelle dynamique, le seul travail à avoir traité la question des anticipations non rationnelles est Buchinsky et Lesly (2007). Sans essayer de déterminer quels types d'anticipations sont utilisés par les agents, ils montrent à l'aide d'un modèle calibré que les choix d'éducation sont sensibles au type d'anticipation utilisé. Ils supposent que les paramètres de la future équation de salaire ne sont pas connus des individus. Pour les prévoir les agents sont supposés estimer des régressions de quantiles sur les salaires des années précédentes, puis utiliser un modèle VAR sur les coefficients de ces régressions pour inférer les coefficients inconnus. Ainsi, dans Buchinsky et Leslie (2007), les anticipations ne sont pas rationnelles dans le sens où les individus ne connaissent pas les paramètres des équations de salaire futures, mais ils sont supposés disposer de plusieurs décennies de CPS, être capables d'estimer des régressions de quantiles et des modèles VAR avec un sampler de Gibbs. Pour poursuivre l'idée de Manski, on voit bien que pour lever l'hypothèse d'anticipation rationnelle, les deux auteurs sont amenés à faire des hypothèses tout aussi difficiles à défendre.

Mais faut-il rejeter entièrement l'hypothèse d'anticipation rationnelle ? Dans le domaine, par exemple, des choix structurels dynamiques de départ à la retraite, Benitez-Silva et Dwyer (2005) ont montré que l'évolution des anticipations était compatible avec l'hypothèse d'anticipation rationnelle. En conclure que c'est également le cas dans le domaine de l'éducation serait bien sûr plus qu'hasardeux. Cependant, des travaux comme Dominitz et Manski (1996) ont montré, en interrogeant les étudiants sur leurs anticipations, qu'il existe une croyance commune que les rendements de l'université en termes de salaires sont strictement positifs et que le salaire aura augmenté entre 30 et 40 ans. Ces anticipations sont cohérentes avec les implications de base d'une équation de Mincer et on peut ainsi se demander si les anticipations rationnelles ne fournissent pas une approximation raisonnable des croyances subjectives. Au-delà de cet argument, les données sur les croyances subjectives sont très rares et c'est au fond la raison majeure pour laquelle la littérature structurelle sur les choix dynamiques d'éducation s'est accomodée de l'hypothèse d'anticipation rationnelle. C'est également la raison pour laquelle nous nous placerons sous cette hypothèse. Conscients de cette limitation, nous considérons également ce point comme une piste de recherche future fondamentale.

Le cadre formel

A chaque date, les agents sont supposés maximiser sur leur cycle de vie (jusqu'à l'âge A, fin de la vie active) l'espérance de la somme actualisée de leurs utilités instantanées. Notons U_{ia}^S, U_{ia}^H, U_{ia}^W les utilités instantanées de l'individu i à l'âge a respectivement pour l'école, la maison et le travail. Notons également d_{ia}^S, d_{ia}^H, d_{ia}^W les indicatrices de choix, c'est-à-dire les variables qui valent 1 si l'individu i choisit à l'âge a d'aller respectivement à l'école, à la maison ou sur le marché du travail, et 0 sinon. En notant Ω_{ia} l'ensemble d'information de l'individu i à l'âge a et β le taux d'escompte, le programme de l'agent i à l'âge a est de choisir entre d_{ia}^S, d_{ia}^H et d_{ia}^W de sorte à maximiser

$$E\left\{\sum_{\tau=a}^{A}\beta^{\tau-a}\sum_{m=S,H,W}U_{i\tau}^m d_{i\tau}^m|\Omega_{ia}\right\} \tag{2.1}$$

Le maximum de cette fonction, par définition, est la fonction valeur $V_{ia}\left(\Omega_{ia}\right)$.

Les équations de Bellman correspondant à ce programme (voir Bellman (1957) ou Stockey et Lucas(1989) pour des textes de référence sur la programmation dynamique) sont

$$V_{ia}^S\left(\Omega_{ia}\right) = E\left(U_{ia}^S|\Omega_{ia}\right) + \beta E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia}, d_{ia}^S = 1\right] \tag{2.2}$$

$$V_{ia}^H\left(\Omega_{ia}\right) = E\left(U_{ia}^H|\Omega_{ia}\right) + \beta E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia}, d_{ia}^H = 1\right] \tag{2.3}$$

$$V_{ia}^W\left(\Omega_{ia}\right) = E\left(U_{ia}^W|\Omega_{ia}\right) + \beta E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia}, d_{ia}^W = 1\right] \tag{2.4}$$

avec

$$V_{ia}\left(\Omega_{ia}\right) = max\left[V_{ia}^{S}\left(\Omega_{ia}\right), V_{ia}^{H}\left(\Omega_{ia}\right), V_{ia}^{W}\left(\Omega_{ia}\right)\right] \tag{2.5}$$

$V_{ia}^{S}\left(\Omega_{ia}\right)$, $V_{ia}^{H}\left(\Omega_{ia}\right)$ et $V_{ia}^{W}\left(\Omega_{ia}\right)$ sont les valeurs qu'obtiendront les individus s'ils font respectivement le choix de l'école, de la maison et du travail. La valeur de l'école, par exemple, est composée d'une part de l'utilité que lui procurera à l'âge a ce choix $\left(E\left(U_{ia}^{S}|\Omega_{ia}\right)\right)$, et d'autre part de la valeur présente de l'espérance du maximum de ce que pourra obtenir l'individu par la suite, c'est-à-dire $E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia}, d_{ia}^{S}=1\right]$. Les valeurs des deux autres choix s'interprètent de la même façon. La valeur à l'âge a $V_{ia}\left(\Omega_{ia}\right)$ vaut alors le maximum des valeurs correspondant à chaque choix, $V_{ia}^{S}\left(\Omega_{ia}\right)$, $V_{ia}^{H}\left(\Omega_{ia}\right)$ et $V_{ia}^{W}\left(\Omega_{ia}\right)$. Ces équations de Bellman sont fondamentales car elles permettent de résoudre le programme de l'agent par récurrence arrière. Il est clair que connaissant l'espérance de la valeur future, les valeurs de chaque choix sont déterminées et les choix peuvent être tranchés. Ainsi, un individu à l'âge a se projettera à l'âge terminal A, calculera les espérances de fonction valeur pour tous les états atteignables, et calculera de proche en proche par récurrence arrière les espérances de fonction valeur qui lui permettront in fine de faire son choix.

2.2.2 Spécifications des utilités instantanées

Notons Educ_{ia} et Exp_{ia} respectivement le nombre d'années d'études et le nombre d'années d'expérience accumulées par l'individu i au moment où il doit faire son choix, c'est-à-dire en début de période a.

A l'âge a, l'utilité instantanée de l'éducation U_{ia}^{S} est définie par

$$\begin{aligned} U_{ia}^{S} &= \mathbb{1}_{(\text{Educ}_{ia}\geq 12)}\delta_C + \mathbb{1}_{(\text{Educ}_{ia}\geq 16)}\delta_M + DQ_{ia}\delta_{DQ} + \nu_i^S + \epsilon_{ia}^S && (2.6)\\ &= Z_{ia}\delta_Z + \nu_i^S + \epsilon_{ia}^S && (2.7)\end{aligned}$$

où $Z_{ia} = \left[\mathbb{1}_{(\text{Educ}_{ia}\geq 12)}\ \mathbb{1}_{(\text{Educ}_{ia}\geq 16)}\ DQ_{ia}\right]$ et $\delta_Z = \left[\delta_C\ \delta_M\ \delta_{DQ}\right]'$. $\mathbb{1}_{(\text{Educ}_{ia}\geq 12)}$ et $\mathbb{1}_{(\text{Educ}_{ia}\geq 16)}$

sont des variables indicatrices qui prennent la valeur 1 lorsqu' en début de période a l'individu i a déjà accumulé respectivement au moins 12 et 16 années d'études. Ce sont des variables qui représentent les coûts objectifs, comme les frais d'inscription, et subjectifs associés aux études à l'université. DQ_{ia} est une variable indicatrice qui prend la valeur 1 lorsqu'en début de période a l'individu i a déjà quitté l'école au moins une fois dans le passé. Cette variable prend en compte de potentiels effets psychologiques qui rendent plus difficile le choix de l'école après l'avoir déjà quittée. δ_C, δ_M et δ_{DQ} sont des paramètres à estimer. ν_i^S est une composante d'hétérogénéité inobservée spécifique à chaque individu, constante dans le temps et indépendante des explicatives observées. Elle s'interprète comme le goût et l'aptitude de l'individu i pour les études et prend en compte les phénomènes de sélection déjà pointés par Willis et Rosen (1979). Enfin, ϵ_{ia}^S est un choc indépendant des autres variables de l'utilité de l'école et indépendant dans le temps. On suppose que $\epsilon_{ia}^S \sim N(0, \sigma_S)$.

L'utilité instantanée de la production domestique est définie par

$$U_{ia}^H = \nu_i^H + \epsilon_{ia}^H \tag{2.8}$$

où ν_i^H est une composante d'hétérogénéité inobservée spécifique à chaque individu et fixe dans le temps qui prend en compte le goût pour le loisir. ϵ_{ia}^H est un choc indépendant de l'effet individuel et indépendant dans le temps. On suppose que $\epsilon_{ia}^H \sim N(0, \sigma_H)$.

Enfin l'utilité instantanée du travail est définie par

$$\begin{aligned} U_{ia}^W = ln\,(w_{ia}) &= \text{Educ}_{ia}\theta_{1i} + \text{Exp}_{ia}\theta_{2i} + \text{Exp}_{ia}^2\theta_{3i} + \nu_i^W + \epsilon_{ia}^W & (2.9) \\ &= X_{ia}\theta_i + \epsilon_{ia}^W & (2.10) \end{aligned}$$

où w_{ia} est le salaire que l'individu i va recevoir à l'âge a s'il décide d'aller sur le marché du travail, $X_{ia} = \left[\text{Educ}_{ia}\ \text{Exp}_{ia}\ \text{Exp}_{ia}^2\ 1\right]$ et $\theta_i = \left[\theta_{1i}\ \theta_{2i}\ \theta_{3i}\ \nu_i^W\right]'$. Ainsi l'utilité du travail est assimilée au logarithme du salaire, lui même modélisé par une équation de Mincer standard. θ_{1i}, θ_{2i} et θ_{3i} sont respectivement les rendements de l'éducation, de l'expérience et de l'expérience au carré. ν_i^W, la constante de cette équation de salaire, représente l'aptitude de l'individu i sur le marché du travail. Il faut noter qu'à l'inverse de la littérature sur les modèles structurels dynamiques de choix d'éducation (sauf Belzil et Hansen (2007)), nous n'autorisons pas uniquement la constante de l'équation de salaire à être spécifique à chaque individu, mais tous ses coefficients. ϵ_{ia}^W est un choc de productivité indépendant dans le temps et des autres variables de l'équation de salaire et nous supposons que $\epsilon_{ia}^W \sim N(0, \sigma_W)$. Nous supposons en outre que les trois chocs ϵ_{ia}^S, ϵ_{ia}^H, ϵ_{ia}^W sont indépendants entre eux.

Pour finir, nous supposons qu'il existe deux types d'individus sur le marché du travail, de sorte que pour tout individu i $\theta_i \in \{\theta_1, \theta_2\}$. Il existe par ailleurs K types d'individus vis-à-vis de leur goût à l'école et pour le loisir : $\left(\nu_i^S, \nu_i^H\right)' \in \left\{\left(\nu_k^S, \nu_k^H\right)', k = 1...K\right\}$. Nous notons TypeSH$_i$ ce type de sorte que le type de l'individu i à l'école et à la maison TypeSH$_i$ vaut k si et seulement si $\left(\nu_i^S, \nu_i^H\right)' = \left(\nu_k^S, \nu_k^H\right)'$. Il existe donc en tout 2*K types dans la population, cardinal du produit cartésien de l'ensemble des θ par l'ensemble des ν.

2.2.3 Déroulement du temps et ensemble d'information

Nous avons supposé que les anticipations sont rationnelles, c'est-à-dire que les anticipations sont objectivement correctes compte tenu de l'ensemble d'information. Il reste donc à déterminer l'information dont disposent les individus au moment de faire leur choix. Usuellement, et Keane et Wolpin (1997) en sont l'exemple typique, les agents sont supposés connaître la valeur de toutes les variables présentes, ils ignorent uniquement les réalisations des chocs futurs. En particulier, ils sont donc supposés connaître leurs rendements de l'éducation et leur aptitude sur le marché du travail. Mais cette hypothèse paraît extrêmement forte. Supposons en effet que les rendements de l'éducation et l'aptitude sur le marché du travail soient hétérogènes. Comment un individu qui n'a jamais fait l'expérience du marché du travail pourrait-il connaître ses propres caractéristiques sur ce marché ? Cette question a été soulevée récemment d'une part par trois auteurs, Cunha, Heckman et Navarro (CHN par la suite) dans une série de travaux synthétisés dans Cunha et Heckman (2007) et d'autre part par Belzil (2007).

Le but explicite de CHN est précisément de déterminer l'ensemble d'information des agents, c'est-à-dire les variables qu'ils connaissent au moment de faire leur choix *et* qu'ils utilisent pour se déterminer. Décrivons brièvement leur méthode. Supposons que les agents doivent choisir entre deux alternatives, arrêter les études au bac et faire des études supérieures. Notons $Y_{0,i}$ la somme actualisée des revenus futurs associés au lycée, $Y_{1,i}$ celle associée à l'enseignement supérieur et C_i les coûts associés à l'enseignement supérieur. Notons Ω_i l'ensemble d'information de l'individu au moment de faire son choix. Etant incertains sur la valeur exacte des revenus futurs et éventuellement des coûts, ils forment des anticipations et décident d'aller à l'université si et seulement si $E\left(Y_{1,i} - C_i - Y_{0,i}|\Omega_i\right) \geq 0$. Ainsi, $E\left(Y_{1,i} - C_i - Y_{0,i}|\Omega_i\right)$ est la seule quantité utilisée par les agents pour faire leur choix. Si l'ensemble d'information est bien spécifié, la partie V_{Ω_i} non prévisible de $Y_{1,i} - C_i - Y_{0,i}$, $Y_{1,i} - C_i - Y_{0,i} - E\left(Y_{1,i} - C_i - Y_{0,i}|\Omega_i\right)$, n'est pas utilisée par les agents et ne doit donc pas être corrélée au choix des individus. Cette idée est la base de la méthode de CHN. Pour tester la correcte spécification d'un ensemble d'information $\widetilde{\Omega}_i$, CHN calculent la quantité $V_{\widetilde{\Omega}_i}$, et si elle est corrélée aux choix, c'est que cet ensemble d'information n'est pas celui utilisé par les agents. La principale difficulté pour implémenter cette idée est que l'économètre n'observe jamais à la fois $Y_{0,i}$ et $Y_{1,i}$ car ces deux quantités sont dans une relation de factuel/contrefactuel. Pour estimer les contrefactuels, CHN utilisent un modèle à facteurs mis au point par Carneiro, Hansen et Heckman (2003). Ils estiment alors que les agents ne connaissent pas les chocs présents sur le marché du travail et qu'ils connaissent le sous-ensemble des facteurs lié à l'aptitude à l'école. Ils calculent également que si les individus disposaient de leur information ex-post, 13.81% des lycéens auraient poursuivi à l'univsersité et 17.15% des diplômés de l'enseignement supérieur auraient arrêté leurs études au lycée. Cette série de papiers a été un grand pas en avant dans la reconnaissance des problèmes d'information dans les choix d'éducation. Cependant, on peut regretter d'une part, comme les auteurs, que leur méthode ne permet pas d'identifier l'ensemble d'information au sens strict mais la partie de cet ensemble utilisée par l'agent pour faire son choix, et d'autre part que la connaissance est une notion binaire, c'est-à-dire que les agents sont supposés connaître ou ne pas connaître certaines variables.

Dans le cadre d'un modèle structurel dynamique de choix d'éducation, belzil (2007) prend également en compte ces problèmes d'information. Dans son modèle, l'imperfection d'information vient de deux sources : les chocs de productivité, présents et futurs, et sa propre aptitude sur le marché du travail. A côté des distributions objectives de ces variables, les individus forment des anticipations subjectives que la structure du modèle permet d'identifier. En particulier, chaque type d'individu forme une anticipation sur sa probabilité d'appartenance à chaque type. Cette caractéristique permet de s'écarter de la vision binaire de la connaissance de CHN. Cependant, dans ce modèle, et comme le reconnaît l'auteur, ces anticipations sont exogènes, ie ne sont pas expliquées par le modèle.

En ce qui concerne notre modèle, nous reprenons l'idée avancée par CHN que les individus ne connaissent pas nécessairement l'ensemble des paramètres liés à leur future productvité, c'est-à-dire leur propre rendement de l'éducation, de l'expérience, leur aptitude sur le marché du travail et les chocs de productivité contemporains. Nous reprenons également l'idée de Belzil (2007) que la connaissance n'est pas un phénomène binaire mais plutôt que les individus forment des anticipations probabilistes sur les quantités qu'ils ne connaissent pas. Ainsi, dans notre modèle, les individus connaissent avec certitude leur aptitude et leur goût à l'école et pour la production domestique ν_i^S et ν_i^H. Autrement dit, ils connaissent leur type à l'école et à la maison. A chaque début de période leur sont révélés les chocs liés à l'école et à la production domestique, ϵ_{ia}^S et ϵ_{ia}^H. Ils connaissent également les paramètres de l'équation de salaire et de l'équation liée à l'école : θ_1, θ_2, σ_W, δ_C, δ_M, δ_{DQ}, σ_S et σ_H. Ainsi, les utilités instantanées de l'école et du loisir sont parfaitement connues des agents au moment de faire leur choix. Par contre, comme chez Belzil (2007), les agents ne connaissent pas avec précision leur type sur le marché du travail, θ_i. Autrement dit l'individu i ne sait pas si θ_i vaut θ_1 ou θ_2. Il ne peut que former une anticipation sur sa probabilité d'être de type 1 et de type 2. Cette hypothèse est cohérente avec l'idée qu'un individu qui n'a jamais été sur le marché du travail ne peut pas connaître parfaitement son type sur ce marché. Ainsi, tant que l'individu n'est pas allé sur le marché du travail, nous supposons qu'il dispose d'une probabilité *a priori* π_{ia} qu'il soit de type 1 sur le marché du travail, c'est-à-dire que θ_i vaut θ_1. D'autre part, le salaire lui-même n'est observé qu'en fin de période. Par conséquent, l'agent ne peut pas non plus observer son choc de productivité ϵ_{ia}^W en début de période. Il ne le connaîtra d'ailleurs pas parfaitement non plus en fin de période puisqu'il ne pourra pas distinguer avec certitude dans son salaire son aptitude du choc de productivité. L'hypothèse que le salaire n'est observé qu'en fin de période paraît raisonnable à plusieurs égards. Elle peut tout d'abord être le reflet des bonus ou primes de fin d'année non prévisibles en début de période. Ces primes peuvent être liées à la productivité du salarié lui-même, qu'il ne connaît pas au moment de faire le choix de quitter l'école pour le travail, ou à la bonne conjoncture dont peut bénéficier une entreprise une année donnée, et donc non prévisible par l'individu en début de période. Cette hypothèse est également cohérente avec l'idée qu'un individu, pour faire son choix, ne va pas jusqu'à engager une coûteuse recherche d'emploi qui lui donnerait une offre ferme de salaire. Au contraire, l'agent forme son anticipation de salaire sans savoir exactement dans quelle entreprise il sera embauché, et il ne connaît donc pas avec précision son salaire. Enfin, cette hypothèse est cohérente avec un des résultats de CHN qui estiment que les individus ne connaissent pas les chocs de productivité au moment du choix. Pour résumer, l'ensemble d'information de l'individu i au début de l'âge a s'écrit

$$\Omega_{ia} = \left\{ Educ_{ia},\, Exp_{ia},\, \nu_i^S,\, \nu_i^H,\, \epsilon_{ia}^S,\, \epsilon_{ia}^H,\, \theta_1,\, \theta_2,\, \pi_{ia},\, \sigma_W,\, \delta_C,\, \delta_M,\, \delta_{DQ},\, \sigma_S,\, \sigma_H \right\} \tag{2.11}$$

Essentiellement, le déroulement du temps et l'ensemble d'information des individus sont les mêmes que ceux utilisés par Belzil (2007). Comme on va le voir dans la section suivante, les deux innovations de ce chapitre résident d'une part dans la modélisation initiale de la probabilité π_{ia} qui permet aux individus de prévoir leur type sur le marché du travail et d'autre part la prise en compte de l'apprentissage vis-à-vis de cette probabilité lors des passages de l'agent sur le marché du travail.

2.2.4 Apprentissage bayésien

La première innovation de ce chapitre est de dépasser le caractère exogène des anticipations du modèle de Belzil (2007). Pour ce faire nous partons du principe que l'information qu'utilise l'individu pour connaître son type sur le marché du travail est son aptitude à l'école. On peut en effet aisément imaginer qu'un étudiant qui se sait bon à l'école fera l'anticipation qu'il sera bon sur le marché du travail. Nous supposons ainsi que les individus connaissent la corrélation *objective* qui existe entre les types à l'école et les types sur le marché du travail. Observant son type à l'école, il peut inférer son type sur le marché du travail. Formellement, *tant que l'agent i n'est pas allé sur le marché du travail*, il forme à l'aide de paramètres ξ qu'il connaît un *a priori* initial sur sa probabilité d'être de type 1 sur le marché du travail défini par le logit

$$\pi_{i0} = P\left(\theta_i = \theta_1 / \mathrm{TypeSH}_i\right) = \frac{exp\left(\xi_1 + \xi_2 \mathbb{1}_{\{\mathrm{TypeSH}_i=2\}} + ... + \xi_K \mathbb{1}_{\{\mathrm{TypeSH}_i=K\}}\right)}{1 + exp\left(\xi_1 + \xi_2 \mathbb{1}_{\{\mathrm{TypeSH}_i=2\}} + ... + \xi_K \mathbb{1}_{\{\mathrm{TypeSH}_i=K\}}\right)} \tag{2.12}$$

Mais dès lors que les agents sont supposés évoluer en information imparfaite, il faut faire le pas suivant d'expliquer comment ils apprennent, autrement dit comment ils révisent cet *a priori initial* π_{i0}. C'est la deuxième innovation de ce chapitre. En économie du travail il existe une importante littérature sur les phénomènes d'apprentissage. Jovanovic (1979) fut le premier papier à introduire l'idée d'un apprentissage sur le marché du travail. Dans son modèle de prospection d'emploi, ni le salarié, ni l'entreprise, ne connaissent la valeur de l'appariement. C'est l'observation de la productivité en fin de période qui permet d'actualiser les croyances sur la valeur de cet appariement. Tant que la valeur de l'appariement reste suffisamment élevée, l'individu reste dans l'entreprise, si elle devient trop faible il accepte d'autres offres. Depuis ce travail fondateur, l'apprentissage a été introduit comme étant un facteur potentiellement important dans les théories visant à expliquer la dynamique des salaires ou l'évolution des carrières au sein des entreprises (voir Gibbons en Waldman (2006) pour un exemple récent). Dans ces travaux l'aptitude du salarié n'est pas observée en début de période, le salarié pouvant avoir éventuellement plus d'informations que l'employeur dans le cas d'information asymétrique. Mais l'idée clé, comme pour Jovanovic (1979), reste que c'est l'observation de la productivité en fin de période qui permet d'actualiser les croyances sur l'aptitude du salarié. Cette actualisation est importante dans leurs modèles car ce n'est qu'au moment où l'aptitude du salarié sera évaluée comme suffisament élevée que ce dernier se verra offrir une promotion, c'est-à-dire un poste qui demande de plus grandes compétences. Cette promotion sera bien sûr assortie d'une augmentation de salaire conséquente.

Ne prenant pas en compte les employeurs dans notre modèle, nous nous intéresserons uniquement à l'apprentissage du côté salarié. Nous reprenons l'idée clé que c'est l'observation de la productivité, donc du salaire, en fin de période, qui permet à l'individu d'actualiser son type sur le marché du travail. Plus précisément, supposons qu'au début de l'âge a l'agent i décide, sur la base de π_{ia}, d'aller sur le marché du travail. A la fin de l'âge a, observant son salaire w_{ia}, il actualise de manière bayésienne sa probabilité π_{ia+1} en calculant

$$\pi_{ia+1} = P\left(\theta_i = \theta_1 / w_{ia}\right) = \frac{P\left(w_{ia}/\theta_i = \theta_1\right)\pi_{ia}}{\pi_{ia} P\left(w_{ia}/\theta_i = \theta_1\right) + \left(1 - \pi_{ia}\right) P\left(w_{ia}/\theta_i = \theta_2\right)}$$

$$= \frac{\pi_{ia} \frac{1}{\sigma_W} \varphi\left(\frac{w_{ia} - X_{ia}\theta_1}{\sigma_W}\right)}{\pi_{ia} \frac{1}{\sigma_W} \varphi\left(\frac{w_{ia} - X_{ia}\theta_1}{\sigma_W}\right) + \left(1 - \pi_{ia}\right) \frac{1}{\sigma_W} \varphi\left(\frac{w_{ia} - X_{ia}\theta_2}{\sigma_W}\right)}$$

$$= \frac{\pi_{ia}\varphi\left(\frac{w_{ia}-X_{ia}\theta_1}{\sigma_W}\right)}{\pi_{ia}\varphi\left(\frac{w_{ia}-X_{ia}\theta_1}{\sigma_W}\right) + (1-\pi_{ia})\,\varphi\left(\frac{w_{ia}-X_{ia}\theta_2}{\sigma_W}\right)} \tag{2.13}$$

où φ est la densité de la loi normale centrée réduite. Pour résumer, tant que les individus n'ont pas été sur le marché du travail, ils disposent d'un *a priori* initial π_{i0} fonction de leur aptitude à l'école. Ensuite, à chaque passage sur le marché du travail, ils actualisent de manière bayésienne leur croyance en utilisant le salaire observé en fin de période. Ce phénomène d'apprentissage explique en partie dans notre modèle pourquoi des individus qui ont déjà quitté l'école peuvent choisir d'y revenir.

2.2.5 Equations de Bellman

Ayant défini les utilités instantanées, le déroulement des événements et l'ensemble d'information des agents il est possible de préciser les équations de Bellman données dans la section 2.2.1. En effet les hypothèses faites permettent d'écrire que

$$\begin{aligned} E\left(U_{ia}^S|\Omega_{ia}\right) &= E\left(Z_{ia}\delta_Z + \nu_i^S + \epsilon_{ia}^S|\Omega_{ia}\right) \\ &= Z_{ia}\delta_Z + \nu_i^S + \epsilon_{ia}^S \end{aligned}$$

car $U_{ia}^S \in \Omega_{ia}$. De même,

$$\begin{aligned} E\left(U_{ia}^H|\Omega_{ia}\right) &= U_{ia}^H \\ &= \nu_i^H + \epsilon_{ia}^H \end{aligned}$$

Enfin,

$$
\begin{aligned}
E\left(U_{ia}^{W}|\Omega_{ia}\right) &= E\left(X_{ia}\theta_i + \epsilon_{ia}^{W}|\Omega_{ia}\right) \\
&= E_{\theta_i,\epsilon_{ia}^{W}}\left(X_{ia}\theta_i + \epsilon_{ia}^{W}|X_{ia}\right) \\
&= X_{ia}\left[\pi_{ia}\theta_1 + \left(1-\pi_{ia}\right)\theta_2\right]
\end{aligned}
$$

On peut donc en déduire les équations de Bellman

$$V_{ia}^{S}\left(\Omega_{ia}\right) = Z_{ia}\delta_Z + \nu_i^S + \epsilon_{ia}^{S} + \beta E_{\epsilon_{ia+1}^{S},\epsilon_{ia+1}^{H}}\left[V_{ia+1}\left(\Omega_{ia+1}\right)/\Omega_{ia}, d_{ia}^{S}=1\right] \tag{2.14}$$

$$V_{ia}^{H}\left(\Omega_{ia}\right) = \nu_i^H + \epsilon_{ia}^{H} + \beta E_{\epsilon_{ia+1}^{S},\epsilon_{ia+1}^{H}}\left[V_{ia+1}\left(\Omega_{ia+1}\right)/\Omega_{ia}, d_{ia}^{H}=1\right] \tag{2.15}$$

$$V_{ia}^{W}\left(\Omega_{ia}\right) = X_{ia}\left[\pi_{ia}\theta_1 + \left(1-\pi_{ia}\right)\theta_2\right] + \beta E_{w_{ia},\epsilon_{ia+1}^{S},\epsilon_{ia+1}^{H}}\left[V_{ia+1}\left(\Omega_{ia+1}\right)/\Omega_{ia}, d_{ia}^{W}=1\right] \tag{2.16}$$

Remarquons que pour les valeurs de l'éducation et de la production domestique, l'espérance est uniquement prise par rapport à ϵ_{ia+1}^{S} et ϵ_{ia+1}^{H}, seules variables utilisées par l'agent l'année suivante mais qu'il ne connaît pas aujourd'hui. Pour la valeur du travail, l'espérance est également prise par rapport à w_{ia}, car l'individu sait que s'il travaille cette année, il observera son salaire à la fin de l'année, l'utilisera pour actualiser sa probabilité d'être de type 1 sur le marché du travail, et se servira de cette information pour déterminer son choix l'année prochaine. Mais il ne connaît pas son salaire en début de période, c'est pourquoi il faut également intégrer par rapport au salaire. On voit donc que dans un modèle avec information imparfaite, la valeur s'écrit d'une manière très différente de celle du modèle en information parfaite de Keane et Wolpin (1997). Les différences viennent du fait que l'utilité instantanée n'est pas parfaitement observée et qu'il faut prendre en compte l'observation du salaire en fin de période pour calculer l'espérance de la valeur future.

2.2.6 Résolution du modèle

Comme tous les modèles structurels dynamiques de choix discret, notre modèle se résout par récurrence arrière (voir Eckstein et Wolpin (1989) et Rust (1994)). Il s'agit dans un premier temps de caculer les fonctions valeurs à la dernière date à laquelle l'individu a la possibilité de choisir. Ensuite, nous utilisons les équations de Bellman pour calculer les valeurs antérieures par récurrence arrière. Pour tenir compte de ''la malédiction de la dimension", nous minimisons la taille de l'espace état en supposant qu'à 30 ans les individus sont obligés d'aller sur le marché du travail, et ce jusqu'à la fin de l'horizon A. Ainsi, 29 ans est le dernier âge auquel les agents ont un choix à effectuer. Notons également que de manière classique, nous discrétisons π_{ia} en $J+1$ valeurs qui divisent l'intervalle $[0,1]$ en J intervalles de même longueur : $\pi_{ia} \in \{\pi_j, j = 1...J+1\}$, où $\forall j \in \{1,...,J+1\}$, $\pi_j = \frac{j-1}{J}$.

Espérance des fonctions valeurs pour le dernier choix possible

Calculons tout d'abord l'espérance de la fonction valeur d'un individu qui choisirait d'aller à l'école à 29 ans.

$$E\left\{V_{i30}\left(\Omega_{i30}\right)|\Omega_{i29}, d_{i29}^S = 1\right\}$$

$$= \sum_{a=30}^{A} \beta^{(a-30)} E\left\{U_{ia}^W|\Omega_{i29}, d_{i29}^S = 1, d_{i30}^W = 1, ..., d_{ia-1}^W = 1\right\}$$

$$= \sum_{a=30}^{A} \beta^{(a-30)} E\left\{X_{ia}\theta_i + \epsilon_{ia}^W|\Omega_{i29}, d_{i29}^S = 1, d_{i30}^W = 1, ..., d_{ia-1}^W = 1\right\}$$

$$= \sum_{a=30}^{A} \beta^{(a-30)} E_{\pi_{ia}}\left\{X_{ia}\left[\pi_{ia}\theta_1 + (1-\pi_{ia})\,\theta_2\right]|\Omega_{i29}, d_{i29}^S = 1, d_{i30}^W = 1, ..., d_{ia-1}^W = 1\right\}$$

Notons à ce stade que lorsque la valeur est calculée du point de vue de l'âge 29, les valeurs des salaires qui se réaliseront sont inconnues, et donc les valeurs des probabilités π_{ia} aussi. C'est pourquoi dans la dernière ligne il convient d'intégrer également par rapport à π_{ia}. La forme étant linéaire nous sommes amenés à calculer $E_{w_{i30},w_{i31},\ldots,w_{ia-1}}(\pi_{ia}|\pi_{i29})$. Or l'hypothèse d'apprentissage bayésien rend très simple ce calcul. En effet une des propriétés fondamentales de l'apprentissage bayésien est que $E_{w_{ia-1}}(\pi_{ia}|\pi_{ia-1}) = \pi_{ia-1}$, formule qui signifie que sans information supplémentaire, l'anticipation de la probabilité *a posteriori* vaut exactement la probabilité *a priori*. On en conclut donc que $E_{w_{i30},w_{i31},\ldots,w_{ia-1}}(\pi_{ia}|\pi_{i29}) = \pi_{i29}$, et que

$$E\left\{V_{i30}(\Omega_{i30})\,|\Omega_{i29}, d_{i29}^S = 1\right\} = \sum_{a=30}^{A} \beta^{(a-30)} X_{ia}\left[\pi_{i29}\theta_1 + (1-\pi_{i29})\,\theta_2\right]$$

où $X_{ia} = [Educ_{i29}+1, Exp_{i29}+(a-30), 1]$. Nous calculons de même que l'espérance de la valeur future d'un individu qui choisit d'aller à la maison à 29 ans vaut

$$E\left\{V_{i30}(\Omega_{i30})\,|\Omega_{i29}, d_{i29}^H = 1\right\} = \sum_{a=30}^{A} \beta^{(a-30)} X_{ia}\left[\pi_{i29}\theta_1 + (1-\pi_{i29})\,\theta_2\right]$$

où $X_{ia} = [Educ_{i29}, Exp_{i29}+(a-30), 1]$. Enfin, l'espérance de la valeur future d'un individu qui choisit de travailler à 29 ans vaut

$$E\left\{V_{i30}(\Omega_{i30})\,|\Omega_{i29}, d_{i29}^W = 1\right\} = \sum_{a=30}^{A} \beta^{(a-30)} X_{ia}\left[\pi_{i29}\theta_1 + (1-\pi_{i29})\,\theta_2\right]$$

où $X_{ia} = [Educ_{i29}, Exp_{i29}+1+(a-30), 1]$. Ces valeurs terminales sont l'amorce de la récurrence arrière qui permet de calculer les espérances des valeurs futures pour les âges antérieurs. Nous allons maintenant déteminer analytiquement cette formule de récurrence. Mais préalablement nous présentons un résultat utile sur les espérances de lois normales tronquées.

Une propriété des lois normales tronquées

Supposons que $(X_1, X_2)'$ est un vecteur gaussien de variance unitaire et de correlation ρ. Notons $\Phi(h,k,p) = P(X_1 < h, X_2 < k)$ et $L(h,k,p) = P(X_1 > h, X_2 > k)$. Soient

$A(h,k,p) = \frac{k-\rho h^2}{\sqrt{1-\rho^2}}$, $B(h,k,p) = \frac{h-\rho k^2}{\sqrt{1-\rho^2}}$, $\varphi(x)$ et $\Phi(x)$ la densité et la fonction de répartition de la loi normale centrée réduite à x. Alors (Tallis, 1961)

$$E\left\{X_1/X_1 > h, X_2 > k\right\} = \frac{\varphi(h)\left\{1-\Phi\left[A(h,k,p)\right]\right\} + \rho\varphi(k)\left\{1-\Phi\left[B(h,k,p)\right]\right\}}{L(h,k,p)}$$

Plus généralement, si $(Y_1, Y_2)'$ est un vecteur gaussien d'espérance nulle, de variances σ_1^2 et σ_2^2 et de correlation ρ,

$$E\left\{Y_1/Y_1 > c_1, Y_2 > c_2\right\} \tag{2.17}$$

$$= \sigma_1 E\left\{X_1/X_1 > \frac{c_1}{\sigma_1}, X_2 > \frac{c_2}{\sigma_2}\right\} \tag{2.18}$$

$$= \sigma_1 \frac{\varphi(\frac{c_1}{\sigma_1})\left\{1-\Phi\left[A(\frac{c_1}{\sigma_1},\frac{c_2}{\sigma_2},p)\right]\right\} + \rho\varphi(\frac{c_2}{\sigma_2})\left\{1-\Phi\left[B(\frac{c_1}{\sigma_1},\frac{c_2}{\sigma_2},p)\right]\right\}}{L(\frac{c_1}{\sigma_1},\frac{c_2}{\sigma_2},p)} \tag{2.19}$$

$$= H(c_1, c_2, \sigma_1, \sigma_2, \rho) \tag{2.20}$$

Equations de récurrences arrières

Pour résoudre le modèle nous devons, à chaque âge a, calculer $E\left\{V_{ia}\left(\Omega_{ia}\right)|\Omega_{ia-1}, d^S_{ia-1} = 1\right\}$, $E\left\{V_{ia}\left(\Omega_{ia}\right)|\Omega_{ia-1}, d^H_{ia-1} = 1\right\}$ et $E\left\{V_{ia}\left(\Omega_{ia}\right)|\Omega_{ia-1}, d^w_{ia-1} = 1\right\}$.

- $E\left\{V_{ia}\left(\Omega_{ia}\right)/\Omega_{ia-1}, d^S_{ia-1} = 1\right\}$

$$
\begin{aligned}
& E\left\{V_{ia}\left(\Omega_{ia}\right)|\Omega_{ia-1}, d_{ia-1}^{S}=1\right\} \\
& = P\left(d_{ia}^{S}=1|\Omega_{ia-1}, d_{ia-1}^{S}=1\right) \\
& * E\left\{V_{ia}^{S}\left(\Omega_{ia}\right)|V_{ia}^{S}\left(\Omega_{ia}\right) > V_{ia}^{H}\left(\Omega_{ia}\right), V_{ia}^{S}\left(\Omega_{ia}\right) > V_{ia}^{W}\left(\Omega_{ia}\right), \Omega_{ia-1}, d_{ia-1}^{S}=1\right\} \\
& + P\left(d_{ia}^{H}=1|\Omega_{ia-1}, d_{ia-1}^{S}=1\right) \\
& * E\left\{V_{ia}^{H}\left(\Omega_{ia}\right)|V_{ia}^{H}\left(\Omega_{ia}\right) > V_{ia}^{S}\left(\Omega_{ia}\right), V_{ia}^{H}\left(\Omega_{ia}\right) > V_{ia}^{W}\left(\Omega_{ia}\right), \Omega_{ia-1}, d_{ia-1}^{S}=1\right\} \\
& + P\left(d_{ia}^{W}=1|\Omega_{ia-1}, d_{ia-1}^{S}=1\right) \\
& * E\left\{V_{ia}^{W}\left(\Omega_{ia}\right)/V_{ia}^{W}\left(\Omega_{ia}\right) > V_{ia}^{S}\left(\Omega_{ia}\right), V_{ia}^{W}\left(\Omega_{ia}\right) > V_{ia}^{H}\left(\Omega_{ia}\right), \Omega_{ia-1}, d_{ia-1}^{S}=1\right\}
\end{aligned}
$$

Soient

$$
\begin{aligned}
x =& Z_{ia}\delta_{Z} + \nu_{i}^{S} + \beta E\left\{V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1}, d_{ia-1}^{S}=1, d_{ia}^{S}=1\right\} \\
y =& \nu_{i}^{H} + \beta E\left\{V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1}, d_{ia-1}^{S}=1, d_{ia}^{H}=1\right\} \\
z =& X_{ia}\left[\pi_{ia}\theta_{1} + \left(1-\pi_{ia}\right)\theta_{2}\right] + \beta E\left\{V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1}, d_{ia-1}^{S}=1, d_{ia}^{w}=1\right\} \\
r =& \sigma_{H} \\
s =& \sigma_{S}
\end{aligned}
$$

Ainsi

$$
\begin{aligned}
P\left(d_{ia}^{S}=1|\Omega_{ia-1}, d_{ia-1}^{S}=1\right) &= P\left\{\epsilon_{ia}^{H} - \epsilon_{ia}^{S} < x-y, -\epsilon_{ia}^{S} < x-z\right\} \\
&= \Phi\left\{\frac{x-y}{\sqrt{r^{2}+s^{2}}}, \frac{x-z}{s}, \frac{s}{\sqrt{r^{2}+s^{2}}}\right\}
\end{aligned}
$$

De même,

$$
P\left(d_{ia}^{H}=1|\Omega_{ia-1}, d_{ia-1}^{S}=1\right) = \Phi\left\{\frac{y-x}{\sqrt{r^{2}+s^{2}}}, \frac{y-z}{s}, \frac{r}{\sqrt{r^{2}+s^{2}}}\right\}
$$

et

$$\begin{aligned} P\left(d_{ia}^{W}=1|\Omega_{ia-1},d_{ia-1}^{S}=1\right) &= P\left\{\epsilon_{ia}^{S}<z-x,\epsilon_{ia}^{H}<z-y\right\} \\ &= \Phi\left\{\frac{z-x}{s}\right\}\Phi\left\{\frac{z-y}{r}\right\} \end{aligned}$$

D'autre part,

$$\begin{aligned} & E\left\{V_{ia}^{S}\left(\Omega_{ia}\right)|V_{ia}^{S}\left(\Omega_{ia}\right)>V_{ia}^{H}\left(\Omega_{ia}\right),V_{ia}^{S}\left(\Omega_{ia}\right)>V_{ia}^{W}\left(\Omega_{ia}\right),\Omega_{ia-1},d_{ia-1}^{S}=1\right\} \\ & = x+E\left\{\epsilon_{ia}^{S}|\epsilon_{ia}^{S}-\epsilon_{ia}^{H}>y-x,\epsilon_{ia}^{S}>z-x\right\} \\ & = x+H\left\{z-x,y-x,s,\sqrt{r^{2}+s^{2}},\frac{s}{\sqrt{r^{2}+s^{2}}}\right\} \end{aligned}$$

De même,

$$\begin{aligned} & E\left\{V_{ia}^{H}\left(\Omega_{ia}\right)|V_{ia}^{H}\left(\Omega_{ia}\right)>V_{ia}^{S}\left(\Omega_{ia}\right),V_{ia}^{H}\left(\Omega_{ia}\right)>V_{ia}^{W}\left(\Omega_{ia}\right),\Omega_{ia-1},d_{ia-1}^{S}=1\right\} \\ & = y+H\left\{z-y,x-y,r,\sqrt{r^{2}+s^{2}},\frac{r}{\sqrt{r^{2}+s^{2}}}\right\} \end{aligned}$$

Et finalement,

$$\begin{aligned} & E\left\{V_{ia}^{W}\left(\Omega_{ia}\right)|V_{ia}^{W}\left(\Omega_{ia}\right)>V_{ia}^{S}\left(\Omega_{ia}\right),V_{ia}^{W}\left(\Omega_{ia}\right)>V_{ia}^{H}\left(\Omega_{ia}\right),\Omega_{ia-1},d_{ia-1}^{S}=1\right\} \\ & = E\left\{z/\epsilon_{it}^{S}<z-x,\epsilon_{it}^{H}<z-y\right\} \\ & = z \end{aligned}$$

Définissons maintenant

$$
\begin{aligned}
&f(x,y,z,r,s)\\
&=\Phi\left(\frac{x-y}{\sqrt{r^2+s^2}},\frac{x-z}{s},\frac{s}{\sqrt{r^2+s^2}}\right)\left[x+H\left(z-x,y-x,s,\sqrt{r^2+s^2},\frac{s}{\sqrt{r^2+s^2}}\right)\right]\\
&+\Phi\left(\frac{y-x}{\sqrt{r^2+s^2}},\frac{y-z}{s},\frac{r}{\sqrt{r^2+s^2}}\right)\left[y+H\left(z-y,x-y,r,\sqrt{r^2+s^2},\frac{r}{\sqrt{r^2+s^2}}\right)\right]\\
&+\Phi\left(\frac{z-x}{s}\right)\Phi\left(\frac{z-y}{r}\right)z
\end{aligned}
$$

Il est alors clair que

$$
\begin{aligned}
&E\left\{V_{ia}\left(\Omega_{ia}\right)|\Omega_{ia-1},d^S_{ia-1}=1\right\}\\
&=f\left\{Z_{ia}\delta_Z+\nu_i^S+\beta E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1},d^S_{ia-1}=1,d^S_{ia}=1\right],\right.\\
&\nu_i^H+\beta E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1},d^S_{ia-1}=1,d^H_{ia}=1\right],\\
&\left.X_{ia}\left[\pi_{ia}\theta_1+\left(1-\pi_{ia}\right)\theta_2\right]+\beta E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1},d^S_{ia-1}=1,d^W_{ia}=1\right],\sigma_H,\sigma_S\right\}
\end{aligned}
$$

Cette équation définit la récurrence arrière utilisée pour calculer $E\left\{V_{ia}\left(\Omega_{ia}\right)|\Omega_{ia-1},d^S_{ia-1}=1\right\}$ à tous les âges a.

- $E\left\{V_{ia}\left(\Omega_{ia}\right)/\Omega_{ia-1},d^H_{ia-1}=1\right\}$

Un calcul similaire donne

$$
\begin{aligned}
&E\left\{V_{ia}\left(\Omega_{ia}\right)|\Omega_{ia-1},d^H_{ia-1}=1\right\}\\
&=f\left\{Z_{ia}\delta_Z+\nu_i^S+\beta E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1},d^H_{ia-1}=1,d^S_{ia}=1\right],\right.\\
&\nu_i^H+\beta E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1},d^H_{ia-1}=1,d^H_{ia}=1\right],\\
&\left.X_{ia}\left[\pi_{ia}\theta_1+\left(1-\pi_{ia}\right)\theta_2\right]+\beta E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1},d^H_{ia-1}=1,d^W_{ia}=1\right],\sigma_H,\sigma_S\right\}
\end{aligned}
$$

- $E\left\{V_{ia}\left(\Omega_{ia}\right)|\Omega_{ia-1},d^W_{ia-1}=1\right\}$

Enfin,

$$
\begin{aligned}
& E\left\{V_{ia}\left(\Omega_{ia}\right)|\Omega_{ia-1}, d_{ia-1}^{W}=1\right\} \\
& = E_{w_{ia-1}}\ \{\ f\left\{Z_{ia}\delta_Z+\nu_i^S+\beta E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1}, d_{ia-1}^{W}=1, d_{ia}^{S}=1\right],\right. \\
& \nu_i^H+\beta E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1}, d_{ia-1}^{W}=1, d_{ia}^{H}=1\right], \\
& \left. X_{ia}\left[\pi_{ia}\theta_1+\left(1-\pi_{ia}\right)\theta_2\right]+\beta E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1}, d_{ia-1}^{W}=1, d_{ia}^{W}=1\right], \sigma_H, \sigma_S\right\}\ \}
\end{aligned}
$$

D'un point de vue technique, seule la dernière relation de récurrence pose problème. Tout d'abord elle nécessite le calcul d'une espérance par simulation. Comme pour tout a $w_{ia} = X_{ia}\theta_i + \epsilon_{ia}^W$, la loi de w_{ia-1} est un mélange gaussien $N\left(X_{ia-1}\theta_1, \sigma_W\right)$ avec probabilité π_{ia-1} et $N\left(X_{ia-1}\theta_2, \sigma_W\right)$ avec probabilité $1-\pi_{ia-1}$. D'autre part, lorsqu'un salaire est tiré, il est utilisé dans la formule d'apprentissage bayésien pour actualiser π_{ia-1} à π_{ia}. Mais de manière générale, π_{ia} ainsi actualisé ne prend pas une valeur de la grille qui sert à discrétiser ces probabilités. Il convient donc d'interpoler lineairement les espérances des valeurs futures dans la récurrence liée au travail. Ainsi, si l'on note $E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1}, d_{ia-1}^{W}=1, d_{ia}^{S}=1\right]$, $E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1}, d_{ia-1}^{W}=1, d_{ia}^{H}=1\right]$ et $E\left[V_{ia+1}\left(\Omega_{ia+1}\right)|\Omega_{ia-1}, d_{ia-1}^{W}=1, d_{ia}^{W}=1\right]$ respectivement $m_S(\pi_{ia})$, $m_H(\pi_{ia})$ and $m_W(\pi_{ia})$, alors $\forall k \in \{S, H, W\}$, $m_k(\pi_{ia})$ est approximé par

$$
m_k\left(\frac{[J\pi_{ia}]}{J}\right)+\frac{m_k\left(\frac{[J\pi_{ia}]+1}{J}\right)-m_k\left(\frac{[J\pi_{ia}]}{J}\right)}{\frac{1}{J}}\left(\pi_{ia}-\frac{[J\pi_{ia}]}{J}\right)
$$

où pour tout réel x, $[x]$ désigne la partie entière de x.

2.3 Identification

L'identification des modèles structurels dynamiques de choix discrets a été étudiée par Rust (1994), Magnac et Thesmar (2002) et Heckman et Navarro (2007). Rust (1994) se place dans le cadre d'un modèle dynamique à horizon infini, sans état absorbant et dans un environnement stationnaire. Il montre que ces modèles sont génériquement non paramétriquement non identifiés. Magnac et Thesmar (2002) précisent le degré de sous-identification. Se plaçant dans un modèle à horizon fini, et en supposant que l'économètre n'observe que les choix et la loi d'évolution des variables d'état, ils montrent que la sous identification réside essentiellement dans la distribution des chocs et le taux d'escompte. Ainsi, lorsque l'économètre est prêt à fixer *a priori* le taux d'escompte et à supposer que la distribution des chocs appartient à une certaine famille paramétrique, le modèle devient identifié. Ils montrent de plus que ce résultat n'est pas fondamentalement modifié lorsque de l'hétérogénéité inobservée est introduite dans le modèle. Pour Heckman et Navarro (2007), ce résultat de sous-identification n'a qu'une portée limitée. Ils pensent en effet que Rust (1994) et Magnac et Thesmar (2002) n'utilisent pas l'ensemble de l'information qui permet l'identification. De fait ils montrent que, dans un modèle à horizon fini avec un état absorbant, l'observation de résultats (les salaires dans notre cas) et une structure de facteurs avec des mesures permettent, sous certaines conditions, d'identifier semi-paramétriquement les préférences des individus, leur ensemble d'information, leurs croyances sur les rendements futurs et même l'actualisation de leurs croyances en fonction d'informations nouvelles qui arriveraient avec le temps.

Notre modèle est différent de celui analysé par Heckman et Navarro (2007), l'éducation n'est pas un état absorbant et nous n'imposons pas de structure à facteurs. Nous ne pouvons donc pas appliquer les résultats de ce travail. Au contraire, nous pensons que voir l'éducation comme un état absorbant réduit l'information qui permet d'identifier les phénomènes d'information et d'apprentissage. En effet, intuitivement, si les données permettent d'observer des individus qui reviennent à l'école après avoir fait le choix de la quitter, c'est que la nouvelle information qu'ils ont reçue (leur salaire) les a amené à reviser leur croyance et a rendu le choix d'une année d'étude à nouveau optimal. Nous observons de tels individus dans nos données et c'est, intuitivement, ce qui nous permet d'identifier l'ensemble d'information des agents et son évolution. Cependant, pour que cette idée soit raisonnable, il faut que le retour à l'école soit uniquement une conséquence de l'apprentissage. Or, on peut se dire par exemple que s'il existe des contraintes de crédit, les agents contraints peuvent décider de quitter l'école provisoirement pour accumuler suffisamment de liquidités afin de poursuivre leurs études par la suite. L'existence de contraintes de liquidités dans le domaine des choix d'éducation a été étudiée par de nombreux travaux, comme Heckman, Lochner et Taber (1998), Card (1995), Cameron et Heckman (1998, 2001), Carneiro et Heckman (2005), Kane (1996), Keane et Wolpin (2001) ou encore Stinebrickner et Stinebrickner (2007). Il semble se dégager de ces différents travaux que des contraintes de crédit pourraient exister, c'est-à-dire que si certains individus avaient eu la possibilité d'emprunter, ils auraient continuer leurs études. Ceci signifie que pour ces individus, la réduction de la consommation présente ou le travail étudiant ne suffisent pas au financement des études. La seule solution qui s'offre à ces individus est de travailler provisoirement à plein temps pour épargner ce qui permettra plus tard de reprendre les études. Mais le coût des études aux Etats-Unis est tel qu'il nous paraît improbable que travailler un ou deux ans permette d'économiser suffisamment pour reprendre ses études durablement. Il nous paraît donc improbable que les individus contraints reviennent à l'école après l'avoir quittée. Nous faisons cette hypothèse, ce qui assure que le retour à l'école est essentiellement un phénomène lié à l'apprentissage des individus.

Il faut également s'intéresser aux étudiants qui travaillent pour financer leurs études. Ces derniers observant un salaire à la fin de chaque période scolaire, on peut se demander si cette information peut leur permettre d'apprendre sur eux-mêmes. Si c'était le cas, notre modèle serait mal spécifié. Mais le travail étudiant étant le plus souvent faiblement qualifié et à temps partiel, nous supposerons que s'il y a effectivement apprentissage, il est négligeable par rapport à celui qui intervient lorsque les agents travaillent à temps plein.

Au-delà des arguments intuitifs qui nous semblent participer à l'identification du modèle, cette dernière est acquise grâce à un taux d'escompte fixé *a priori*, aux hypothèses paramétriques faîtes sur les lois des chocs et à la forme paramétrique qui découle de l'hypothèse d'apprentissage bayésien.

2.4 Estimation

Comme Keane et Wolpin (1997) ou Belzil et Hansen (2002) par exemple, nous estimons le modèle par maximum de vraisemblance. A chaque âge a entre 16 et 28 ans, nous observons pour chaque individu i (d_{ia}, w_{ia}), où d_{ia} prend les valeurs S, H ou W selon que l'individu i a choisi à l'âge a d'aller à l'école, de rester à la maison ou d'aller sur le marché du travail. w_{ia} est le salaire qu'a gagné l'individu i s'il a choisi d'aller travailler à l'âge a. Nous maximisons la vraisemblance conditionnellement à l'ensemble d'information initial Ω_{i16}

$$P\left[(d_{i16}, w_{i16}), (d_{i17}, w_{i17}), ..., (d_{i28}, w_{i28}) |\Omega_{i16}\right]$$

$$= \sum_{k=1}^{K} P\left[(d_{i16}, w_{i16}), (d_{i17}, w_{i17}), ..., (d_{i28}, w_{i28}) |\Omega_{i16}, \text{TypeSH}_i = k\right] P\left(\text{TypeSH}_i = k/\Omega_{i16}\right)$$

$$= \sum_{k=1}^{K} \prod_{a=16}^{28} P\left[(d_{ia}, w_{ia}) \,|\, (d_{ia-1}, w_{ia-1}), ..., (d_{i16}, w_{i16}), \Omega_{i16}, \text{TypeSH}_i = k\right] P\left(\text{TypeSH}_i = k|\Omega_{i16}\right)$$

Il faut remarquer que le problème des conditions initiales, c'est-à-dire l'éventualité d'une corrélation entre Ω_{i16} et l'hétérogénéité inobservée par l'économètre TypeSH_i, est pris en compte. Nous pensons en effet, comme Keane et Wolpin (1997), qu'en particulier le niveau d'éducation atteint à 16 ans a peu de chances d'être indépendant de l'hétérogénéité inobservée. Nous modélisons cette éventuelle corrélation par un logit multinomial

$$P\left(\text{TypeSH}_i = k/\Omega_{i16}\right) = P\left(\text{TypeSH}_i = k/\mathbb{1}_{\{Educ_{i16} \leq 9\}}\right) = \frac{exp\left(\mu_{k1} + \mu_{k2}\mathbb{1}_{\{Educ_{i16} \leq 9\}}\right)}{\sum_{j=1}^{K} exp\left(\mu_{j1} + \mu_{j2}\mathbb{1}_{\{Educ_{i16} \leq 9\}}\right)} \tag{2.21}$$

où, pour définir le $\text{TypeSH}_i = 1$ comme modalité de référence, nous supposons que $\mu_{11} = \mu_{12} = 0$.

Pour résumer, les paramètres à estimer sont les δ_Z de l'équation de l'école (3 paramètres), le support de θ dans l'équation de salaire (2*4 paramètres), le support de l'hétérogénéité inobservée des équations de l'école et de la maison (2*K paramètres), les variances des chocs (3 paramètres), les μ qui donnent la répartition des TypeSH selon le niveau initial d'éducation (2*(K-1) paramètres), et enfin les ξ qui décrivent comment les agents forment leur première anticipation en fonction de leur type à l'école et à la maison (K paramètres). Il y a donc en tout 12+5K paramètres à estimer. Avec K=4, il y en a 32.

La vraisemblance s'écrit alors

$$P\left[(d_{ia} = H, w_{ia} = .) / (d_{ia-1}, w_{ia-1}), ..., (d_{i16}, w_{i16}) / \Omega_{i16}, type\right]$$
$$= P\left(V_{ia}^H \geq V_{ia}^S, V_{ia}^H \geq V_{ia}^W / ...\right)$$

$$= P\left\{\epsilon_{ia}^S - \epsilon_{ia}^H \leq \nu_i^H + \beta E\left(V_{ia+1}/\Omega_{ia}, d_{ia}^H = 1\right) - \left[Z_{ia}\delta_Z + \nu_i^S + \beta E\left(V_{ia+1}/\Omega_{ia}, d_{ia}^S = 1\right)\right],\right.$$
$$-\epsilon_{it}^H \leq \nu_i^H + \beta E\left(V_{ia+1}/\Omega_{ia}, d_{ia}^H = 1\right)$$
$$\left.- \left\{X_{ia}\left[\pi_{ia}\theta_1 + (1 - \pi_{ia})\theta_2\right] + \beta E\left(V_{ia+1}/\Omega_{ia}, d_{ia}^w = 1\right)\right\} / ...\right\}$$

$$P\left[(d_{ia} = S, w_{ia} = .) / (d_{ia-1}, w_{ia-1}), ..., (d_{i16}, w_{i16}) / \Omega_{i16}, type\right]$$
$$= P\left(V_{ia}^S \geq V_{ia}^H, V_{ia}^S \geq V_{ia}^W / ...\right)$$

$$= P\left\{\epsilon_{ia}^H - \epsilon_{ia}^S \leq Z_{ia}\delta_Z + \nu_i^S + \beta E\left(V_{ia+1}/\Omega_{ia}, d_{ia}^S = 1\right) - \left[\nu_i^H + \beta E\left(V_{ia+1}/\Omega_{ia}, d_{ia}^H = 1\right)\right],\right.$$
$$-\epsilon_{ia}^S \leq Z_{ia}\delta_Z + \nu_i^S + \beta E\left(V_{ia+1}/\Omega_{ia}, d_{ia}^S = 1\right)$$
$$\left.- \left\{X_{ia}\left[\pi_{ia}\theta_1 + (1 - \pi_{ia})\theta_2\right] + \beta E\left(V_{ia+1}/\Omega_{ia}, d_{ia}^w = 1\right)\right\} / ...\right\}$$

$$P\left[(d_{ia} = W, w_{ia}) / (d_{ia-1}, w_{ia-1}), ..., (d_{i16}, w_{i16}) / \Omega_{i16}, type\right]$$
$$= P\left(w_{ia}/d_{ia} = W, ...\right) P\left(d_{ia} = W/...\right)$$

$$P\left(d_{ia}=W/...\right)=P\left(V_{ia}^{W}\geq V_{ia}^{S},V_{ia}^{W}\geq V_{ia}^{H}/...\right)$$

$$=P\left\{\epsilon_{ia}^{S}\leq X_{ia}\left[\pi_{ia}\theta_{1}+\left(1-\pi_{ia}\right)\theta_{2}\right]+\beta E\left(V_{ia+1}/\Omega_{ia},d_{ia}^{w}=1\right)-\left[Z_{ia}\delta_{Z}+\nu_{i}^{S}+\beta E\left(V_{ia+1}/\Omega_{ia},d_{ia}^{S}=1\right)\right],\right.$$

$$\left.\epsilon_{ia}^{H}\leq X_{ia}\left[\pi_{ia}\theta_{1}+\left(1-\pi_{ia}\right)\theta_{2}\right]+\beta E\left(V_{ia+1}/\Omega_{ia},d_{ia}^{w}=1\right)-\left[\nu_{i}^{H}+\beta E\left(V_{ia+1}/\Omega_{ia},d_{ia}^{H}=1\right)\right]/...\right\}$$

Ces trois probabilités sont simplement des fonctions de répartition de normales bivariées prises à des points particuliers qui dépendent de la solution du modèle. Enfin

$$P\left(w_{ia}/d_{ia}=W,...\right)$$

$$=P\left(w_{ia}/...\right)$$

$$=\pi_{ia}\frac{1}{\sigma_{w}}\varphi\left(\frac{w_{ia}-X_{ia}\theta_{1}}{\sigma_{w}}\right)+\left(1-\pi_{ia}\right)\frac{1}{\sigma_{w}}\varphi\left(\frac{w_{ia}-X_{ia}\theta_{2}}{\sigma_{w}}\right)$$

Notons pour finir que cette vraisemblance se calcule en trois phases. Dans un premier temps, en fonction des paramètres, les ensembles d'information sont déterminés à chaque date. Nous calculons en particulier l'anticipation initiale π_{i0} en fonctoin des ξ puis à chaque âge a π_{ia} en fonction des salaires éventuellement observés en utilisant la formule d'actualisation de l'apprentissage bayésien. Muni de ces ensembles d'informations nous sommes capables d'utiliser les formules de récurrence qui servent à résoudre le modèle et ainsi de calculer les espérances de valeurs futures qui apparaissent dans la vraisemblance. Notons que les π_{ia} précédemment obtenus n'appartenant pas en général à la grille des π, nous calculons à nouveau les espérances de fonction valeur par interpolation linéaire. Enfin, nous pouvons calculer la vraisemblance.

2.5 Données

L'immense majorité des travaux de la littérature structurelle s'intéressant aux choix dynamiques d'éducation utilisent tous la même base de données, le National Longitudinal Survey of Youth, 1979. Notre travail s'inscrivant pleinement dans cette littérature, nous utilisons la même base de données. La définition des variables est essentiellement la même que celle de Keane et Wolpin (1997).

2.5.1 La base de données et la définition des variables

La base de données existante la plus adaptée à notre modèle est le très étudié National Longitudinal Survey of Youth dans sa version de 1979 (NLSY 79). Cette base est une des bases favorites des économistes qui estiment des modèles structurels dynamiques de choix d'éducation. Le NLSY 79 consiste en 12686 individus, avec approximativement une moitié d'hommes, qui avaient entre 14 et 21 ans au premier janvier 1979. Pour ce qui nous concerne, les individus sont interrogés chaque année sur leur activité. Nous sommes ainsi capables de savoir si un individu a choisi d'aller à l'école, de travailler ou ni l'un ni l'autre. Des informations sont également collectées sur les revenus des individus. En particulier, les salaires et le temps de travail sont renseignés.

Pour définir les variables clé du modèle, nous adoptons les mêmes conventions que Keane et Wolpin (1997). Cette définition se fait de manière séquentielle. Tout d'abord, nous codons un individu comme ayant choisi l'école s'il a *validé* une année d'étude supplémentaire. Ainsi, ceux qui ont passé un an à étudier mais qui n'ont pas eu leur année ne sont pas considérés comme étudiants. Précisons également qu'un individu qui a eu son année et qui travaille pendant ses études est considéré uniquement comme étudiant, et ce même s'il travaille à plein temps. Si un individu n'est pas codé comme ayant choisi l'éducation, il est considéré comme ayant choisi de travailler s'il a travaillé au moins deux tiers des semaines de l'année (en excluant les mois de juillet et aôut) et en moyenne au moins vingt heures par semaines. Son salaire est alors calculé en ramenant le salaire total de l'année reporté par l'individu sur 50 semaines. Un individu qui ne serait ni classé à l'école ni sur le marché du travail se verrait codé comme étant à la maison.

Nous restreignons notre analyse aux hommes blancs dont on peut observer l'histoire sans valeur manquante entre l'âge de 16 ans et l'âge de 28 ans. Les salaires sont en effet essentiels à chaque fois qu'un individu a choisi de travailler car ils sont nécessaires à l'actualisation des croyances de l'agent. Considérant que notre modèle ne s'applique pas aux individus qui choisissent de poursuivre leurs études en PHD, nous supprimons ces derniers de l'échantillon en fixant un nombre d'années d'études maximal à 18 années. Nous éliminons également les individus qui reportent des trajectoires incohérentes temporellement qui nous interdisent donc tout classement. Pour finir, nous excluons de l'échantillon les valeurs extrêmes des salaires, ce qui nous conduit finalement à un échantillon de $591 * 13 = 7683$ individus-dates.

2.5.2 Statistiques descriptives

La table 2.1 donne la distribution des niveaux d'éducation atteints à 28 ans, observés et simulés par le modèle. On peut voir que le mode de cette distribution (43.6%) correspond à 12 années d'études, c'est-à-dire à la fin du lycée. Le deuxième mode (20.1%) correspond à 16 années d'études, l'équivalent du M1. Le reste de la distribution se répartit relativement uniformément entre 11 et 18 années d'études. Dans la table 2.2 on peut voir que de manière très classique, la proportion d'individus qui choisissent les études décroît avec l'âge ; parallèlement, la proportion d'individus qui choisissent d'aller sur le marché du travail croît ; enfin ceux qui choisissent la production domestique représentent environ 20% entre 18 et 22 ans et 15% par la suite.

Les retours à l'école participant de manière essentielle à l'identification du modèle, et en particulier à celle des paramètres relatifs à l'apprentissage des individus, il est intéressant de décrire ce phénomène dans les données. Par définition, un individu effectue un retour à l'école lorsqu'il a quitté l'école, effectué un passage sur le marché du travail et décide de revenir s'éduquer. Ainsi, un individu qui quitte l'école, reste à la maison puis revient à l'école n'est pas considéré comme ayant effectué un retour. On peut lire dans la table 2.3 que ce phénomène représente 25% de l'échantillon. Cette proportion loin d'être négligeable montre que les retours jouent leur rôle dans l'identification du modèle. Les trois quarts des individus qui effectuent un retour ne reviennent qu'une seule fois, le quart restant en effectue deux. A quels niveaux d'études se sont arrêtés ceux qui reviennent ? La table 2.4 nous montre que le premier retour s'effectue majoritairement pour les individus qui s'étaient arrêtés à la fin du lycée (37.8%) et qui veulent donc obtenir un diplôme universitaire ; ensuite la répartition est uniforme dans les niveaux supérieurs. Le deuxième retour se fait pour des individus qui veulent augmenter leur nombre d'années d'études à l'université. Dans la table 2.5, on voit qu'avant le premier retour, presque la moitié des individus (43.9%) n'ont travaillé qu'une année, les autres se répartissant entre 2 et 6 années. Entre le premier et le deuxième retour, les individus travaillent soit une (52.8%) soit deux (30.6%) années. Pendant le premier retour (table 2.6), les individus s'éduquent soit une (64.6%) soit deux (28.3%) années, alors que durant le deuxième retour, 80% d'entre eux ne font qu'une année d'études supplémentaire.

2.6 Résultats

Les résultats présentés dans cette section sont basés sur l'estimation du modèle avec 4 types d'individu à l'école et à la maison (K=4) et un taux d'actualisation β fixé à 0.9. Dans un premier temps nous évaluerons la qualité de l'ajustement du modèle, nous présenterons ensuite les estimations des paramètres du modèle, nous verrons comment on peut interpréter les types d'individus résultant de l'estimation du modèle, et nous présenterons enfin les résultats des analyses contrefactuelles.

2.6.1 Qualité de l'ajustement du modèle

Nous nous demandons dans cette section si le modèle permet par simulation de reproduire les statistiques descriptives décrites dans la section précédente. Pour simuler des trajectoires, nous procédons de la manière suivante. Nous utilisons comme donnée de départ les niveaux d'éducation initiaux (à 16 ans) des individus de l'échantillon. A partir des valeurs estimées des μ, nous utilisons le logit multinomial (2.21) pour simuler des TypeSH (Ecole-Maison) en fonction des niveaux d'éducation initiaux. Ensuite nous utilisons les estimations des ξ et le logit (2.12) pour simuler un θ_i à chaque individu. Nous calculons en utilisant (2.12) l'*a priori* initial des individus sur leur probabilité que leur θ_i soit égal à θ_1. Nous partons alors de l'âge de 16 ans pour simuler les trajectoires des individus. Nous simulons les valeurs des chocs relatifs à l'âge de 16 ans ϵ_{i16}^S, ϵ_{i16}^H et ϵ_{i16}^W en utilisant les estimations des écarts types σ_S, σ_H et σ_W. Nous lançons le modèle structurel qui nous donne les choix des individus à 16 ans. Si l'individu i a fait le choix du marché du travail, nous actualisons son *a priori* $\pi_{i16} = \pi_{i0}$ en π_{i17} en utilisant la simulation de son salaire et la formule d'apprentissage bayésien (2.13). Cette opération est réalisée jusqu'à l'âge de 28 ans. Ces trajectoires peuvent être comparées aux données réelles pour évaluer la capacité du modèle à reproduire les traits de l'échantillon observé.

L'objectif initial d'un tel modèle étant d'expliquer les choix d'éducation, regardons si la distribution finale des niveaux d'éducation est bien reproduite. Nous pouvons lire sur la dernière colonne de la table 2.1 que c'est bien le cas. Il existe uniquement une légère confusion entre 15 et 16 années d'études. Ce résultat n'est pas surprenant dans la mesure où les modèles structurels dynamiques de choix d'éducation ont déjà prouvé par le passé leur grande capacité à reproduire les données. Nous voyons de plus dans la table 2.2 que les proportions de choix à chaque âge sont également bien reproduites, sauf peut-être à l'âge de 17 ans.

Qu'en est-il des statistiques relatives au retour à l'école ? La première ligne de la table 2.3 nous montre que dans les données simulées, 39% des individus effectuent un retour à l'école, alors que ce chiffre est de 25% dans les données. Ce défaut doit être corrigé. Par contre on voit sur le bas de cette même table que le nombre de retours est bien reproduit. La table 2.4 montre que les niveaux auquels les individus se sont arrêtés avant de revenir sont également mal ajustés, en revanche la distribution du nombre d'années travaillées avant les retours de la table 2.5 est correcte dans les données simulées. Dans la table 2.6, la proportion d'individus qui font deux années d'études supplémentaires est légèrement sous-estimée.

Il apparaît donc dans l'ensemble que les statistiques qui concernent les choix et les niveaux atteints sont bien reproduites mais que celles relatives aux retours à l'école peuvent être améliorées.

2.6.2 Les paramètres et leurs interprétations

La table 2.7 présente les estimations des paramètres du modèle et leurs écarts-types[2]. Les deux premiers paramètres sont les estimations des rendements de l'éducation. Le premier correspond à ceux des individus de type $\theta_i = \theta_1$, le second à ceux des individus de type $\theta_i = \theta_2$. Les premiers valent environ 10.5% alors que les seconds valent environ 8.5%. La première remarque est que ces valeurs sont tout à fait raisonnables pour des rendements de l'éducation. Cependant, Keane et Wolpin (1997) trouvent des rendements de l'éducation de 7% pour les cols blancs et 2.4% pour les cols bleus. Nos estimations sont donc sensiblement plus élevées et moins discriminantes. D'ailleurs les écarts-types sont tels qu'on ne peut pas rejeter l'égalité statistique de ces deux paramètres. Il n'en reste pas moins que ces estimations, bien que peu précises, sont telles que le premier type a des rendements de l'éducation plus élevés que le second. On voit ensuite que les rendements de l'expérience sont concaves et que c'est le carré de l'expérience qui différencie les deux types, de telle sorte que les rendements de l'expérience du type $\theta_i = \theta_1$ diminuent moins avec le temps que ceux du type $\theta_i = \theta_2$. La constante de l'équation de salaire, par contre, discrimine fortement les deux types, celle du type $\theta_i = \theta_1$ étant fortement inférieure à celle du type $\theta_i = \theta_2$. Ces résultats montrent que la distinction entre les deux types sur le marché du travail se fait essentiellement sur la base de la constante, c'est-à-dire l'aptitude inobservée des individus

[2]Les écarts-types sont calculés par bootstrap, avec 50 tirages et 40 itérations par tirage.

sur le marché du travail. Les individus du type $\theta_i = \theta_1$ ont une aptitude moins élevée que ceux du type $\theta_i = \theta_2$, mais ont des rendements de l'éducation et de l'expérience légèrement supérieurs.

Les paramètres concernant l'utilité instantanée de l'école sont tous significatifs et ont les mêmes signes que ceux estimés par Keane et Wolpin (1997). Le signe négatif de la variable qui indique si l'individu a déjà quitté l'école au moment de faire son choix signifie qu'il existe des coûts psychologiques qui font que lorsqu'on a déjà quitté l'école il est plus coûteux de faire le choix d'y revenir que lorsqu'on ne l'a jamais quittée. Les signes négatifs des indicatrices de l'enseignement supérieur montrent que plus on poursuit ses études, plus les coûts objectifs (comme les frais d'inscription) et subjectifs augmentent. Dans le bas de la table on trouve les estimations de la constante de l'utilité instantanée de l'école, pour chacun des 4 TypeSH, et la constante de l'utilité instantanée de la production domestique. Ces constantes représentent les goûts et les aptitudes (inobservés par l'économètre mais connus de l'individu) des agents pour l'école et la production domestique. Les types 1 et 3 ont un goût pour les études plus élevé que les individus des types 2 et 4. Au sein de ces groupes, le goût pour le loisir est différent. Le type 1 a un goût pour le loisir plus prononcé que le type 3. Il en est de même pour les types 4 et 2 respectivement. On voit donc qu'il existe une forte hétérogénéité des individus quant à leurs goûts et aptitudes pour l'école et le loisir.

Les estimations des paramètres μ nous renseignent sur les conditions initiales, c'est-à-dire sur la corrélation entre le niveau d'éducation initial (à 16 ans) des individus et leur type école-maison. Ces paramètres étant ceux d'un logit multinomial, il est plus aisé d'interpréter ces résultats en recalculant les tableaux croisés sur la base des estimations des paramètres. C'est l'objet de la table 2.8. On y voit que les individus qui initialement ont le plus faible niveau d'étude sont sur-représentés parmi le type 2 (ceux qui ont un faible goût pour l'école) et sous-représentés parmi le type 1 (ceux qui ont goût élevé pour les études). La prise en compte des conditions initiales donne donc des résultats tout à fait cohérents.

Les estimations des ξ décrivent la corrélation entre les 4 types école-maison et les 2 types sur le marché du travail. Leur interprétation est facilitée par la lecture de la table 2.9 dans laquelle on voit clairement que le TypeSH2 est associé aux au type $\theta_i = \theta_1$, les autres types étant associés au type $\theta_i = \theta_2$. Autrement dit, seuls les individus qui ont un goût faible pour les études et pour le loisir sont plus souvent des individus qui ont une faible aptitude sur le marché du travail.

Pour aller plus avant dans la description des types d'individus estimés par le modèle, il est intéressant de les croiser avec les statistiques descriptives. Dans la table 2.10 on voit les niveaux d'éducation finaux atteints selon le type. On note que le type sur le marché du travail a très peu d'influence, alors que les TypeSH 1 et 3 font beaucoup plus d'études que les types 2 et 4. Ceci reflète le goût et les aptitudes de ces types pour l'éducation. La table 2.11 nous donne la distribution des retours conditionnellement aux types. On y voit clairement que les individus de TypeSH 1 et 3 (goût élevé pour les études) reviennent plus souvent que les autres, et qu'au sein de ces individus, ceux qui sont de type $\theta_i = \theta_1$ reviennent légèrement plus souvent que ceux du type $\theta_i = \theta_2$.

Le modèle nous donne également la possibilité d'évaluer la précision des anticipations initiales des individus concernant leur type sur le marché du travail (les π_{i0}). Ces anticipations initiales peuvent être lues dans la colonne de droite de la table 2.9. Les individus de TypeSH 1, 2, 3 et 4 se donnent initialement respectivement 40.7%, 75.0%, 34.2% et 36.9% de chances d'être de type $\theta_i = \theta_1$. La proportion dans la population d'individus de type $\theta_i = \theta_1$ étant de 52.6%, on voit que leur TypeSH leur donne effectivement une information. Comme nous simulons le type des individus sur le marché du travail, nous le connaissons parfaitement, et ce depuis l'âge de 16 ans pour chaque individu. Ainsi, nous observons ce type, à l'inverse des individus qui ne peuvent que l'apprendre une fois rentré sur le marché du travail, et ceci nous permet de calculer la proportion d'individus qui se trompent dans leur prévision initiale. Si nous forçons les individus à 16 ans à faire une prévision de leur type sur le marché du travail, ils répondront qu'ils sont de type $\theta_i = \theta_1$ si et seulement leur π_{i0} est supérieur à 0.5. En comparant cette prévision et leur vrai type, nous pouvons calculer le pourcentage d'individus qui se trompent. Cette proportion est de 34.9%, ce qui est loin d'être négligeable.

2.6.3 Analyse contrefactuelle

Un des avantages majeurs des modèles structurels est qu'ils permettent de réaliser des analyses contrefactuelles. Pour ce qui nous concerne, il est intéressant de se demander quel est l'impact de l'imperfection d'information sur les choix d'éducation. Pour répondre à cette question nous pouvons simuler quels seraient les choix d'éducation des individus, soumis aux mêmes chocs que dans la simulation précédente, mais dans un cadre d'information parfaite. Ainsi, au lieu d'affecter π_{i0} comme anticipation initiale des individus, nous leur affectons la probabilité 1 s'ils sont de type $\theta_i = \theta_1$, et 0 s'ils sont de type $\theta_i = \theta_2$. Autrement dit nous supposons qu'ils connaissent parfaitement leur type sur le marché du travail dès l'âge de 16 ans. La comparaison des trajectoires en information parfaite et en information imparfaite nous permet d'évaluer l'impact de l'information imparfaite sur les choix d'éducation.

Le premier résultat est qu'en information parfaite, 21.7% des individus ne feraient pas le même nombre d'années d'études. Ce chiffre est loin d'être négligeable. Parmi eux 56.9% feraient moins d'études. La table 2.12 détaille ce résultat et montre que ceux qui feraient des études différentes feraient soit 1 soit 2 années en plus ou en moins mais surtout 1. On peut lire dans la table 2.13 à quel niveau d'études se font ces changements. Alors qu'on aurait pu s'attendre à trouver une masse d'individus autour du nombre d'années d'études correspondant au passage à l'université (12 années), on voit que les changements sont répartis sur l'ensemble des niveaux d'étude. Ce résultat est sans doute du au fait que l'on ne prend pas en compte la convexité des rendements de l'éducation (Belzil et Hansen, 2002). D'autre part, en ce qui concerne les retours à l'école, on voit sur les tables 2.3 à 2.6 que la connaissance parfaite du type sur le marché du travail ne changerait fondamentalement rien.

Ceux qui se trompent dans la prévision initiale de leur type sur le marché du travail sont-ils sur-représentés parmi ceux qui s'éduqueraient différemment en information parfaite ? Comme le montre la table 2.14, la réponse est oui. Ce résultat est logique dans la mesure où ce sont ceux dont les anticipations initiales sont les plus mauvaises qui se trompent le plus dans l'anticipation des salaires futures et qui sont donc amenés à faire des choix différents. Enfin, il est intéressant de décrire les différences d'éducation conditionnellement au type des individus. En premier lieu on remarque sur la table 2.15 que conditionnellement au type école-maison, il existe une corrélation entre le type sur le marché du travail et les différences d'éducation. On voit clairement que les individus du type $\theta_i = \theta_1$ sont sur-représentés parmi ceux qui feraient plus d'études en information parfaite, alors que les individus du type $\theta_i = \theta_2$ sont sur-représentés parmi ceux qui en feraient moins. L'interprétation de ce résultat est claire. Rappelons-nous que les individus du type $\theta_i = \theta_1$ ont des rendements de l'éducation et de l'expérience légèrement supérieurs à ceux des individus du type $\theta_i = \theta_2$, et ont une abilité sur le marché du travail plus faible. Ainsi, lorsqu'un individu du type $\theta_i = \theta_1$ se trompe sur son type sur le marché du travail, il croît avoir des rendements de l'éducation plus faibles que ce qu'ils sont réellement et une aptitude sur le marché du travail plus forte. L'éducation devient donc moins rentable pour eux. C'est ce que l'on observe en constatant dans les simulations qu'ils s'éduqueraient plus s'ils connaissaient parfaitement leur type. Le raisonnement symétrique s'applique pour les individus de l'autre type.

Pour résumer les enseignements de cette analyse contrefactuelle, on peut dire que l'inefficacité engendrée par les problèmes d'information se traduisent par des niveaux d'éducation sous-optimaux pour une part non-négligeable de la population (environ 20%). Cependant les choix observés diffèrent peu des choix optimaux (la plupart du temps une année en plus ou en moins). On peut enfin noter que la distribution de cette inefficacité parmi les types d'individus est très intelligible et fournit une preuve de la cohérence interne du modèle.

2.7 Conclusion

L'objet de ce chapitre est de participer à la littérature sur les modèles structurels dynamiques de choix d'éducation en travaillant sur la structure de l'ensemble d'information. Nous remettons en cause l'idée selon laquelle les agents connaissent parfaitement leurs caractéristiques sur le marché du travail (rendements de l'éducation, de l'expérience, et aptitude) au moment de faire leur choix, alors même qu'ils n'ont jamais fait l'expérience du marché du travail. A l'inverse nous supposons que des anticipations sont formées sur la base des goûts et des aptitudes à l'école, dont on peut raisonnablement pensé qu'ils sont connus. Ce n'est qu'en observant leur productivité sur le marché du travail à travers leur salaire que les individus révisent leur croyance de manière bayésienne.

Au delà des formes fonctionnelles et des hypothèses paramétriques postulées dans le modèle, nous pensons que la part non négligeable des individus de l'échantillon qui font un retour à l'école après être passés sur le marché du travail participe à l'identification des paramètres associés au phénomène d'apprentissage. Bien que perfectible en ce qui concerne les retours à l'école, la qualité de l'ajustement du modèle estimé est relativement bonne par ailleurs. L'estimation du modèle structurel permet de simuler les choix des individus sous l'hypothèse contrefactuelle que les caractéristiques sur le marché du travail sont parfaitement observées à partir du premier choix.

Le premier enseignement de ce travail est que l'information parfaite est rejetée par les données. Il apparaît donc nécessaire de prendre en compte cette dimension dans la modélisation des choix d'éducation. Le deuxième enseignement est que ces problèmes d'information sont tels que 20% des individus font des choix différents de ceux qu'ils feraient en information parfaite. Ces choix ne diffèrent cependant que d'une ou deux années d'études. La question de savoir si l'inefficacité engendrée est importante n'est donc pas simple, car si elle touche une part non négligeable de la population, elle ne l'affecte que peu. Il apparaît clairement que ce sont les goûts et les aptitudes pour l'éducation qui font que les individus font plus ou moins d'études. Cependant il se peut que d'un point de vue macroéconomique, cette inefficacité, bien que marginale pour chaque individu, ait des effets non nuls par exemple en terme de consommation ou d'épargne car touchant une part non négligeable de la population. Si tel était le cas il faudrait mettre en place des politiques publiques visant à réduire ces problèmes d'information. Une telle politique serait par exemple de favoriser les stages en entreprise pendant les études ou développer l'apprentissage. En faisant l'expérience du marché du travail pendant leurs études, les individus disposeraient ainsi d'une information plus précise sur leurs salaires futurs.

TAB. 2.1 – Distribution du niveau d'éducation final

Niveau d'éducation	Observé	Simulé-Observé
7	0.00508	−0.00508
8	0.01015	−0.01015
9	0.01184	−0.00846
10	0.01523	−0.01184
11	0.03215	0.02707
12	0.43655	0.01015
13	0.07783	−0.00169
14	0.09306	0.00677
15	0.04399	0.05753
16	0.20135	−0.07614
17	0.04230	0.01354
18	0.03046	−0.00169

Note : La colonne ''Observé" reporte les statistiques issues du NLSY, la colonne ''Simulé-Observé" donne l'écart entre les distributions observées dans les données et simulées par le modèle.

Tab. 2.2 – Distribution des choix à chaque âge

	Education		Travail		Maison	
	Observé	Simulé-Observé	Observé	Simulé-Observé	Observé	Simulé-Observé
16	0.93063	−0.05076	0.00338	0.02200	0.06599	0.02876
17	0.88325	−0.14044	0.03723	0.06937	0.07953	0.07107
18	0.47039	0.04061	0.29272	−0.03046	0.23689	−0.01015
19	0.34856	0.06430	0.38748	−0.03723	0.26396	−0.02707
20	0.28088	0.05753	0.46531	−0.00846	0.25381	−0.04907
21	0.23350	−0.01354	0.53976	0.03553	0.22673	−0.02200
22	0.14044	−0.00338	0.64129	0.00846	0.21827	−0.00508
23	0.08629	0.02369	0.72420	−0.00338	0.18951	−0.02030
24	0.06430	0.00338	0.78173	−0.05922	0.15398	0.05584
25	0.04230	0.01523	0.80880	−0.03553	0.14890	0.02030
26	0.04569	0.01692	0.80711	−0.01015	0.14721	−0.00677
27	0.03046	0.01184	0.82572	0.01015	0.14382	−0.02200
28	0.03723	0.00677	0.83080	0.01523	0.13198	−0.02200

Note : La colonne ''Observé" reporte les statistiques issues du NLSY, la colonne ''Simulé-Observé" donne l'écart entre les distributions observées dans les données et simulées par le modèle.

TAB. 2.3 – Les retours à l'école et leur nombre

		Observé	Simulé	Information Parfaite
A fait au moins un retour		0.25042	0.39086	0.40778
Nombre de	1	0.75676	0.80519	0.82573
retours parmi	2	0.24324	0.18182	0.17012
les retours	3	0	0.01299	0.00415

Note : Un individu est supposé avoir fait un retour à l'école s'il l'a quittée et s'il a effectué un passage sur le marché du travail avant d'y revenir. La colonne ''Observé'' reporte les statistiques issues du NLSY, la colonne ''Simulé'' reporte l'ajustement du modèle et la dernière colonne donne les résultat de l'analyse contrefactuelle dans laquelle on suppose que les individus connaissent toujours parfaitement leur type sur le marché du travail.

TAB. 2.4 – Distribution du niveau d'éducation au moment du retour à l'école

	Niveau d'éducation	Observé	Simulé	Information parfaite
au premier retour	8	0.00676	0.00433	0.00415
	9	0.00676	0.02597	0.01245
	10	0.02703	0.16883	0.15768
	11	0.08784	0.24675	0.31950
	12	0.37838	0.15152	0.10373
	13	0.12838	0.11688	0.11203
	14	0.11486	0.09524	0.12033
	15	0.08784	0.12554	0.11203
	16	0.14865	0.04762	0.04564
	17	0.01351	0.01732	0.01245
au deuxième retour	10	0	0.08889	0.04762
	11	0	0.40000	0.42857
	12	0	0.02222	0.07143
	13	0.25000	0.24444	0.16667
	14	0.16667	0.08889	0.09524
	15	0.33333	0.11111	0.14286
	16	0.16667	0.02222	0.02381
	17	0.08333	0.02222	0.02381

Note : Un individu est supposé avoir fait un retour à l'école s'il l'a quittée et s'il a effectué un passage sur le marché du travail avant d'y revenir. La colonne ''Observé" reporte les statistiques issues du NLSY, la colonne ''Simulé" reporte l'ajustement du modèle et la dernière colonne donne les résultat de l'analyse contrefactuelle dans laquelle on suppose que les individus connaissent toujours parfaitement leur type sur le marché du travail.

TAB. 2.5 – Distribution du nombre d'années travaillées avant de retourner à l'école

	Années travaillées	Observé	Simulé	Information parfaite
Avant le premier retour	1	0.43919	0.39827	0.39004
	2	0.16216	0.21212	0.18257
	3	0.11486	0.16017	0.13693
	4	0.08108	0.06061	0.08299
	5	0.08108	0.05628	0.06639
	6	0.05405	0.05628	0.04979
	7	0.02703	0.03030	0.04979
	8	0.02027	0.00866	0.02490
	9	0.00676	0.01299	0.01245
	10	0.01351	0.00433	0.00415
Entre le premier et le deuxième retour	1	0.52778	0.40000	0.38095
	2	0.30556	0.08889	0.16667
	3	0.11111	0.20000	0.16667
	4	0.05556	0.11111	0.16667
	5	0	0.06667	0.02381
	6	0	0.06667	0.07143
	7	0	0.06667	0.02381

Note : Un individu est supposé avoir fait un retour à l'école s'il l'a quittée et s'il a effectué un passage sur le marché du travail avant d'y revenir. La colonne ''Observé'' reporte les statistiques issues du NLSY, la colonne ''Simulé'' reporte l'ajustement du modèle et la dernière colonne donne les résultat de l'analyse contrefactuelle dans laquelle on suppose que les individus connaissent toujours parfaitement leur type sur le marché du travail.

Tab. 2.6 – Distribution du nombre d'années d'études pendant le retour à l'école

		Observé	Simulé	Information parfaite
Premier Retour	1	0.64567	0.85849	0.88182
	2	0.28346	0.12736	0.11364
	3	0.04724	0.00943	0
	4	0.02362	0.00472	0.00455
Deuxième Retour	1	0.80000	0.91176	0.96774
	2	0.20000	0.08824	0.03226

Note : Un individu est supposé avoir fait un retour à l'école s'il l'a quittée et s'il a effectué un passage sur le marché du travail avant d'y revenir. La colonne ''Observé" reporte les statistiques issues du NLSY, la colonne ''Simulé" reporte l'ajustement du modèle et la dernière colonne donne les résultat de l'analyse contrefactuelle dans laquelle on suppose que les individus connaissent toujours parfaitement leur type sur le marché du travail.

TAB. 2.7 – Estimations des paramètres et écarts-types

Equation de salaire				
	Type	Paramètre	Estimation	Ecart-type
Education	$\theta_i = \theta_1$	θ_{11}	0.103	0.013
Education	$\theta_i = \theta_2$	θ_{12}	0.085	0.010
Expérience	$\theta_i = \theta_1$	θ_{21}	0.116	0.009
Expérience	$\theta_i = \theta_2$	θ_{22}	0.103	0.006
Expérience Carrée / 100	$\theta_i = \theta_1$	θ_{31}	-0.154	0.078
Expérience Carrée / 100	$\theta_i = \theta_2$	θ_{32}	-0.312	0.049
Constante	$\theta_i = \theta_1$	ν_1^W	7.257	0.181
Constante	$\theta_i = \theta_2$	ν_2^W	8.102	0.152
ln(Ecart-type des chocs)		$\ln(\sigma_W)$	-1.041	0.007
Utilité instantanée de l'éducation				
	Type	Paramètre	Estimation	Ecart-type
A déjà quitté l'école		δ_{DQ}	-0.7498	0.094
Education ≥ 13		δ_C	-1.538	0.137
Education ≥ 17		δ_M	-0.534	0.241
Constante	TypeSH 1	ν_1^S	10.974	0.201
Constante	TypeSH 2	ν_2^S	8.777	0.144
Constante	TypeSH 3	ν_3^S	10.350	0.200
Constante	TypeSH 4	ν_4^S	9.936	0.280
ln(Ecart-type des chocs)		$\ln(\sigma_S)$	-0.447	0.071
Utilité instantanée de la production domestique				
	Type	Paramètre	Estimation	Ecart-type
Constante	TypeSH 1	ν_1^H	9.3478	0.115
Constante	TypeSH 2	ν_2^H	8.838	0.143
Constante	TypeSH 3	ν_3^H	8.635	0.217
Constante	TypeSH 4	ν_4^H	9.583	0.177
ln(Ecart-type des chocs)		$\ln(\sigma_H)$	-0.444	0.073
Conditions initiales				
		Paramètre	Estimation	Ecart-type
		μ_{21}	0.241	0.279
		μ_{22}	1.680	0.158
		μ_{31}	0.037	0.230
		μ_{32}	0.188	0.105
		μ_{41}	-0.867	0.220
		μ_{42}	0.882	0.127
Anticipation initiale				
		Paramètre	Estimation	Ecart-type
		ξ_1	-0.376	0.127
		ξ_2	1.201	0.187
		ξ_3	-0.279	0.174
		ξ_4	-0.204	0.180

Tab. 2.8 – Distribution des TypeSH conditionnellement au niveau d'éducation initial

Education initiale	TypeSH			
	1	2	3	4
> 9	0.25918	0.36327	0.26327	0.11429
≤ 9	0.12871	0.61386	0.16832	0.08911
Echantillon	0.23689	0.40609	0.24704	0.10998

Tab. 2.9 – Distribution du type sur le marché du travail conditionnellement au TypeSH

TypeSH	2	1
1	0.59286	0.40714
2	0.25000	0.75000
3	0.65753	0.34247
4	0.63077	0.36923
Echantillon	0.47377	0.52623

Tab. 2.10 – Distribution des niveaux d'éducation finaux conditionnellement aux types

TypeSH	Type travail	8	9	10	11	12	13	14	15	16	17	18
1	1	0.000	0.000	0.000	0.000	0.010	0.046	0.096	0.153	0.372	0.209	0.111
1	2	0.000	0.000	0.000	0.000	0.017	0.052	0.100	0.151	0.396	0.173	0.108
2	1	0.002	0.002	0.023	0.146	0.811	0.013	0.000	0.000	0.000	0.000	0.000
2	2	0.000	0.005	0.047	0.195	0.734	0.016	0.000	0.000	0.000	0.000	0.000
3	1	0.000	0.000	0.000	0.000	0.035	0.168	0.263	0.286	0.203	0.039	0.004
3	2	0.000	0.000	0.000	0.001	0.064	0.218	0.218	0.292	0.168	0.028	0.007
4	1	0.000	0.000	0.013	0.032	0.775	0.146	0.032	0.000	0.000	0.000	0.000
4	2	0.000	0.002	0.014	0.083	0.752	0.134	0.012	0.000	0.000	0.000	0.000

Tab. 2.11 – Distribution des retours conditionnellement aux types

TypeSH	Type travail	Pas de retour	retour
1	1	0.45426	0.54574
1	2	0.55966	0.44034
2	1	0.63859	0.36141
2	2	0.67917	0.32083
3	1	0.53527	0.46473
3	2	0.62737	0.37263
4	1	0.71560	0.28440
4	2	0.72549	0.27451

Tab. 2.12 – Distribution des différences de niveau d'étude en information parfaite et imparfaite

Différence	Distribution
-5	0.00034
-4	0.00254
-3	0.00812
-2	0.03029
-1	0.08240
0	0.78274
1	0.06227
2	0.02183
3	0.00694
4	0.00237
5	0.00017

Note : En information parfaite nous supposons que les individus connaissent toujours parfaitement leur type sur le marché du travail. Cette table donne la distribution de la différence des niveaux d'études atteints en information parfaite par rapport à ceux atteints en information imparfaite.

TAB. 2.13 – Distribution des niveaux d'éducation en information imparfaite conditionnellement aux différences d'éducation information parfaite/inparfaite

	8	9	10	11	12	13	14	15	16	17	18
−5	0	0	0	0	0	0	0	0	0	0.5	0.5
−4	0	0	0	0	0	0	0	0.067	0.533	0.2	0.2
−3	0	0	0	0	0.021	0	0	0.313	0.375	0.229	0.063
−2	0	0	0	0.017	0.134	0.017	0.151	0.19	0.257	0.095	0.14
−1	0	0	0.012	0.053	0.23	0.125	0.088	0.154	0.17	0.121	0.045
0	0	0.002	0.013	0.07	0.491	0.071	0.066	0.09	0.128	0.044	0.025
1	0.005	0.003	0.035	0.207	0.106	0.095	0.122	0.193	0.158	0.076	0
2	0	0.008	0.008	0	0.054	0.225	0.419	0.186	0.101	0	0
3	0	0	0	0	0.122	0.512	0.293	0.073	0	0	0
4	0	0	0	0	0.286	0.286	0.429	0	0	0	0
5	0	0	0	0	0	1	0	0	0	0	0

Note : En information parfaite nous supposons que les individus connaissent toujours parfaitement leur type sur le marché du travail. Le chiffre 0.125, intersection de la ligne -1 et de la colonne 13 signifie que parmi ceux qui feraient une année d'étude en moins en information parfaite, 12.5% font 13 années d'études en information imparfaite. Ils feraient donc 12 années d'études en information parfaite.

Tab. 2.14 – Distribution des modifications d'études en information parfaite/imparfaite conditionnellement aux erreurs d'anticipations

	Etudes différentes	
Erreur d'anticipation	Non	Oui
Non	0.83407	0.16593
Oui	0.68717	0.31283
Echantillon	0.78274	0.21726

Note : En information parfaite nous supposons que les individus connaissent toujours parfaitement leur type sur le marché du travail. Les anticipations considérées interviennent à l'âge initial et sont relatives au type sur le marché du travail.

TAB. 2.15 – Distribution des modifications d'études en information parfaite/imparfaite conditionnellement aux types d'individus

		Un an de plus en information parfaite	
TypeSH	Type travail	Non	Oui
1	1	0.79690	0.20310
1	2	0.92251	0.07749
2	1	0.94863	0.05137
2	2	0.99306	0.00694
3	1	0.53942	0.46058
3	2	0.99368	0.00632
4	1	0.85780	0.14220
4	2	0.97794	0.02206
		Un an de moins en information parfaite	
TypeSH	Type travail	Non	Oui
1	1	0.84031	0.15969
1	2	0.81673	0.18327
2	1	0.98447	0.01553
2	2	0.79444	0.20556
3	1	1.00000	0.00000
3	2	0.71368	0.28632
4	1	0.97248	0.02752
4	2	0.93382	0.06618

Chapitre 3

Introduire de l'Hétérogénéité Inobservée Dans la Mobilité des Salaires : une Approche Semi-Paramétrique

3.1 Introduction

La mobilité des salaires décrit comment les individus se déplacent dans l'échelle des revenus. Son étude est traditionnellement justifiée par son effet égalisateur[1], qui signifie intuitivement qu'une hausse des inégalités de revenu (les riches deviennent plus riches par rapport aux pauvres) est moins un problème si la mobilité s'accroît (c'est-à-dire s'il devient plus facile de passer de pauvre à riche). Cependant, il existe une autre raison qui justifie de s'intéresser à l'étude de la mobilité des revenus, les conséquences de la structure de la mobilité des salaires sur la segmentation de la population. Pour clarifier cette idée

[1] Voir Atkinson et al. (1992) pour une introduction à la mobilité des salaires.

prenons un exemple très simple. Supposons que l'on soit intéressé par la manière dont les individus bougent par rapport à la médiane de la distribution des salaires. Supposons que la matrice de transition inconditionnelle est telle que les individus en-dessous de la médiane ont une probabilité de 0.6 de rester en-dessous d'elle (et donc une probabilité de 0.4 de monter) et que ceux initialement au-dessus ont une probabilité de 0.6 d'y rester (et donc une probabilité de 0.4 de descendre). Supposons maintenant qu'il existe dans le processus de mobilité de l'hétérogénéité qui se traduit par deux types d'individus, chacun (50% de la population) étant associé à une matrice de transition spécifique, la somme pondérée des matrices de transition conditionnelle étant égale à la matrice inconditionnelle. Supposons que les individus du type 1 ont une probabilité de 1 de rester en-dessous de la médiane et de 0.8 de descendre quand ils sont initialement au-dessus : lorsqu'ils sont en-dessous ils y restent et lorsqu'ils sont au-dessus la plupart d'entre eux descendent. D'une certaine façon les individus du type 1 sont fortement associés au bas de la distribution, ils y sont attirés et il est clair que si le processus se déroule indéfiniment, tous les travailleurs du type 1 se retrouveront en-dessous de la médiane. Supposons maintenant symétriquement que les individus du type 2 ont des probabilités de 1 de rester au-dessus de la médiane et de 0.8 de monter lorsqu'ils sont initialement en-dessous : ils sont attirés vers le haut de la distribution et sont tous au-dessus à l'équilibre stationnaire. Dans cet exemple très stylisé, il est clair que l'introduction de l'hétérogénéité dans le processus de mobilité rend la population totalement segmentée à l'équilibre stationnaire : tous les individus du type 1 sont en-dessous de la médiane et tous ceux du type 2 sont au-dessus.

Le but de ce chapitre est de suivre cette idée : y a-t-il de l'hétérogénéité dans la mobilité des salaires, et si oui quelles sont ses conséquences sur la segmentation de la population. A notre connaissance ce travail est le premier à soulever cette question dans ces termes, et il n'y a donc pas de littérature spécifique sur le sujet. Cependant, certains papiers sur la mobilité des salaires fournissent quelques réponses préliminaires. En ce qui concerne les Etats-Unis, la question de l'hétérogénéité *observée* a été étudiée par Gittleman and Joyce

(1995), Schiller (1994) et plus récemment par Kopczuk et al. (2007). Gittleman and Joyce (1995) trouvent que les jeunes, les moins éduqués et les afro-américains sont plus mobiles que ceux qui sont plus vieux, plus éduqués ou blancs. Schiller (1994) montre que dans les années 80 les jeunes femmes ont plus de mobilité descendante et moins de mobilité ascendante que les jeunes hommes. Kopczuk et al. (2007) confirment le résultat que les femmes montent moins fréquemment que les hommes dans l'échelle des salaires. Ainsi nous savons aujourd'hui que la structure de la mobilité des salaires n'est pas la même selon les différents groupes démographiques. Dans Weber (2002), l'auteur étudie la présence de dépendance d'état dans la mobilité des salaires et montre que si l'hétérogénéité *inobservée* n'est pas prise en compte, la dépendance d'état est surestimée. Ce papier est donc le premier à montrer la présence d'hétérogénéité inobservée dans le processus de mobilité. Cependant, sa description n'est pas fournie et la question de la segmentation ne peut donc pas être investiguée.

Dans ce chapitre, pour introduire de l'hétérogénéité dans les matrices de transition entre quintiles de salaire, la dynamique des quintiles de salaire est, comme dans Weber (2002), directement modélisée par un logit multinomial dynamique avec hétérogénéité inobservée. D'une part, les paramètres de dépendance d'état (c'est-à-dire ceux associés au quintile retardé), sont interprétés comme une mesure de la mobilité corrigée des caractéristiques individuelles. D'autre part, l'hétérogénéité inobservée, qui est spécifique à chaque individu, reflète des caractéristiques constantes dans le temps, observables (comme le sexe, la couleur de peau ou l'éducation) et inobservables.

Les principaux résultats sont les suivants. Tout d'abord, la présence de dépendance d'état ne peut pas être rejetée et nous montrons qu'elle prend la forme d'une rigidité à la baisse, traduisant certains mécanismes du marché du travail comme les comportements de prospection d'emploi ou les problèmes d'information asymétriques entre employés et employeurs. Ensuite, l'hétérogénéité inobservée joue un rôle majeur dans la mobilité des revenus, ce qui signifie que le processus de mobilité est spécifique à chaque individu et qu'il n'est donc pas suffisant de calculer des matrices de transition inconditionnelles pour étudier la mobilité des revenus. Chaque individu est fortement associé à une zone spécifique de la distribution des quintiles et est attiré vers elle, ce qui se traduit par une segmentation de la distribution des quintiles. A l'équilibre stationnaire, les individus passent ainsi la majorité de leur temps dans cette zone spécifique. De plus cette segmentation est fortement corrélée avec le sexe, la race et l'éducation : les hommes, les blancs et les plus éduqués se trouvent plus souvent dans les types associés au haut de la distribution des quintiles et les femmes, les non-blancs et les moins éduqués à ceux du bas. De manière cohérente avec la constance relative du degré de mobilité aux Etats-Unis entre 1970 et 2000, nous trouvons que la structure segmentée de la distribution des quintiles n'a pas changé sur la période. Enfin, l'introduction d'hétérogénéité inobservée nous permet de montrer que pour cette partition de la population, la mobilité des salaires réduit fortement les inégalités de long terme intra-groupes mais quasiment pas les inégalités inter-groupes.

Dans la section suivante nous décrivons les données utilisées et nous fournissons des statistiques descriptives, dans la section 3.3 nous présentons le modèle, son identification et la manière dont nous l'estimons, nous discutons ensuite les résultats dans la section 3.4 et nous concluons dans la section 3.5.

3.2 Description des données utilisées

3.2.1 La base de données

La problématique étudiée n'est pas spécifique à un pays en particulier, et la question du choix de la base de données se pose pleinement. Il est naturel de se demander en premier lieu si des données françaises pourraient être utilisées. Une étude de la mobilité en France effectuée par Buchinsky et al (2003) utilise les *Déclarations Annuelles de Données Sociales* (DADS) couplées à *l'Echantillon Démographique Permanent*. Ces données ont le grand avantage d'être administratives, ce qui minimise l'erreur de mesure, et la taille importante de l'échantillon permet des analyses fines. Cependant, ces données ne donnent aucune information sur les fonctionnaires, sur les chômeurs et les inactifs. Utiliser ces données implique donc de travailler sur une sous-population sélectionnée. Au contraire le Panel Study of Income Dynamics, représentatif de la population américaine, bien que de taille restreinte et sujet à l'erreur de mesure, permet de ne faire aucune sélection et de pouvoir contrôler des transitions entre l'emploi et le non-emploi (chômage, inactivité), et c'est la raison pour laquelle nous avons choisi d'utiliser ces données plutôt que des données françaises. Ces données sont une des sources favorites des économistes qui travaillent sur des problématiques relatives au salaire car les individus sont suivis sur de nombreuses années, et des informations sont collectées sur les salaires, les changements d'emploi, le non-emploi, ou encore les entreprises pour ne citer que cela. Ces informations longitudinales permettent par exemple de s'intéresser à l'évolution des inegalités de salaire, à celle de la mobilité salariale, ou encore à l'offre de travail. Le PSID est un panel représentatif de la population américaine qui a débuté en 1968 et qui est toujours actif aujourd'hui. L'unité d'observation est le ménage, les enquêtes étaient annuelles jusqu'en 1997 et sont réalisées depuis tous les deux ans. Il y avait 4800 ménages en 1968 et 7000 en 2001. En 1968, le PSID était composé de deux échantillons distincts, l'échantillon SRC, représentatif de la population américaine, et l'échantillon SEO, surreprésentant les ménages à faible revenu. Ainsi, les faibles revenus sont surreprésentés dans l'échantillon global et nous nous restreignons à l'échantillon SRC, sans utiliser de poids, comme le recommandent Moffit et Gottschalk (2002).

Pour introduire de l'hétérogénéité inobservée, il est nécessaire de disposer de données de panel. D'une part, pour limiter l'attrition, le pas de la mobilité doit être le plus court possible ; d'autre part, l'identification de l'hétérogénéité inobservée nécessite du mouvement dans les rangs des individus et donc un pas le plus grand possible[2]. Un pas de deux ans

[2] Il est bien connu dans la littérature sur la mobilité des salaires que cette dernière est une fonction croissante du pas avec lequel elle est mesurée. Ceci traduit simplement le fait que lorsque

est un bon compromis et est donc adopté dans ce travail. Par conséquent, dans ce chapitre, la mobilité doit être comprise comme une mobilité de court terme. Pour minimiser l'erreur de mesure dans les salaires, nous utilisons le revenu du travail annuel plutôt que le salaire horaire car le nombre d'heures travaillées est mal renseigné dans le PSID (voir Bound et al. (1994) pour une étude de validation du PSID). Pour les mêmes raisons, les quintiles de salaires sont préférés aux déciles ou aux ventiles[3].

Dans la mesure où la motivation principale de l'étude de la mobilité est son effet égalisateur, il est naturel de scinder la période 1970-2000 de manière cohérente avec l'évolution des inégalités de salaire. Or on sait (Katz et Autor, 1999) qu'aux Etats-Unis ces dernières ont été relativement stables dans les années 70, puis ont littéralement explosé dans les années 80 et ont continué à légèrement augmenter dans les années 90. La période est ainsi découpée en trois sous-périodes, les années 70 (1970, 1972, 1974, 1976, 1978), les années 80 (1980, 1982, 1984, 1986, 1988) et les années 90 (1990, 1992, 1994, 1996, 1998). Pour chaque décennie, nous retenons les individus qui n'ont aucune donnée manquante dans la décennie. La méthode économétrique utilisée est très consommatrice de données (Maximum de Vraisemblance Conditionnel). Pour maximiser la taille de l'échantillon, la seule restriction faite est de ne conserver que les 25-60 ans à chaque date de la décennie pour éviter les étudiants salariés. De plus cette méthode ne permet pas d'inclure d'explicatives qui varient dans le temps, c'est pourquoi les quintiles de salaire ne sont pas calculés directement sur les salaires mais avec les résidus de la régression des salaires sur une constante, l'expérience et l'expérience au carré. Pour prendre en compte le non-emploi nous suivons Buchinsky and Hunt (1999) et nous créons un quintile zero qui inclut les individus inactifs, au chômage ou travaillant à temps partiel (au plus 1200 heures par an dans les années 70, 1300 dans les années 80 et 1400 dans les années 90[4]). Enfin, nous éliminons de l'échantillon les indi-

la mobilité est mesurée avec un pas plus large les individus ont plus de temps pour bouger, et donc la mobilité augmente.

[3]De plus, étudier la mobilité entre déciles ou entre ventiles impliquerait l'estimation d'un trop grand nombre de paramètres étant données les tailles restreintes de nos échantillons.

[4]les seuils ont été choisis de telle sorte que 25 % des individus travaillent moins. Comme le

vidus qui effectuent des transitions trop peu fréquentes dans les données pour permettre l'identification de paramètres associés, à savoir les transitions 1-4, 1-5, 2-5, 5-2, 5-1 ou 4-1. L'échantillon final est composé de 2005*5 individus-années dans les années 70, 2642*5 dans les années 80 et 3104*5 dans les années 90[5]. Les femmes représentent chaque décennie environ 55% de l'échantillon (54.6 dans les années 70, 54.1 dans les années 80 et 54.7 dans les années 90), les non-blancs environ 10% (10.2, 8.9 et 8.7), et la part des diplômés d'université est croissante (38.3, 52.8 et 58.8), ce qui reflète l'augmentation générale du niveau d'éducation sur la période.

3.2.2 Statistiques descriptives

Pour décrire la mobilité relative des salaires, un des outils les plus appropriés est la matrice de transition entre quintiles. La table 3.1 présente les matrices de transition pour les années 70, 80 et 90. Ces résultats sont très classiques. Tout d'abord, pour toutes les périodes, la transition la plus probable est la stabilité. Il y a plus de mobilité dans le milieu de la distribution qu'aux extrêmes ; la vaste majorité des mouvements se font dans un quintile adjacent ; la mobilité ascendante est légèrement supérieure à la mobilité descendante dans le bas de la distribution et la mobilité descendante est plus grande que la mobilité ascendante dans le haut de la distribution. Enfin, la mobilité des salaires est restée assez stable sur la période, ce qui implique que la mobilité des salaires n'a pas compensé l'augmentation des inégalités de salaire sur la période.

nombre d'heures travaillées a augmenté entre 1970 et 2000, ces seuils ont également augmenté.

[5]Les tailles d'échantillons augmentent parce que dans le PSID les enfants de l'échantillon original sont suivis.

Tab. 3.1 – Matrices de transition inter-quintiles, 1970-1998

Quintile d'origine	Quintile d'arrivée 0	1	2	3	4	5	Total	Direction Bas	Stable	Haut	Total
Années 70											
0	83.7	9.8	3.1	2.0	0.8	0.6	100.0	0.0	83.7	16.3	100.0
1	19.9	57.0	19.7	3.4	0.0	0.0	100.0	19.9	57.0	23.1	100.0
2	9.1	12.0	54.4	20.6	3.9	0.0	100.0	21.1	54.4	24.5	100.0
3	5.4	3.3	14.9	50.8	20.0	5.7	100.0	23.5	50.8	25.7	100.0
4	3.6	0.0	3.7	17.3	57.8	17.7	100.0	24.5	57.8	17.7	100.0
5	2.4	0.0	0.0	4.3	16.5	76.7	100.0	23.3	76.7	0.0	100.0
Années 80											
0	74.8	13.0	6.2	3.4	1.6	1.0	100.0	0.0	74.8	25.2	100.0
1	20.9	53.5	20.9	4.7	0.0	0.0	100.0	20.9	53.5	25.6	100.0
2	10.6	14.1	49.6	21.7	4.1	0.0	100.0	24.7	49.6	25.7	100.0
3	9.3	2.9	13.9	50.3	21.0	2.5	100.0	26.2	50.3	23.5	100.0
4	6.0	0.0	3.0	15.5	57.1	18.4	100.0	24.5	57.1	18.4	100.0
5	3.8	0.0	0.0	3.1	14.6	78.5	100.0	21.5	78.5	0.0	100.0
Années 90											
0	73.3	12.9	5.3	4.4	2.2	1.9	100.0	0.0	73.3	26.7	100.0
1	22.0	56.9	17.0	4.0	0.0	0.0	100.0	22.0	56.9	21.1	100.0
2	11.0	13.1	54.1	18.1	3.6	0.0	100.0	24.1	54.1	21.8	100.0
3	10.5	2.9	15.0	51.1	17.6	2.9	100.0	28.4	51.1	20.5	100.0
4	7.7	0.0	2.7	15.7	58.4	15.6	100.0	26.0	58.4	15.6	100.0
5	7.2	0.0	0.0	2.0	14.6	76.2	100.0	23.8	76.2	0.0	100.0

Notes : PSID, 25-60 ans, salaires annuels. 2005 $*$ 5 observations dans les années 70, 2642 $*$ 5 dans les années 80 et 3104 $*$ 5 dans les années 90. Le quintile 0 comprend les individus inactifs, chômeurs ou travaillant à temps partiel (au plus 1200 heures dans les années 70, 1300 dans les années 80 et 1400 dans les années 90).

3.3 Cadre Econométrique

3.3.1 Environnement théorique

Plusieurs théories classiques du marché du travail sont utiles pour comprendre la mobilité des salaires. Ces modèles sont synthétisés dans Schiller (1977), Agell et Bennmarker (2007) et Cardoso (2006), et sont brièvement résumés dans cette section.

La stabilité des salaires peut être expliquée par plusieurs facteurs. Les premiers facteurs sont spécifiques à chaque individu. La population dans chaque quintile à chaque période est sélectionnée sur la base de caractéristiques fixes dans le temps, et a donc une certaine propension à rester dans le même quintile lors de la période suivante. Ces caractéristiques peuvent être liées aux investissements initiaux en capital humain général ou à d'autres caractéristiques comme le sexe ou la race, qui peuvent être sources de discrimination. Des caractéristiques qui varient dans le temps peuvent également générer de la stabilité : si ceux qui ont obtenu initialement les meilleurs diplômes accumulent plus de capital humain via l'expérience, les rangs initiaux sont renforcés. Les seconds facteurs sont communs à tous les individus. Les modèles de salaire d'efficience montrent que l'information imparfaite sur le marché du travail peut mener à de la rigidité des salaires à la baisse : les entreprises ne peuvent pas diminuer les salaires, même en cas de choc négatif de demande, parce que les salariés pourraient diminuer leur effort ou changer de firme. Deuxièmement, les facteurs institutionnels comme les négociations collectives, les législations sur la protection de l'emploi ou l'existence d'un salaire minimum national peuvent être source d'immobilité.

D'autre part, deux facteurs peuvent expliquer la mobilité des salaires. Les premiers sont les chocs de productivité, les seconds sont liés aux investissements en capital humain en cours d'emploi : si la formation en entreprise implique un sacrifice de revenus présents, on peut attendre des mouvements relatifs dans les rangs de revenus lorsque les individus soldent les gains ou les pertes dues à leurs décisions d'investissement.

3.3.2 Modéliser les niveaux ou les rangs ?

Pour modéliser une matrice de transition entre quintiles, deux options sont possibles : soit modéliser directement la dynamique des quintiles, soit modéliser la dynamique des salaires eux-mêmes et en déduire les matrices de transition. Etudions la deuxième possibilité. La littérature sur la dynamique des salaires s'est dans un premier temps principalement développée sous l'hypothèse que les salaires suivent un processus ARMA(1,1), comme par exemple Burkhauser et al (1997) ou encore Browning et al (2006). Supposons donc que $\forall i = 1...N$ et $\forall t = 1...T$

$$y_{it} = \mu_{it} + \beta y_{it-1} + \epsilon_{it} + \theta\epsilon_{it-1}$$

où y_{it} est le log-salaire de l'individu i à la date t ; μ_{it} est une constante spécifique à chaque individu autorisée à varier dans le temps ; β est le coefficient AR reflétant la dépendance d'état du processus ; θ est le coefficient MA qui mesure comment se propage le choc ; ϵ_{it} est un processus indépendant dans le temps, de moyenne nulle, et indépendant des réalisations passées des log-salaires. Usuellement, (voir par exemple Burkhauser et al (1997) ou Moffit et Gottschalk (1998)), les paramètres de ce modèle sont obtenus en utilisant des techniques de distance minimum qui minimisent la distance entre les moments observés et les moments déduits du modèle, fonctions des paramètres. Mais pour dériver une matrice de transition, il faut supposer que les chocs suivent une certaine loi. La première idée naturelle, celle poursuivie par Moffit et Gottschalk (1998), est de supposer que les chocs sont normaux, et ils montrent alors que le processus de mobilité est entièrement déterminé par la corrélation entre les salaires. En effet, sous l'hypothèse de normalité des chocs, la loi jointe de deux log-salaires successifs est une normale bivariée :

$$\begin{bmatrix} y_{it} \\ y_{it-1} \end{bmatrix} \sim N\left(\begin{bmatrix} m_{it} \\ m_{it-1} \end{bmatrix}, \begin{bmatrix} \sigma_{it} & \rho_{it}\sigma_{it}\sigma_{it-1} \\ \rho_{it}\sigma_{it}\sigma_{it-1} & \sigma_{it-1} \end{bmatrix}\right)$$

Ainsi, si l'on note $F_{y_{it}}$ et Φ la fonction de répartition de y_{it} et celle de la loi normale centrée réduite respectivement, et Q le nombre de tranches avec lequel on découpe l'intervalle [0,1] (Q=5 pour des quintiles), la probabilité $P_{k/j}$ que l'individu i soit dans le quintile j à la date $t-1$ et dans le quintile k à la date t vaut

$$P_{k/j} = P\left[\frac{k-1}{Q} \leq F_{y_{it}}(y_{it}) \leq \frac{k}{Q} \mid \frac{j-1}{Q} \leq F_{y_{it-1}}(y_{it-1}) \leq \frac{j}{Q}\right]$$

$$= QP\left[\frac{k-1}{Q} \leq F_{y_{it}}(y_{it}) \leq \frac{k}{Q}, \frac{j-1}{Q} \leq F_{y_{it-1}}(y_{it-1}) \leq \frac{j}{Q}\right]$$

$$= QP\left[\frac{k-1}{Q} \leq \Phi\left(\frac{y_{it}-m_{it}}{\sigma_{it}}\right) \leq \frac{k}{Q}, \frac{j-1}{Q} \leq \Phi\left(\frac{y_{it-1}-m_{it-1}}{\sigma_{it-1}}\right) \leq \frac{j}{Q}\right]$$

$$= QP\left[\Phi^{-1}\left(\frac{k-1}{Q}\right) \leq \frac{y_{it}-m_{it}}{\sigma_{it}} \leq \Phi^{-1}\left(\frac{k}{Q}\right), \Phi^{-1}\left(\frac{j-1}{Q}\right) \leq \frac{y_{it-1}-m_{it-1}}{\sigma_{it-1}} \leq \Phi^{-1}\left(\frac{j}{Q}\right)\right]$$

$$= Q\int_{\Phi^{-1}\left(\frac{j-1}{Q}\right)}^{\Phi^{-1}\left(\frac{j}{Q}\right)} \int_{\Phi^{-1}\left(\frac{k-1}{Q}\right)}^{\Phi^{-1}\left(\frac{k}{Q}\right)} g_{it}(u,v)dudv$$

où g_{it} est la densité d'une loi normale bivariée d'espérance nulle, de variances unitaires et de coefficient de correlation ρ_{it}. Ce calcul confirme que chaque élément de la matrice de transition inter-quintiles ne dépend que de cette corrélation. Ainsi, cette spécification gaussienne est très parcimonieuse, mais elle fait porter une contrainte très forte sur la matrice de transition. En effet, g_{it} étant symétrique ($g_{it}(u,v) = g_{it}(v,u)$), il suffit dans $P_{k/j}$ de renommer u en v et v en u et d'intervertir les intégrales pour se rendre compte que $P_{k/j} = P_{j/k}$: l'hypothèse de normalité du processus a pour conséquence de rendre la matrice de transition symétrique conditionnellement à ρ_{it}. Cette hypothèse de symétrie n'est a priori pas fondée[6] et il faut donc enrichir le modèle pour s'en écarter.

Introduire de l'hétérogénéité individuelle dans les coefficients β et θ ne suffit pas. En conservant la structure ARMA, Meghir et Pistaferri (2004) et Browning et al (2006) brisent la symétrie en introduisant un effet ARCH dans la dynamique des chocs, c'est-à-dire en introduisant une hypothèse de la forme $E(\epsilon_{it}^2) = \alpha\epsilon_{it-1}^2$. Si cet enrichissement a bien l'effet voulu, il a, selon Browning et al (2006), l'inconvénient de reposer sur l'hypothèse ad hoc de normalité de ϵ_{it} conditionnellement aux réalisations passées des log-salaires. D'autre part, Bonhomme et Robin (2008) proposent une modélisation de la dynamique des salaires basée sur une décomposition à l'aide de copules. Cette spécification très parcimonieuse de la dynamique des rangs leur permet de modéliser les densités marginales de y_{it} et y_{it-1} de manière très flexible. Ils modélisent les chocs par des mélanges de loi normales et l'hétérogénéité inobservée par une loi discrète avec un grand nombre de points de support. Bien que la famille de copules paramétriques qu'ils utilisent soit symétrique (Plackett), l'asymétrie présente dans les marges leur permet d'obtenir une matrice de transition non symétrique.

Ainsi, la littérature récente sur la dynamique des salaires a montré qu'il était possible de modéliser des matrices de transition inter-quintiles conditionnelles non-symétriques. Mais les spécifications utilisées reposent sur des lois paramétriques. Dans la mesure où notre objectif n'est pas en soi de modéliser la dynamique des salaires mais spécifiquement les matrices de transition, il est légitime de se demander si restreindre l'objet de la modélisation peut permettre des spécifications moins paramétriques. La réponse à cette question se trouve dans la littérature sur les modèles qualitatifs dynamiques. Les quintiles de salaires étant par nature ordonnés, il pourrait être envisageable d'employer un modèle ordonné dynamique, mais les probabilités de transition induites n'ont pas des formes facilement interprétables et à notre connaissance il n'est pas possible d'estimer de tels modèles de manière semi-paramétrique. En revanche, Magnac (2000) et Honore et Kyriazidou (2000) ont montré qu'il est possible, en modélisant de manière très transparente les probabilités

[6] Et on verra par la suite qu'elle est rejetée par les données.

de transition, d'implémenter des méthodes semi-paramétriques dans le cadre d'un logit multinomial dynamique avec hétérogénéité inobservée. Nous empruntons donc cette voie.

3.3.3 Le modèle

Nous modélisons la dynamique des quintiles de salaire par un logit multinomial dynamique avec hétérogénéité inobservée. Plus précisément, notons y_{it} ($y_{it} = 0...5$) le quintile de l' individu i ($i = 1...N$) à la date t ($t = 1...T$). Nous supposons que

$$\forall k = 0...5\,,\, \forall t = 2...T\,,\; y_{it} = k \,\text{si et seulement si}\, y^*_{ikt} = \max_{j=0...5}\{y^*_{ijt}\} \text{ où}$$

$$\forall k = 0...5\,,\, \forall t = 2...T\,,\; y^*_{ikt} = \sum_{j=0}^{5} \delta_{jk} \mathbb{1}_{\{y_{it-1}=j\}} + \alpha_{ik} + \epsilon_{ikt} \tag{3.1}$$

Comme dans tout modèle qualitatif, un état de référence doit être choisi et certains paramètres fixés à zéro. Nous choisissons le quintile 0 et nous supposons que $\forall j = 0...5$, $\delta_{j0} = 0$, $\forall i = 1...N$, $\alpha_{i0} = 0$ et $\forall k = 0...5$ $\delta_{0k} = 0$. Parce que nous supprimons de l'échantillon tous les individus qui réalisent une transition extrême (1-4, 1-5, 2-5, 5-2, 5-1 or 4-1), ces transitions doivent être impossibles dans le modèle et c'est pourquoi nous supposons que $\delta_{14} = \delta_{15} = \delta_{25} = \delta_{41} = \delta_{51} = \delta_{52} = -\infty$. Nous supposons que les ϵ's sont distribués comme une loi de valeurs extrêmes de type I, et qu'ils sont indépendants entre alternatives, entre individus et dans le temps, et indépendants des α's. Nous notons $\forall i = 1...N$, $\alpha_i = (\alpha_{i0}, ..., \alpha_{i5})'$. Par conséquent, la probabilité qu'un individu i soit dans le quintile k à la période t ($t = 2...T$), sachant qu'il était dans le quintile j à la période $t-1$ vaut

$$P(y_{it} = k \,/ y_{it-1} = j\,,\, \alpha_i) = \frac{\exp(\delta_{jk} + \alpha_{ik})}{\sum_{l=0}^{5} \exp(\delta_{jl} + \alpha_{il})} \tag{3.2}$$

ce qui implique que la matrice de transition est hétérogène entre individus. Enfin, nous supposons que la loi de y_{i1} est une fonction non spécifiée de α_i, ce qui impose de prendre en compte le problème des conditions initiales, c'est-à-dire la corrélation qui existe entre le premier quintile observé et l'hétérogénéité inobservée.

D'un point de vue statistique, ce modèle peut être considéré comme un ''équivalent non-linéaire" des ARMA utilisés dans la littérature sur la dynamique des salaires : le côté droit de (3.1) est composé du retard de la variable dépendante (la composante AR), un effet fixe et un choc transitoire[7].

L'interprétation des paramètres est facilitée en voyant que

$$\frac{P(y_{it} = k\, / y_{it-1} = j\, ,\, \alpha_i)}{P(y_{it} = l\, / y_{it-1} = j\, ,\, \alpha_i)} = \exp((\delta_{jk} - \delta_{jl}) + (\alpha_{ik} - \alpha_{il})) \tag{3.3}$$

et avec $\alpha_{ik} = \alpha_{il}$,

$$\frac{P(y_{it} = k\, / y_{it-1} = j\, ,\, \alpha_i)}{P(y_{it} = l\, / y_{it-1} = j\, ,\, \alpha_i)} = \exp(\delta_{jk} - \delta_{jl}) \tag{3.4}$$

α_{ik} représente la propension individuelle de l'individu i fixe dans le temps à se trouver dans le quintile k, et fait donc référence dans la section 3.3.1 aux caractéristiques spécifiques à chaque individu et fixes dans le temps qui poussent à la stabilité, comme l'éducation initiale, le sexe, la race, etc. Ces paramètres d'hétérogénéité inobservée capturent des effets de sélection. Pour prendre par exemple l'éducation, il est clair qu'un bon diplôme devrait augmenter la propension d'être dans un quintile élevé, et diminuer celle d'être dans un

[7]Cependant, les travaux comme Burhauser et al (1997) trouvent également une composante MA(1) dans la dynamique des salaires et notre hypothèse AR(1) est donc quelque peu restrictive. La présence éventuelle d'une composante MA dans la dynamique des quintiles de salaire fait partie des pistes de recherches futures.

quintile faible. L'effet de ces variables est donc probablement différent selon le quintile et il est donc nécessaire que les effets individuels α_{ik} soient spécifiques à chaque quintile, c'est-à-dire indicés par k dans (3.1). L'équation (3.3) montre que, pour des δ's fixés, plus α_{ik} est supérieur à α_{il}, plus l'individu i a de chances d'arriver dans le quintile k plutôt que dans le quintile l. Maintenant, si $\alpha_{ik} = \alpha_{il}$, c'est-à-dire pour un individu qui a la même propension individuelle d'être dans les quintiles k et l, l'équation (3.4) montre que $\delta_{jk} - \delta_{jl}$ peut être interprété comme une différence de (logarithme de) probabilités. En ce sens, les paramètres de dépendance d'état fournissent une mesure de la mobilité des salaires corrigée des caractéristiques individuelles. Les estimations de ces paramètres permettront d'aller plus loin dans leur interprétation économique.

3.3.4 Identification et Estimation

Pour estimer ce modèle, une procédure en deux étapes est implémentée. Dans la première étape, les paramètres de dépendance d'état (les δ's) sont estimés avec une technique semi-paramétrique, et ces estimations sont utilisées dans une deuxième étape pour estimer les paramètres d'hétérogénéité inobservée (les α's).

Dans la première étape nous appliquons la méthode de maximum de vraisemblance conditionnel développée par Magnac (2000). L'idée est de conditionner la vraisemblance par une statistique suffisante telle que la vraisemblance conditionnelle ne dépend plus des effets individuels. Cette méthode est très attirante car elle résoud directement le problème des conditions initiales (Heckman, 1981b) et elle permet d'obtenir les paramètres de dépendance d'état sans faire la moindre hypothèse sur la loi de l'hétérogéhéité inobservée[8]. Cependant cette technique ne permet pas de prendre en compte de covariables variant dans le temps[9].

[8]Une procédure en une étape nécessitant, pour estmier les δ's, de spécifier la loi de l'hétérogénéité inobservée et d'approximer la corrélation entre cette dernière et les conditions initiales, ces caractéristiques justifient l'utilisation d'une méthode en deux étapes.

[9]Pour inclure des variables qui varient dans le temps, il existe la méthode semi-paramétrique d'Honoré et Kyriazidou (2000), mais la taille trop petite de notre échantillon ne nous permet pas de l'utiliser.

En particulier nous n'incluons dans le modèle ni de tendance temporelle ni de variable d'expérience, ce qui pose problème. Pour surmonter cette difficulté, les effets du temps et de l'expérience sont éliminés préalablement des salaires : les quintiles ne sont pas calculés directement sue les salaires, mais sur les résidus de la régression des log-salaires sur une constante, l'expérience et l'expérience au carré[10].

En notant Y_{ik} le nombre de fois ou l'individu i a été dans le quintile k entre les périodes 2 et $T-1$ de la décennie, Magnac (2000) montre dans son annexe B que

$$P\left(y_{i2},...,y_{iT-1}\,/y_{i1},Y_{i0},...,Y_{i5},y_{iT}\right)=\frac{\exp\sum_{k>0}\sum_{j}\left(\sum_{t>1}\mathbb{1}_{\{y_{it-1}=j\}}\mathbb{1}_{\{y_{it}=k\}}\delta_{jk}\right)}{\sum_{B}\exp\sum_{k>0}\sum_{j}\left(\sum_{t>1}\mathbb{1}_{\{y_{it-1}=j\}}\mathbb{1}_{\{y_{it}=k\}}\delta_{jk}\right)} \tag{3.5}$$

où $B=\{b=(y_{i2},...,y_{iT-1})\,/\forall\,k>0,\ \sum_{t=2}^{T-1}\mathbb{1}_{\{y_{it-1}=j\}}=Y_{ik}\}$ est l'ensemble de toutes les histoires possibles qui sont compatibles avec le nombre de visites dans chaque quintile entre 2 et $T-1$. Cette vraisemblance conditionnelle ne dépend plus de α_i, ce qui veut dire intuitivement que les individus qui ont le même nombre de visites dans chaque quintile, mais pas aux mêmes dates, ont le même niveau d'hétérogénéité inobservée. Autrement dit, si l'on voulait faire ''à la main''des groupes d'individus qui ont des matrices de transitions similaires , on regrouperait ceux qui ont à peu près le même nombre de visites dans chaque quintile. C'est l'hypothèse principale qui permet d'identifier les δ's.

Pour estimer la loi de l'hétérogénéité inobservée, ie la loi de α_i, des hypothèses doivent être faites et donc des techniques à effets aléatoires doivent être utilisées. Pour minimiser l'impact des hypothèses sur les distributions, nous suivons Heckman et Singer (1984) et nous choisissons une loi discrète, le nombre de points de support (ou nombre de types d'individus) étant déterminé par un processus itératif croissant que nous décidons d'arrêter lorsque

[10]Ces régressions sont réalisées année par année.

l'ajustement du modèle est satisfaisant. Pour résoudre le problème des conditions initiales nous adoptons la solution proposée par Wooldridge (2005) : nous maximisons la vraisemblance conditionnellement aux conditions initiales et nous laissons la loi de α_i en dépendre pour approximer la corrélation qui existe entre les deux. Plus précisément, nous maximisons

$$P\left(y_{iT}, ..., y_{i2}\, / y_{i1}, \hat{\delta}\right) = \sum_{l=1}^{L} \left\{ P(\alpha_i = \alpha^l / y_{i1}) \prod_{t=2}^{T} P(y_{it}/y_{it-1}, \alpha_i = \alpha^l, y_{i1}, \hat{\delta}) \right\} \tag{3.6}$$

ou l est l'un des L types et $\hat{\delta}$ représente l'estimation de δ de première étape. Il faut noter que la vraisemblance est conditionnelle à y_{i1} et que le terme $P(\alpha_i = \alpha^l/y_{i1})$ montre que la loi de α_i dépend de y_{i1}. Les paramètres à déterminer sont $\forall\ l = 1...L$ et $\forall\ k = 0...5$, α^l et $P(\alpha_i = \alpha^l/y_{i1} = k)$ (que nous notons π_{lk}). Cette vraisemblance est maximisée via un algorithme EM standard, en itérant les deux étapes suivantes :

- Etape E

Pour des valeurs initiales de l'étape $r-1$, $(\alpha_l)^{(r-1)}$ et $\pi_{lk}^{(r-1)}$, pour chaque type $l = 1...L$ et chaque individu i dans l'échantillon calculer la probabilité a posteriori w_{il} que l'individu i appartienne au type l :

$$P(\alpha_i = (\alpha^l)^{(r-1)}/y_{iT}, ..., y_{i2}, y_{i1}) = \frac{\pi_{l y_{i1}}^{(r-1)} \prod_{t=2}^{T} P(y_{it}/y_{it-1}, \alpha_i = (\alpha^l)^{(r-1)}, y_{i1})}{\sum_{j=1}^{L} \left\{ \pi_{j y_{i1}}^{(r-1)} \prod_{t=2}^{T} P(y_{it}/y_{it-1}, \alpha_i = (\alpha^j)^{(r-1)}, y_{i1}) \right\}}$$

- Etape M

Actualiser la loi de l'hétérogénéité inobservée en moyennant les probabilités a posteriori obtenues dans l'étape E : $\forall\ l = 1...L$, $k = 0...5$,

$$\pi_{lk}^{(r)} = \frac{\sum\limits_{i/y_{i1}=k} w_{il}}{\sum\limits_{i=1}^{N} \mathbb{1}_{\{y_{i1}=k\}}}$$

Enfin, actualiser le support de l'hétérogénéité inobservée : $\forall\ l = 1...L$,

$$\alpha_{il}^{(r)} = \underset{a}{\operatorname{argmax}} \sum_{i=1}^{N}\sum_{t=2}^{T} w_{il} \ln P(y_{it}/y_{it-1}, \alpha_i = a, y_{i1})$$

Cette dernière expression est la vraisemblance d'un logit multinomial pondérée par les probabilités a priori de l'étape E et est donc aisément maximisée.

L'estimation du modèle nous permet de caculer les matrices de transition spécifiques à chaque individu associées à chaque type d'individu estimé en utilisant (1.2). Pour caractériser chaque matrice de transition, il est intéressant de déterminer ce qui se passe lorsque le processus se déroule indéfiniment. Pour ce faire nous calculons pour chaque matrice les distributions d'équilibre stationnaire, qui ne se déforment plus d'années en années, c'est-à-dire les distributions de quintile Q telles que $Q = M'Q$, où M est la matrice de transition spécifique. Ainsi, pour chaque matrice de transition, la distribution d'équilibre stationnaire est le vecteur propre de norme 1 de la valeur propre 1 associée à la transposée de la matrice de transition[11]. Ces distributions nous montrent dans quelle zone les individus bougent à l'état stationnaire.

[11]La somme des éléments d'une ligne d'une matrice de transition valant 1, une telle matrice admet toujours une valeur propre valant 1.

3.4 Résultats

Pour commencer, rappelons que comme la méthode d'estimation semi-paramétrique utilisée ne permet pas d'inclure de variables explicatives qui varient dans le temps, les quintiles de salaires ne sont pas calculés à partir des salaires eux-mêmes mais avec les résidus de la régression des logs-salaires sur l'expérience et l'expérience au carré. Ainsi, nous contrôlons l'expérience de manière préalable et les résultats présentés dans cette section doivent être compris à niveau d'expérience donné.

3.4.1 Nombre de types et ajustement du modèle

Le nombre de points de support de l'hétérogénéité inobservée est choisi chaque décennie de sorte à maximiser la qualité de l'ajustement du modèle avec les matrices de transition observées (Table 3.1). Nous comparons les matrices de transition observées avec les matrices de transition simulées que nous calculons en utilisant les paramètres estimés et l'équation (3.2). Le processus est itératif dans le sens où nous commençons par estimer le modèle avec deux types, et si nous observons que l'ajustement n'est pas satisfaisant, nous essayons avec trois, etc. Nous arrêtons lorsque l'ajustement est satisfaisant.

Pour les années 70 et 80 l'ajustement est très bon avec 5 et 6 types. Mais dans les années 70, avec 5 types, l'un d'eux correspond à des individus qui passent l'essentiel de leur temps dans les quintiles 2 et 5, et le même phénomène survient dans les années 80 avec un type d'individus qui partagent leur temps entre les quintiles 1 et 5. Un tel type d'individu n'a aucune interprétation économique, et nous préférons donc pour les années 70 et 80 les modèles avec 6 types. Pour les années 90, il n'y a pas d'hésitation possible : l'ajustement du modèle à 6 types est très bon et meilleur que les autres. Les matrices de transition simulées sont présentées dans les Tables 3.2, 3.3 et 3.4. Elles doivent être comparées aux matrices de transition observées entre parenthèses. Il est clair que les ajustements de ces trois modèles sont très bons.

Tab. 3.2 – Qualité de l'ajustement du modèle, années 70

Quintile	Quintile d'arrivée						
d'origine	0	1	2	3	4	5	Total
0	85.2	9.2	2.5	1.9	0.7	0.5	100
	(83.7)	(9.8)	(3.1)	(2.0)	(0.8)	(0.6)	
1	19.8	56.1	20.4	3.7	0.0	0.0	100
	(19.9)	(57.0)	(19.7)	(3.4)	(0.0)	(0.0)	
2	7.9	12.3	53.8	21.9	4.1	0.0	100
	(9.1)	(12.0)	(54.4)	(20.6)	(3.9)	(0.0)	
3	5.9	3.7	15.0	50.0	20.1	5.3	100
	(5.4)	(3.3)	(14.9)	(50.8)	(20.0)	(5.7)	
4	4.0	0.0	3.7	15.9	58.1	18.3	100
	(3.6)	(0.0)	(3.7)	(17.3)	(57.8)	(17.7)	
5	2.4	0.0	0.0	5.2	18.9	73.4	100
	(2.4)	(0.0)	(0.0)	(4.3)	(16.5)	(76.7)	

Notes : PSID, 25-60 ans, salaires annuels. Entre parenthèses la matrice de transition observée. La matrice de transition simulée est calculée en utilisant les estimations des paramètres et l'équation (3.2).

3.4.2 Dépendance d'état

Les estimations des paramètres de dépendance d'état sont présentées dans les tables 3.5, 3.6 et 3.7. Pour interpréter les résultats, la mesure pertinente est la différence entre les δ's et non les δ's eux-mêmes. Par exemple, si $\delta_{21} \leq \delta_{22}$, un individu (fictif) tel que $\alpha_{i1} = \alpha_{i2}$[12] initialement dans le quintile 2 a une probabilité plus grande d'être stable plutôt que de descendre d'un quintile. Ainsi, la comparaison entre les δ's peut être interprétée comme une comparaison de probabilités pour des individus qui ont les mêmes propensions individuelles. En ce sens, les paramètres de dépendance d'état fournissent une mesure de la mobilité des salaires corrigée des caractéristiques individuelles. Le premier résultat des tables 3.5, 3.6 et 3.7 est que tous les coefficients sont positifs et significativement différents de zéro[13] : la dépendance d'état est telle que la transition vers le quintile 0 est la moins probable. Maintenant si nous nous concentrons sur les quintiles 1 à 5, il est intéressant de

[12] C'est-à-dire un individu qui a la même propension individuelle à être dans les quintiles 1 et 2.

[13] sauf δ_{31} dans les années 80 et δ_{53} dans les années 90 qui ne sont pas significativement différents de 0.

Tab. 3.3 – Qualité de l'ajustement du modèle, années 80

Quintile d'origine	Quintile d'arrivée 0	1	2	3	4	5	Total
0	77.1	11.9	5.7	3.4	1.3	0.7	100
	(74.8)	(13.0)	(6.2)	(3.4)	(1.6)	(1.0)	
1	24.2	52.0	20.1	3.7	0.0	0.0	100
	(20.9)	(53.5)	(20.9)	(4.7)	(0.0)	(0.0)	
2	10.7	14.7	50.4	20.4	3.7	0.0	100
	(10.6)	(14.1)	(49.6)	(21.7)	(4.1)	(0.0)	
3	9.4	2.4	14.4	51.6	19.9	2.3	100
	(9.3)	(2.9)	(13.9)	(50.3)	(21.0)	(2.5)	
4	4.7	0.0	3.6	15.5	58.1	18.0	100
	(6.0)	(0.0)	(3.0)	(15.5)	(57.1)	(18.4)	
5	2.9	0.0	0.0	3.4	14.1	79.6	100
	(3.8)	(0.0)	(0.0)	(3.1)	(14.6)	(78.5)	

Notes : PSID, 25-60 ans, salaires annuels. Entre parenthèses la matrice de transition observée. La matrice de transition simulée est calculée en utilisant les estimations des paramètres et l'équation (3.2).

Tab. 3.4 – Qualité de l'ajustement du modèle, années 90

Quintile d'origine	Quintile d'arrivée 0	1	2	3	4	5	Total
0	75.1	11.4	4.8	4.6	2.2	1.8	100
	(73.3)	(12.9)	(5.3)	(4.4)	(2.2)	(1.9)	
1	22.9	57.1	16.4	3.6	0.0	0.0	100
	(22.0)	(56.9)	(17.0)	(4.0)	(0.0)	(0.0)	
2	11.5	13.5	54.0	17.8	3.2	0.0	100
	(11.0)	(13.1)	(54.1)	(18.1)	(3.6)	(0.0)	
3	10.2	3.2	15.5	49.6	18.4	3.1	100
	(10.5)	(2.9)	(15.0)	(51.1)	(17.6)	(2.9)	
4	7.7	0.0	2.6	14.7	58.1	16.9	100
	(7.7)	(0.0)	(2.7)	(15.7)	(58.4)	(15.6)	
5	7.6	0.0	0.0	2.1	14.0	76.3	100
	(7.2)	(0.0)	(0.0)	(2.0)	(14.6)	(76.2)	

Notes : PSID, 25-60 ans, salaires annuels. Entre parenthèses la matrice de transition observée. La matrice de transition simulée est calculée en utilisant les estimations des paramètres et l'équation (3.2).

tester l'égalité des paramètres d'une même ligne. Ces tests (de Wald) sont présentés dans la table 3.8. Pour les trois décennies, l'hypothèse que les δ sont égaux sur toute la ligne est rejetée pour tous les quintiles d'origine[14]. Ce résultat montre que la présence de dépendance d'état ne peut pas être rejetée. Ce résultat est cohérent avec Weber (2002). Pour voir si la dépendance d'état vient de la mobilité ascendante ou descendante, nous réalisons des tests sur des sous-parties de chaque ligne. Deux résultats se dégagent de l'ensemble de ces tests. Premièrement, l'hypothèse que le paramètre δ_{jj} est égal aux paramètres δ_{jk}, $k < j$ est rejetée avec force. Deuxièmement, l'hypothèse que le paramètre δ_{jj} est égal au paramètres δ_{jk}, $k > j$ ne peut la plupart du temps pas être rejetée. Ces deux résultats montrent que la dépendance d'état rend la baisse moins fréquente que la stabilité alors qu'elle n'a pas d'influence sur le rapport entre la hausse et la stabilité. La dépendance d'état s'apparente ainsi à une rigidité à la baisse. Ceci fait écho au phénomène bien connu de rigidité des salaires à la baisse[15], expliqué par trois types de mécanisme du marché du travail. D'une part les comportements de prospection d'emploi des salariés font qu'ils ne changent volontairement de situation que lorsque ça leur est profitable. D'autre part, les modèles de salaire d'efficience montrent qu'en raison de problèmes d'information, les entreprises ne sont pas incitées à baisser les salaires, même en cas de choc négatif, car les salariés pourraient réduire leur effort ou même quitter l'entreprise, ce qui obligerait cette dernière à se lancer dans une coûteuse phase de recrutement, tant en termes de prospection que de formation. Enfin, il existe des facteurs institutionnels comme les négociations collectives ou les législations sur la protection de l'emploi.

[14] Avec un peu moins de force pour le quintile 1.

[15] Voir par exemple Agell et Benmarker (2007) pour une étude récente.

TAB. 3.5 – Paramètres de dépendance d'état, années 70

Quintile d'origine	Quintile d'arrivée 0	1	2	3	4	5
0	0	0	0	0	0	0
	–	–	–	–	–	–
1	0	1.70	2.32	1.53	$-\infty$	$-\infty$
	–	(0.21)	(0.29)	(0.41)	–	–
2	0	1.57	3.13	2.91	2.60	$-\infty$
	–	(0.29)	(0.34)	(0.37)	(0.48)	–
3	0	1.30	2.36	2.99	3.22	3.12
	–	(0.44)	(0.37)	(0.42)	(0.48)	(0.60)
4	0	$-\infty$	1.87	2.73	3.79	3.67
	–	–	(0.46)	(0.47)	(0.56)	(0.62)
5	0	$-\infty$	$-\infty$	2.99	3.76	4.22
	–	–	–	(0.62)	(0.68)	(0.79)

Notes : le chiffre à l'intersection de la ligne j et de la colonne k correspond à l'estimation de δ_{jk}. Par hypothèse, $\delta_{j0} = \delta_{0k} = 0$, et $\delta_{14} = \delta_{15} = \delta_{25} = \delta_{41} = \delta_{51} = \delta_{52} = -\infty$. Les écarts-types se lisent entre parenthèses.

TAB. 3.6 – Paramètres de dépendance d'état, années 80

Quintile d'origine	Quintile d'arrivée 0	1	2	3	4	5
0	0	0	0	0	0	0
	–	–	–	–	–	–
1	0	1.31	1.38	0.80	$-\infty$	$-\infty$
	–	(0.16)	(0.18)	(0.27)	–	–
2	0	1.19	2.17	1.74	1.81	$-\infty$
	–	(0.21)	(0.21)	(0.23)	(0.33)	–
3	0	-0.06	1.22	2.18	2.56	2.12
	–	(0.52)	(0.27)	(0.29)	(0.33)	(0.51)
4	0	$-\infty$	1.18	1.98	3.55	3.45
	–	–	(0.31)	(0.29)	(0.36)	(0.48)
5	0	$-\infty$	$-\infty$	2.04	3.01	4.51
	–	–	–	(0.41)	(0.42)	(0.57)

Notes : le chiffre à l'intersection de la ligne j et de la colonne k correspond à l'estimation de δ_{jk}. Par hypothèse, $\delta_{j0} = \delta_{0k} = 0$, et $\delta_{14} = \delta_{15} = \delta_{25} = \delta_{41} = \delta_{51} = \delta_{52} = -\infty$. Les écarts-types se lisent entre parenthèses.

Tab. 3.7 – Paramètres de dépendance d'état, années 90

Quintile d'origine	Quintile d'arrivée 0	1	2	3	4	5
0	0	0	0	0	0	0
	–	–	–	–	–	–
1	0	1.42	1.51	0.93	$-\infty$	$-\infty$
	–	(0.15)	(0.18)	(0.25)	–	–
2	0	1.27	2.53	1.97	1.97	$-\infty$
	–	(0.18)	(0.21)	(0.21)	(0.30)	–
3	0	0.73	1.82	2.23	2.49	2.23
	–	(0.25)	(0.21)	(0.22)	(0.26)	(0.36)
4	0	$-\infty$	1.10	1.52	2.75	2.62
	–	–	(0.29)	(0.24)	(0.29)	(0.32)
5	0	$-\infty$	$-\infty$	0.63	2.02	2.87
	–	–	–	(0.34)	(0.30)	(0.34)

Notes : le chiffre à l'intersection de la ligne j et de la colonne k correspond à l'estimation de δ_{jk}. Par hypothèse, $\delta_{j0} = \delta_{0k} = 0$, et $\delta_{14} = \delta_{15} = \delta_{25} = \delta_{41} = \delta_{51} = \delta_{52} = -\infty$. Les écarts-types se lisent entre parenthèses.

Tab. 3.8 – Tests de Wald de présence de dépendance d'état

H0	DF	Années 70 Wald Stat	Années 70 P value	Années 80 Wald Stat	Années 80 P value	Années 90 Wald Stat	Années 90 P value
$\delta_{11} = \delta_{12} = \delta_{13}$	2	6.441	0.040	4.694	0.096	5.112	0.078
$\delta_{11} = \delta_{12}$	1	4.800	0.028	0.107	0.743	0.216	0.642
$\delta_{21} = \delta_{22} = \delta_{23} = \delta_{24}$	3	31.877	0.000	25.103	0.000	43.749	0.000
$\delta_{22} = \delta_{23} = \delta_{24}$	2	1.359	0.507	4.126	0.127	7.823	0.020
$\delta_{22} = \delta_{23}$	1	0.427	0.514	3.997	0.046	7.231	0.007
$\delta_{22} = \delta_{21}$	1	30.484	0.000	24.925	0.000	43.741	0.000
$\delta_{31} = \delta_{32} = \delta_{33} = \delta_{34} = \delta_{35}$	4	16.900	0.002	43.218	0.000	39.701	0.000
$\delta_{33} = \delta_{34} = \delta_{35}$	2	0.316	0.854	2.426	0.297	1.500	0.472
$\delta_{33} = \delta_{34}$	1	0.316	0.574	2.004	0.157	1.318	0.251
$\delta_{31} = \delta_{32} = \delta_{33}$	2	14.176	0.001	35.050	0.000	33.566	0.000
$\delta_{32} = \delta_{33}$	1	3.716	0.054	18.887	0.000	3.899	0.048
$\delta_{42} = \delta_{43} = \delta_{44} = \delta_{45}$	3	16.338	0.001	49.700	0.000	39.560	0.000
$\delta_{44} = \delta_{45}$	1	0.058	0.809	0.057	0.811	0.196	0.658
$\delta_{42} = \delta_{43} = \delta_{44}$	2	15.044	0.001	47.498	0.000	36.189	0.000
$\delta_{43} = \delta_{44}$	1	6.084	0.014	30.194	0.000	27.541	0.000
$\delta_{53} = \delta_{54} = \delta_{55}$	2	5.298	0.071	23.384	0.000	36.544	0.000
$\delta_{54} = \delta_{55}$	1	0.738	0.390	11.818	0.001	7.963	0.005

Notes : PSID, 25-60 ans, salaires annuels.

3.4.3 Hétérogénéité inobservée

Avec les estimations des δ's, des α's et l'expression (3.2) des probabilités de transition déduites du modèle, nous calculons une matrice de transition spécifique à chaque type (ie pour chaque valeur discrète de α_i). Ces matrices sont présentées pour les années 70, 80 et 90 dans les tables 3.9, 3.10 et 3.11. Considérons par exemple la matrice de transition spécifique dans la partie inférieure gauche de la table 3.11. Ce type représente 18.8% de la population. Lorsque ces individus sont initialement dans le quintile 3, le quintile de destination le plus probable est le quintile 3 (59.9%). Quand ils sont initialement dans le quintile 2, la destination la plus probable est également de monter dans le quintile 3 (50.1%), et en fait pour tous les quintiles d'origine (sauf le quintile 0 qui est un quintile particulier), le quintile 3 est la destination la plus probable. Les individus de ce type peuvent être caractérisés par le fait qu'ils sont comme ''attirés" vers le quintile 3, quelque soit le quintile d'où ils partent. C'est pourquoi nous labélisons ce type le type 3. En fait, ce phénomène d'attraction est présent pour tous les types de la table 3.11. Par exemple, dans la matrice de la partie supérieure droite de la table, il est clair que la colonne correspondant au quintile 2 est le mode de chaque distribution conditionnelle[16], ce qui veut dire que cette matrice de transition attire vers le quintile 2 et peut donc être labélisée type 2. Ces résultats signifient que, dans les années 90, chaque individu, étant données ses caractéristiques individuelles fixes dans le temps, est attiré vers un quintile spécifique. Les tables 3.9 et 3.10 donnent exactement les mêmes interprétations pour les années 70 et 80. Ces résultats montrent que l'on peut associer à chaque type d'individu un quintile vers lequel il est sans cesse attiré.

Ces résultats ont des conséquences sur la manière de modéliser la dynamique des salaires. Comme nous l'avons expliqué dans la section 3.3.2, les modélisations de la dynamique des salaires basées sur des processus ARMA avec des chocs gaussiens[17] mènent à supposer

[16] Notons cependant que ceci n'est pas vrai pour le quintile d'origine 5. Cela vient du fait que par hypothèse, la transition 5-2 n'est pas autorisée ($\delta_{52} = -\infty$), et donc la probabilité de transition 5-2 est nulle. Le quintile de destination le plus probable est donc le quintile adjacent le quintile 3.

[17] Voir par exemple Burkhauser et al (1997) ou Moffit et Gottschalk (1998).

que les matrices de transition inter-quintiles conditionnelles sont symétriques. Or cette hypothèse est clairement rejetée par nos résultats. Il paraît donc indispensable de développer des modèles qui brisent cette symétrie, comme l'ont fait Meghir et Pistaferri (2004) et Browning et al (2006) en introduisant des effets ARCH ou Bonhomme et Robin (2008) en utilisant des modèles à copules.

Pour voir plus précisément les zones dans lesquelles les individus sont attirés, en d'autres termes les zones dans lesquelles les individus évoluent la plupart du temps à l'équilibre stationnaire, il est utile de calculer, pour chaque matrice de transition spécifique, les distributions d'équilibre stationnaire des quintiles[18]. Ces distributions sont présentées dans la table 3.12. Par exemple, pour les années 90, à l'équilibre stationnaire, nous voyons que 13.4% des individus du type 2 sont dans le quintile 0, et 12.6%, 60.3%, 13.2%, 0.4% et 0.0% dans les quintiles 1, 2, 3, 4 et 5 respectivement. Cela signifie que la majorité des individus du type 2 sont chaque année, à l'équilibre stationnaire, dans le quintile 2, ou en d'autres termes que les individus du type 2 passent la plupart de leur temps, à l'équilibre stationnaire, dans le quintile 2. La table 3.12 montre que ce résultat est vrai pour tous les types, et pour les trois décennies. Ceci confirme que la nature de l'hétérogénéité inobservée est d'attirer les individus vers un quintile spécifique. Ainsi, la présence d'hétérogénéité dans la mobilité des salaires se traduit par une segmentation de la population. La signification de cette segmentation est que chaque individu passe la plupart du temps dans le même quintile, en gravitant autour de temps en temps.

Pour donner une interprétation économique aux types estimés, nous présentons dans la Table 3.13 le croisement des types inobservés avec le sexe, la race (blanc/non-blanc) et l'éducation (Diplômé de l'enseignement supérieur ou non). On y voit sans ambiguités que

[18]Voir section 3.3.4.

Tab. 3.9 – Matrices de transitions par type, années 70

Quintile	Dest Q, Type 0 (30.1%)							Dest Q, Type 1 (13.0%)							Dest Q, Type 2 (14.4%)						
d'origine	0	1	2	3	4	5	Total	0	1	2	3	4	5	Total	0	1	2	3	4	5	Total
0	94.5	4.3	0.2	0.7	0.1	0.2	100.0	55.3	40.8	2.8	0.6	0.0	0.5	100.0	52.4	20.1	22.8	4.3	0.4	0.0	100.0
1	76.3	19.2	1.9	2.7	0.0	0.0	100.0	17.8	72.2	9.2	0.8	0.0	0.0	100.0	12.6	26.5	56.0	4.8	0.0	0.0	100.0
2	70.3	15.5	4.0	9.8	0.6	0.0	100.0	17.0	60.3	19.6	3.2	0.0	0.0	100.0	6.9	12.8	69.0	10.5	0.8	0.0	100.0
3	71.5	12.0	1.9	10.7	1.1	2.9	100.0	21.6	58.3	11.6	4.4	0.0	4.1	100.0	11.3	15.8	52.1	18.3	2.4	0.0	100.0
4	81.4	0.0	1.3	9.4	2.1	5.7	100.0	55.0	0.0	18.0	8.7	0.1	18.2	100.0	18.3	0.0	51.7	23.0	6.9	0.0	100.0
5	77.1	0.0	0.0	11.5	2.0	9.4	100.0	56.1	0.0	0.0	11.4	0.1	32.4	100.0	33.4	0.0	0.0	54.3	12.1	0.1	100.0

Quintile	Dest Q, Type 3 (15.1%)							Dest Q, Type 4 (17.5%)							Dest Q, Type 5 (9.9%)						
d'origine	0	1	2	3	4	5	Total	0	1	2	3	4	5	Total	0	1	2	3	4	5	Total
0	41.3	7.4	13.4	31.9	5.5	0.6	100.0	50.9	3.5	6.0	10.5	22.1	7.0	100.0	46.0	0.0	9.6	3.3	3.7	37.4	100.0
1	11.3	11.1	37.4	40.3	0.0	0.0	100.0	28.3	10.8	33.9	27.0	0.0	0.0	100.0	28.8	0.1	61.5	9.6	0.0	0.0	100.0
2	4.0	3.4	29.3	56.2	7.1	0.0	100.0	7.3	2.4	19.7	27.8	42.7	0.0	100.0	12.2	0.0	58.3	16.1	13.3	0.0	100.0
3	4.1	2.7	14.3	63.6	13.9	1.3	100.0	4.9	1.2	6.1	19.9	52.9	15.1	100.0	4.0	0.0	8.9	5.7	8.1	73.3	100.0
4	4.7	0.0	9.8	55.3	27.6	2.6	100.0	3.4	0.0	2.6	10.7	65.2	18.2	100.0	2.6	0.0	3.5	2.8	9.3	81.8	100.0
5	4.3	0.0	0.0	66.5	24.9	4.2	100.0	3.0	0.0	0.0	12.3	56.4	28.3	100.0	1.6	0.0	0.0	2.3	5.7	90.4	100.0

Notes : PSID, 25-60 ans, salaires annuels.

Tab. 3.10 – Matrices de transitions par type, années 80

Quintile	Dest Q, Type 0 (25.7%)							Dest Q, Type 1 (9.1%)							Dest Q, Type 2 (19.0%)						
d'origine	0	1	2	3	4	5	Total	0	1	2	3	4	5	Total	0	1	2	3	4	5	Total
0	90.7	7.3	1.4	0.4	0.1	0.1	100.0	41.5	52.6	5.4	0.4	0.1	0.0	100.0	39.0	22.1	28.8	8.9	0.7	0.5	100.0
1	73.0	21.9	4.4	0.7	0.0	0.0	100.0	16.0	75.4	8.2	0.3	0.0	0.0	100.0	15.3	32.3	44.7	7.8	0.0	0.0	100.0
2	70.0	18.5	9.4	1.8	0.3	0.0	100.0	15.7	65.3	17.9	0.8	0.3	0.0	100.0	9.3	17.3	60.3	12.1	1.0	0.0	100.0
3	84.4	6.4	4.4	3.4	0.7	0.8	100.0	36.4	43.4	16.0	2.8	1.3	0.0	100.0	15.7	8.4	39.3	31.5	3.5	1.6	100.0
4	87.8	0.0	4.4	2.9	1.9	3.1	100.0	63.0	0.0	26.6	4.0	6.3	0.0	100.0	16.5	0.0	39.7	27.3	10.0	6.5	100.0
5	86.9	0.0	0.0	3.0	1.1	8.9	100.0	88.7	0.0	0.0	6.0	5.1	0.1	100.0	23.5	0.0	0.0	41.3	8.3	26.8	100.0

Quintile	Dest Q, Type 3 (16.3%)							Dest Q, Type 4 (16.7%)							Dest Q, Type 5 (13.2%)						
d'origine	0	1	2	3	4	5	Total	0	1	2	3	4	5	Total	0	1	2	3	4	5	Total
0	34.2	4.1	19.6	36.4	5.6	0.1	100.0	37.0	18.2	5.1	15.8	19.4	4.5	100.0	56.8	0.1	6.1	1.7	10.5	24.9	100.0
1	16.4	7.3	37.3	38.9	0.0	0.0	100.0	23.1	42.4	12.6	21.9	0.0	0.0	100.0	66.9	0.3	28.3	4.5	0.0	0.0	100.0
2	7.4	2.9	37.3	44.9	7.4	0.0	100.0	10.6	17.1	12.8	25.7	33.8	0.0	100.0	30.8	0.1	28.9	5.3	34.9	0.0	100.0
3	6.8	0.8	13.3	64.2	14.7	0.2	100.0	7.4	3.4	3.5	27.8	50.4	7.5	100.0	13.0	0.0	4.7	3.4	31.4	47.3	100.0
4	6.1	0.0	11.4	47.0	34.9	0.6	100.0	3.8	0.0	1.7	11.6	68.5	14.5	100.0	4.6	0.0	1.6	1.0	29.6	63.2	100.0
5	7.8	0.0	0.0	63.9	26.2	2.1	100.0	3.8	0.0	0.0	12.6	40.9	42.7	100.0	2.2	0.0	0.0	0.5	8.4	88.8	100.0

Notes : PSID, 25-60 ans, salaires annuels.

Tab. 3.11 – Matrices de transitions par type, années 90

Quintile	Dest Q, Type 0 (19.1%)							Dest Q, Type 1 (17.9%)							Dest Q, Type 2 (14.7%)						
d'origine	0	1	2	3	4	5	Total	0	1	2	3	4	5	Total	0	1	2	3	4	5	Total
0	93.5	4.1	1.4	0.4	0.0	0.5	100.0	51.8	41.0	5.7	1.0	0.5	0.0	100.0	41.5	16.4	32.0	9.8	0.3	0.0	100.0
1	79.3	14.5	5.3	1.0	0.0	0.0	100.0	20.7	68.0	10.3	1.0	0.0	0.0	100.0	14.9	24.4	51.8	8.9	0.0	0.0	100.0
2	72.7	11.3	13.5	2.5	0.1	0.0	100.0	18.5	52.0	25.6	2.6	1.3	0.0	100.0	7.3	10.2	69.9	12.3	0.4	0.0	100.0
3	78.2	7.1	7.2	3.4	0.1	4.0	100.0	27.5	45.3	18.8	5.1	3.2	0.0	100.0	11.3	9.3	53.7	24.7	0.9	0.0	100.0
4	87.5	0.0	3.9	1.9	0.1	6.6	100.0	63.5	0.0	21.1	5.8	9.5	0.0	100.0	22.2	0.0	51.3	24.0	2.4	0.1	100.0
5	90.3	0.0	0.0	0.8	0.1	8.8	100.0	90.1	0.0	0.0	3.4	6.5	0.0	100.0	66.8	0.0	0.0	29.6	3.5	0.2	100.0
Quintile	Dest Q, Type 3 (18.8%)							Dest Q, Type 4 (15.4%)							Dest Q, Type 5 (14.1%)						
d'origine	0	1	2	3	4	5	Total	0	1	2	3	4	5	Total	0	1	2	3	4	5	Total
0	44.9	5.2	10.0	32.7	6.3	0.8	100.0	39.1	3.2	2.8	15.1	32.8	7.0	100.0	47.3	0.7	2.4	3.2	7.7	38.7	100.0
1	23.0	11.2	23.2	42.6	0.0	0.0	100.0	37.9	12.8	12.1	37.3	0.0	0.0	100.0	68.5	4.3	15.4	11.9	0.0	0.0	100.0
2	9.6	4.0	26.7	50.1	9.7	0.0	100.0	9.1	2.6	8.1	25.4	54.8	0.0	100.0	30.0	1.6	18.7	14.7	35.1	0.0	100.0
3	8.9	2.2	12.3	59.9	15.2	1.6	100.0	5.9	1.0	2.6	21.1	59.8	9.7	100.0	8.7	0.3	2.7	5.5	17.2	65.7	100.0
4	13.4	0.0	9.0	44.6	29.6	3.5	100.0	5.4	0.0	1.1	9.5	70.8	13.1	100.0	6.6	0.0	1.0	2.1	16.8	73.6	100.0
5	26.6	0.0	0.0	36.3	28.2	8.9	100.0	8.9	0.0	0.0	6.5	56.4	28.2	100.0	6.0	0.0	0.0	0.8	7.3	86.0	100.0

Notes : PSID, 25-60 ans, salaires annuels.

Tab. 3.12 – Distributions stationnaires des quintiles, par type

	Quintile						
Années 70							
Type	0	1	2	3	4	5	Total
0	93.2	5.2	0.4	1.0	0.1	0.2	100
1	28.6	61.9	8.2	1.0	0.0	0.3	100
2	15.4	16.3	58.1	9.4	0.8	0.0	100
3	7.0	3.0	16.7	58.4	13.8	1.2	100
4	7.2	0.6	3.6	12.9	57.7	18.0	100
5	3.5	0.0	1.9	2.7	6.1	85.8	100
Années 80							
Type	0	1	2	3	4	5	Total
0	88.8	8.8	1.8	0.5	0.1	0.1	100
1	21.6	69.5	8.5	0.4	0.1	0.0	100
2	16.4	19.5	48.6	13.7	1.2	0.5	100
3	9.6	1.5	18.4	55.0	15.4	0.2	100
4	8.0	4.3	2.7	15.4	53.4	16.2	100
5	6.2	0.0	0.8	0.7	11.3	80.9	100
Années 90							
Type	0	1	2	3	4	5	Total
0	92.4	4.7	1.8	0.5	0.0	0.5	100
1	30.0	57.6	10.8	1.3	0.4	0.0	100
2	13.4	12.6	60.3	13.2	0.4	0.0	100
3	16.1	2.8	13.5	51.0	15.0	1.6	100
4	9.3	0.5	1.4	11.3	63.4	14.0	100
5	10.6	0.1	0.5	1.3	8.4	79.2	100

Notes : PSID, 25-60 ans, salaires annuels.

les hommes, les blancs et les diplômés d'université sont sur-représentés dans les types de numéro élevé et sous-représentés dans ceux de numéro faible. Cela signifie que les hommes, les blancs et les plus éduqués bougent plus souvent dans le haut de la distribution des quintiles alors que les femmes, les non-blancs et les moins éduqués dans le bas. Ainsi, cette segmentation reflète les investissements initiaux en capital humain qui se matérialisent par le biais de l'affectation des individus aux emplois en fonction de leur éducation (Rosen (1982), Waldman (1984)). L'éducation des individus les affecte à un certain quintile, autour duquel ils gravitent en fonction des chocs de productivité et des opportunités de changements d'entreprises, de secteur ou encore d'occupation.

3.4.4 Exploitation du modèle

L'estimation du modèle permet de répondre à deux questions supplémentaires. Ayant distingué la dépendance d'état de l'hétérogénéité inobservée dans la mobilité des salaires, on peut légitimement se demander quelles sont les contributions respectives de ces deux composantes. D'autre part, ayant identifié des groupes d'individus, on peut se demander comme Buchinsky et Hunt (1999) dans quelles proportions l'effet égalisateur de la mobilité salariale agit sur les composantes inter-groupes et intra-groupes de l'inégalité salariale.

Quelles sont les contributions respectives de la dépendance d'état et de l'hétérogénéité inobservée dans la stabilité des salaires ?

Les tables 3.9, 3.10 et 3.11 montrent que pour chaque type un seul quintile peut être considéré comme stable, le quintile autour duquel les individus gravitent. Par exemple dans le type 3 des années 90 (table 3.11), les pourcentages de stabilité valent, du quintile 0 à 5, 11.2%, 26.7%, 59.9%, 29.6% et 8.9% : il est clair que ce type est le groupe de stabilité du quintile 3. Ainsi, l'idée est d'étudier la stabilité de chaque quintile au sein de son type associé, c'est-à-dire la stabilité dans le quintile k parmi les individus du type k. Pour calculer

Tab. 3.13 – Corrélation entre les types inobservés et les caractéristiques observables

	Type						
Années 70							
	0	1	2	3	4	5	Total
Femme	53.2	20.0	15.3	7.3	3.6	0.5	100.0
Homme	2.2	4.6	13.3	24.5	34.2	21.2	100.0
Non-blanc	29.3	23.4	17.6	13.2	12.7	3.9	100.0
Blanc	30.2	11.8	14.1	15.3	18.0	10.6	100.0
Lycée	34.3	17.8	15.4	14.3	13.8	4.4	100.0
Université	23.3	5.3	12.9	16.4	23.3	18.8	100.0
Années 80							
	0	1	2	3	4	5	Total
Femme	44.4	12.3	21.6	12.7	7.0	2.0	100.0
Homme	3.6	5.3	16.0	20.6	28.1	26.4	100.0
Non-blanc	26.7	18.2	23.3	16.1	11.9	3.8	100.0
Blanc	25.6	8.2	18.6	16.3	17.2	14.1	100.0
Lycée	30.7	14.4	22.6	14.0	13.1	5.3	100.0
Université	21.3	4.3	15.9	18.4	19.9	20.2	100.0
Années 90							
	0	1	2	3	4	5	Total
Femme	30.4	25.2	15.4	16.5	7.8	4.7	100.0
Homme	5.4	9.2	13.7	21.5	24.6	25.5	100.0
Non-blanc	21.1	26.7	16.3	17.8	14.1	4.1	100.0
Blanc	18.9	17.1	14.5	18.9	15.5	15.1	100.0
Lyceé	20.2	27.1	19.2	17.3	11.1	5.1	100.0
Université	18.3	11.5	11.5	19.8	18.4	20.5	100.0

Notes : PSID, 25-60 ans, salaires annuels.

les contributions respectives de la dépendance d'état et de l'hétérogénéité inobservée dans la stabilité des salaires, le logarithme de l'équation (3.3)

$$\ln\left\{\frac{P(y_{it}=k\,/y_{it-1}=k\,,\,\text{Type}_i=k)}{P(y_{it}=j\,/y_{it-1}=k\,,\,\text{Type}_i=k)}\right\}=(\delta_{kk}-\delta_{kj})+(\alpha_k^k-\alpha_j^k) \qquad (3.7)$$

avec $j \neq k$ est utilisé[19]. Cette dernière expression représente la mesure dans laquelle les individus du type k ont plus de chances de rester dans le quintile k plutôt que de bouger dans le quintile j. Elle est égale à la somme de $(\delta_{kk}-\delta_{kj})$, un terme de dépendance d'état, et de $(\alpha_k^k-\alpha_j^k)$, un terme d'hétérogénéité inobservée. La part S_{kj} de la dépendance d'état dans la stabilité dans le quintile k (par rapport au mouvement dans le quintile j) peut ainsi être mesurée par

$$S_{kj}=\frac{(\delta_{kk}-\delta_{kj})}{(\delta_{kk}-\delta_{kj})+(\alpha_k^k-\alpha_j^k)}$$

Comme on l'a vu, la dépendance d'état agit dans la plupart des cas comme une rigidité à la baisse, et donc en général elle n'explique pas pourquoi la stabilité dans le quintile de référence est plus fréquente que la mobilité ascendante. Ceci est donc entièrement dû aux caractéristiques individuelles comme le sexe, la race ou l'éducation, ici capturées par l'hétérogénéité inobservée. Par contre, comme le montre la table 3.14, la dépendance d'état explique en général entre 25% et 45% de l'écart entre stabilité et mobilité descendante selon les quintiles d'origine et la période considérée. Les mécanismes du marché du travail ne sont donc pas un frein à la mobilité ascendante et contribuent pour une part importante à diminuer les risques de mobilité descendante.

[19] α_j^k représente la composante j du vecteur α_i pour les individus du type k.

Tab. 3.14 – Part de la dépendance d'état dans le logarithme du ratio stabilité/baisse

Quintile	Baisse d'un quintile			Baisse de deux quintiles		
d'origine	Années 70	Années 80	Années 90	Années 70	Années 80	Années 90
2	36.7	30.2	39.1	–	–	–
3	27.2	44.8	25.5	28.3	31.0	28.8
4	37.7	45.7	42.9	32.8	38.7	26.0
5	28.6	50.8	35.6	23.6	29.6	38.1

Notes : PSID, 25-60 ans, salaires annuels.

La mobilité réduit-elle les inégalités intra-groupes autant que les inégalités inter-groupes ?

Une des caractéristiques fondamentales de la mobilité des salaires est d'égaliser les revenus de long terme. Par là, nous voulons dire que plus les individus bougent dans l'échelle des salaires moins grandes sont les inégalités de salaires moyens sur le cycle de vie. Prenons un exemple très simple pour illustrer cette idée. Supposons qu'il y a deux individus, 1 et 2, deux périodes, 1 et 2, qu'il existe deux salaires $\underline{w}$ and $\overline{w}$ avec $\underline{w} < \overline{w}$, et que ces deux salaires restent constants entre les deux périodes. Supposons que lors de la première période les individus 1 et 2 reçoivent respectivement $\underline{w}$ et $\overline{w}$. Deux cas sont alors possibles : soit il n'y a aucune mobilité salariale entre les deux périodes, auquel cas les individus conservent leur rang et leur salaire en période 2 ; soit il existe de la mobilité salariale, auquel cas les individus échangent leur salaire entre les deux périodes. Dans le premier cas, sans mobilité salariale, le salaire moyen de l'individu 1 sur les deux périodes vaut $\underline{w}$ et celui de l'individu 2 vaut $\overline{w}$. Dans le deuxième cas, avec mobilité salariale, le salaire moyen sur les deux périodes des deux individus est égal et vaut $\frac{\underline{w}+\overline{w}}{2}$, de sorte que la dispersion des salaires moyens sur les deux périodes est nulle. Il est donc clair que la distribution des salaires moyens sur les deux périodes est moins dispersée s'il y a de la mobilité.

Dans la mesure où nous avons identifié des groupes d'individus, il est intéressant de se demander comme Buchinsky et Hunt (1999) si la mobilité réduit plus les inégalités entre groupes ou les inégalités au sein de chaque groupe. Pour cela nous utilisons la méthode de

Buchinsky et Hunt (1999). Nous utilisons deux indices d'inégalité décomposables en des composantes inter et intra groupes. Comme ces deux auteurs nous considérons d'une part un indice de la famille des mesures d'entropie généralisée

$$I_c = \frac{1}{c(c-1)} \sum_{i=1}^{n} z_i \left[\left(\frac{w_i}{\sum_{j=1}^{n} z_j w_j} \right)^c - 1 \right], \quad c \neq 0, 1$$

où w_i et z_i sont respectivement le salaire et le poids de l'individu i, et le paramètre c indique à quelle queue de distribution l'indice doit être le plus sensible (plus c est élevé plus l'indice est sensible à l'inégalité dans la partie haute de la distribution). Nous choisissons comme Buchinsky et Hunt c=2, de sorte que notre premier indice d'inégalité est égal au carré du coefficient de variation divisé par deux. Si l'on suppose qu'il existe K groupes d'individus, cet indice peut se décomposer en une composante intra et une composante inter selon la formule

$$\begin{aligned} I_c &= I_c^W + I_c^B \\ &= \sum_k v_k \left(\frac{\overline{w}_k}{\overline{w}} \right)^c I_c^k + \sum_k v_k \left[\left(\frac{\overline{w}_k}{\overline{w}} \right)^c - 1 \right] \end{aligned}$$

où $\overline{w}_k$ et v_k sont la moyenne et le poids du groupe k, $\overline{w}$ le salaire moyen dans la population et I_c^k l'indice d'inégalité dans le kième groupe. Le deuxième indice que nous utilisons est tout simplement la variance, qui peut être décomposée en la somme d'une variance inter et d'une variance intra-groupes. Muni de nos indices d'inégalités, nous définissons l'indice de mobilité par

$$M = 1 - \frac{I\left(\frac{1}{T} \sum_t w_t \right)}{\sum_t \eta_t I(w_t)}$$

où w_t représente le vecteur des salaires individuels de la date t et η_t indique la part des salaires gagnés l'année t dans le total des salaires gagnés sur les T périodes. Ainsi, cet indice de mobilité mesure la proportion dans laquelle l'inégalité des salaires moyens est inférieure à la moyenne des inégalités. On peut réécrire cette expression sous la forme d'une décomposition en mobilité inter et intra-groupes, pondérées par la part de l'inégalité inter et intra-groupes dans la variance totale, S_T^B et S_T^W :

$$M = \left[1 - \frac{I^B\left(\frac{1}{T}\sum_t w_t\right)}{\sum_t \eta_t I^B(w_t)}\right]\frac{\sum_t \eta_t I^B(w_t)}{\sum_t \eta_t I(w_t)} + \left[1 - \frac{I^W\left(\frac{1}{T}\sum_t w_t\right)}{\sum_t \eta_t I^W(w_t)}\right]\frac{\sum_t \eta_t I^W(w_t)}{\sum_t \eta_t I(w_t)}$$
$$= M^B S_T^B + M^W S_T^W$$

Les résultats de cette décomposition sont présentés dans la table 3.15. On remarque tout d'abord que la mobilité réduit de l'ordre de 13% les inégalités de long terme, et que cette intensité a très peu évolué entre 1970 et 2000. Ce niveau est cohérent avec les résultats de Buchinsky et Hunt (1999). On voit ensuite que la part de l'inégalité intra-groupes représente autour de 25% de l'inégalité totale. Ce résultat est très différent de celui trouvé dans Buchinsky et Hunt (1999), qui attribuent environ cette part à l'inégalité inter-groupes. Cette différence vient du fait que nos groupes d'individus ne sont pas définis de la même manière. Leurs groupes sont basés sur le croisement de caractéristiques uniquement observables comme le sexe, la race, l'éducation et l'expérience, alors que les nôtres sont formés sur la base de caractéristiques inobservables. Il est donc normal que notre typologie explique beaucoup plus la variance totale des salaires. Concernant la mobilité, nous montrons qu'elle réduit de moins d'un pourcent les inégalités entre groupes, alors qu'elle réduit de 50% à 60% les inégalités au sein de chaque groupe. Nous partageons avec Buchinsky et Hunt (1999) le résultat que la mobilité est principalement intra-groupes, mais nous estimons un degré de mobilité intra-groupe bien supérieur, puisqu'ils trouvent une mobilité intra-groupe aux alentours de 25%. Bien que différents, ces résultats sont compatibles car la part de la mobilité intra-groupes est beaucoup plus faible dans notre cas que celle de Buchinsky et

Tab. 3.15 – Effet égalisateur de la mobilité salariale

		Avec l'indice d'inégalité basé sur le coefficient de variation	Avec l'indice d'inégalité égal à la variance des salaires
Années 70	Part de l'inégalité intra dans l'inégalité totale	25.6	22.0
	Mobilite totale	13.5	13.5
	Mobilite intra	52.7	60.8
	Mobilite inter	0.06	0.20
Années 80	Part de l'inégalité intra dans l'inégalité totale	26.8	22.3
	Mobilite totale	13.9	15.0
	Mobilite intra	50.7	58.8
	Mobilite inter	0.5	2.6
Années 90	Part de l'inégalité intra dans l'inégalité totale	27.1	19.1
	Mobilite totale	13.3	11.7
	Mobilite intra	48.4	58.2
	Mobilite inter	0.2	0.7

Notes : PSID, 25-60 ans, salaires annuels.

Hunt (1999). Notre méthode économétrique nous permet donc de trouver une partition de la population pour laquelle la mobilité intra-groupe est très importante.

3.5 Conclusion

Dans ce chapitre nous étudions la structure de la mobilité des salaires aux Etats-Unis entre 1970 et 2000. L'idée est de décomposer la matrice de transition inter-quintiles inconditionnelle en une somme pondérée de matrices spécifiques à chaque individu et de décrire leur nature. Pour réaliser cette décomposition, nous introduisons de l'hétérogénéité inobservée dans les matrices de transition inter-quintiles. Nous choisissons de modéliser directement la dynamique des quintiles de salaire plutôt que la dynamique des salaires eux-mêmes car cette réduction de l'objet à modéliser permet d'affaiblir les hypothèses paramétriques en hypothèses semi-paramétriques. Ainsi, nous modélisons la dynamique des quintiles de salaire par un logit multinomial avec hétérogénéité inobservée, qui nous permet de distinguer la dépendance d'état des caractéristiques individuelles constantes dans le temps. Dans une première étape, pour que les estimations soient robustes à toute spécification de l'hétérogénéité inobservée, les paramètres de dépendance d'état sont estimés par la technique semi-paramétrique du maximum de vraisemblance conditionnel. Dans la deuxième étape, le support de l'hétérogénéité inobservée est supposé discret, le problème des conditions initiales est pris en compte et la vraisemblance est maximisée via un algorithme EM standard.

L'estimation des paramètres de dépendance d'état montre que cette dernière s'apparente à une rigidité à la baisse, qui peut s'expliquer par certains mécanismes du marché du travail comme les comportements de prospection d'emploi ou les problèmes d'information asymétrique entre employés et employeurs. Cette rigidité à la baisse explique entre 25% et 45% de l'écart entre la stabilité et la mobilité descendante. Ensuite l'estimation du modèle entier montre, en identifiant 6 types d'individus, que l'hétérogénéité inobservée joue un rôle crucial dans la mobilité des salaires. Les matrices de transition spécifiques estimées ne sont pas symétriques, ce qui nous permet de rejeter avec force les spécifications gaussiennes ARMA de la dynamique des salaires. La nature des matrices de transition spécifiques est ''d'attirer" chaque individu vers un quintile particulier, et donc la traduction de cette hétérogénéité est une segmentation de la population. Nous montrons que cette segmentation est très corrélée avec le sexe, la race et l'éducation. Enfin, nous montrons que la mobilité des salaires ne réduit quasiment pas l'inégalité de long terme inter-groupe, alors qu'elle réduit d'environ 50% l'inégalité intra-groupe.

Conclusion générale

Cette thèse présente trois essais en microéconométrie de l'éducation. Nous nous proposons de suivre un individu dans le long parcours qui le mène du système éducatif au marché du travail. Dans le premier chapitre, nous étudions quelques propriétés de la fonction de production éducative dans l'enseignement supérieur ; nous travaillons dans le deuxième chapitre sur les problèmes d'information que peuvent rencontrer les étudiants au moment de décider de la poursuite de leurs études ; nous nous intéressons enfin dans le troisième chapitre aux effets des investissements en capital humain en termes de mobilité salariale.

Le premier chapitre permet de mettre en évidence qu'il existe, dans le cadre particulier de l'université étudiée, des effets de pairs et des effets enseignants de même amplitude. Plus précisément, nous trouvons que les moins bons étudiants bénéficient de la présence dans leur groupe des meilleurs étudiants, sans que cela nuise à ces derniers. Des simulations montrent qu'instituer des groupes de niveau ferait chuter de manière substantielle la note moyenne des moins bons étudiants. Au-delà des informations intéressantes que nous apportent ces données originales dans le cadre d'une littérature naissante, nous pensons que ce premier chapitre permet deux avancées significatives. D'une part l'emploi de l'Analyse des Correspondances Multiples permet de créer un indice synthétique de qualité des étudiants sur la base de nombreuses caractéristiques qualitatives ; la pertinence de cet indice est démontrée par le fait qu'il donne les mêmes résultats que l'autre indice utilisé, basé simplement sur la note totale des étudiants en première année. D'autre part, nous ne nous contentons pas de décrire le processus d'affectation des étudiants dans les groupes pour justifier son caractère aléatoire, mais nous proposons un moyen de le tester. Ce test a une portée bien plus générale que nos données et pourrait être réutilisé dans bien d'autres situations.

Le deuxième chapitre se place dans le cadre des modèles structurels dynamiques de choix d'éducation. Nous participons à remettre en cause l'idée selon laquelle les individus au moment de faire leur choix d'éducation, connaissent parfaitement leurs futures caractéristiques sur le marché du travail alors même qu'il n'ont peut-être encore jamais travaillé. Nous proposons deux pistes pour améliorer la modélisation de cette idée. D'une part nous supposons qu'au moment de faire leur choix les individus connaissent toujours parfaitement leur aptitude à l'école, qu'ils connaissent la corrélation objective qui existe entre cette dernière et les caractéristiques des individus sur le marché du travail et qu'ils utilisent cette corrélation pour faire une prévision de leurs futures caractéristiques. D'autre part, puisque l'information n'est pas parfaite, nous suggérons de modéliser la façon dont elle est apprise. Nous supposons ainsi qu'à chaque passage sur le marché du travail, les individus actualisent de manière bayésienne leurs croyances grâce à l'observation en fin de période de leur salaire, reflet de leur productivité. L'estimation du modèle structurel, par maximum de vraisemblance, rejette l'hypothèse d'information parfaite et permet de montrer qu'environ 20% des individus feraient des choix différents s'ils connaissaient toujours parfaitement leurs futures caractéristiques sur le marché du travail. Les différences de niveau d'éducation ne seraient cependant en général que d'une année d'études en plus ou en moins.

Le dernier chapitre nous permet de proposer une description de la structure de la mobilité des salaires basée sur l'hétérogénéité inobservée. Grâce à l'emploi de la méthode semi-paramétrique du maximum de vraisemblance conditionnel, nous avons la certitude que l'estimation des paramètres de dépendance d'état du modèle est robuste à toute spécification de l'hétérogénéité inobservée. Plutôt que de conclure à des niveaux de mobilité différents entre groupes d'individus, nos estimations mettent à jour une structure gravitationnelle de la mobilité salariale : chaque type d'individus gravite autour d'un quintile particulier, quintile qui dépend des caractéristiques de chaque individu comme le sexe, la race ou le niveau d'éducation. Cette structure se traduit par une segmentation de la distribution des quintiles de salaires. Ces résultats ont deux conséquences importantes. Tout d'abord, le caractère asymétrique des matrices de transition conditionnelles estimées permet de rejeter les modélisations ARMA standard gaussiennes de la dynamique des salaires. Ensuite, la typologie mise à jour permet de confirmer que la mobilité salariale ne réduit quasiment pas les inégalités salariales de long terme entre groupes, mais qu'au sein de chaque groupe elle les réduit plus ce qui avait été précédemment estimé sur la base de caractéristiques observables uniquement.

Bien sûr, ces trois chapitres ont des limites qui constituent autant de pistes de recherches futures. La principale limite du premier chapitre est que la qualité des étudiants n'est envisagée que sous l'angle de caractéristiques observables. On ne peut donc pas rejeter a priori la présence d'erreur de mesure. Il faudrait ainsi trouver une façon d'identifier et d'estimer les effets de pairs sur la base d'inobservables. Dans le deuxième chapitre, une des hypothèses centrales est que les individus connaissent toujours parfaitement leur goût et leur aptitude à l'école, supposés constant dans le temps. Mais il se peut qu'il existe également de l'apprentissage vis-à-vis de cette caractéristique : bien souvent l'aptitude des étudiants n'est révélée qu'au fur et à mesure du parcours scolaire lorsque le niveau d'études s'élève. Il serait ainsi intéressant d'intégrer également dans le modèle de l'apprentissage sur le goût et l'aptitude à l'école, avec la nouvelle difficulté de déterminer comment les individus définissent leur a priori sur cette nouvelle dimension. Enfin, dans le chapitre 3, la petite taille de notre échantillon nous a contraint à ne pas pouvoir étudier directement le rôle de l'expérience dans la mobilité. Or il existe aujourd'hui des données américaines administratives dont la grande taille donnerait la possibilité de mettre en oeuvre des techniques semi-paramétriques qui autorisent la présence de variables explicatives qui varient dans le temps.

Bibliographie

Agell, J. and H. Benmarker (2007). Wage Incentives and Wage Rigidity : A Representative View from Within, *Labour Economics*, 14, 347-369.

Arcidiaccono et al. (2007), Estimating Spillovers using panel data, with an application to the classroom, mimeo.

Arnott and Rouse (1987), Peer group effects and educational attainment, *Journal of Public Economics*, 32, 287-305.

Arrow, K. (1973). Higher Education as a Filter, *Journal of Public Economics*, 2, 193-216.

Atkinson, A., Bourguignon, F. and Morisson, C. (1992). Empirical Studies of Earnings Mobility, *Philadelphia : Harwood Academic Publisher*.

Becker, G. (1964). Human Capital. A theoritical and Empirical Analysis, with Special Reference to Education, *The University of Chicago Press, Chicago*.

Bellman, R. (1957). Dynamic Programming, *Princeton, New-Jersey, Princeton University Press*.

Belzil, C. (2007). Subjective Beliefs and Schooling Decisions, *IZA Discussion Paper* No 2820.

Belzil, C. and Hansen, J. (2007). A Structural Analysis of the Correlated Random Coefficient Wage Regression Model with an Application to the OLS-IV Puzzle, *Journal of Econometrics*, 140 (2), 333-948.

Belzil, C. and Hansen, J. (2002). Unobserved Ability and the Returns to Scooling, *Econometrica*, 70, 2075-2092.

Ben Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings, *Journal of Political Economy*, 75, 352-365.

Benitez-Silva, H. and Dwyer, D.S. (2005). The Rationality of Retirement Expectations and the Role of New Information, *Review of Economics and Statistics*, 87(3), 587-592.

Bonhomme, S. and Robin, J.M. (2007). Assessing the Equalizing Force of Mobility Using Short panels : France, 1990-2000, forthcoming *Review of Economic Studies*.

Bound, J., Brown, C., Duncan, G. and Rodgers, W. (1994). Evidence on the validity of cross-sectional and longitudinal labor market data, *Journal of Labor Economics*, 12, 3, 345-368.

Browning, M., Ejrnaes, M. and Alvares, J. (2006). Modelling Income Processes with Lots of Heterogeneity, *University of Oxford Working Paper*, No 285.

Buchinsky, M. and Hunt, J. (1999). Wage mobility in the United-States, *The Review of Economics and Statistics*, 81, 3, 351-368.

Buchinsky, M., Fields, G., Fougère, D. and Kramarz, F. (2003). Francs or Ranks ? Earnings Mobility in France, 1967-1999, *CEPR Discussion Paper*, No 3937.

Buchinsky, M. and Leslie, P. (2007). Educational Attainment and the Changing U.S. Wage Structure : Dynamic Implications Without Rational Implications, *mimeo*.

Burkhauser, R., Hotz, D. and Rhody, S. (1997). Labour earnings mobility and inequality in the United-States and Germany during the growth years of the 1980s, *International Economic Review*, 38, 4, 775-794.

Cameron, S.V. and Heckman, J.J. (1998). Life Cycle Schooling and Dynamic Selection Bias : Models and Evidence for Five Cohorts of American Males, *Journal of Political Economy*, April, 262-333.

Cameron, S.V. and Heckman, J.J. (2001). The Dynamics of Educational Attainment for Black, Hispanic and White Males , *Journal of Political Economy*, 109, 455-499.

Cameron, S.V. and Taber, C. (2004). Estimation of Educational Borrowing Constraints Using Returns to Schooling , *Journal of Political Economy*, 112 (1), 132-182.

Card, D. (1995). Earning, Schooling and Ability Revisited, *Research in Labor Economics*, 14, 23-48.

Cardoso, A.R. (2006). Wage Mobility : Do Institutions Make a Difference ?, *Labour Economics*, 13, 387-404.

Carneiro, P., Hansen, K. and Heckman, J.J. (2003). 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.

Carneiro, P. and Heckman, J.J. (2002). The Evidence on Credit Constraints in Post-Secondary Schooling, *The Economic Journal*, 112, October.

Cunha, F. and Heckman, J.J. (2007). Identifying and Estimating the Distributions of *Ex Post* and *Ex Ante* Returns to Schooling, *Labour Economics*, 14 (6), 870-893.

De Giorgi et al. (2006), Be careful of the books you reas as of the company you keep. Evidence on peer efects in educational chohices, IZA Discussion Papers 2833.

De Paola and Scoppa (2007), Peer froup effects on the academic performance of Italian students, mimeo.

Dominitz, J. and Manski, C.F. (2006). Eliciting Student Expectations of the Returns to Schooling, *Journal of Human Resources*, 31, 1-26.

Eckstein, Z. and Wolpin, K. (1989). The Specification and Estimation of Dynamic Stochastic Discrete Choice Models, *Journal of Human Resources*, 24, 562-598.

Eckstein, Z. and Wolpin, K. (1999). Youth Employment and Academic Performance in High School, *Econometrica*, 67 (6), 1295-1339.

Epple and Romano (1998), Competition between private and public schools, vouchers and peer group effects, *American Economics Review*, 88, 33-62.

Ge, S. (2007), Women's College Decisions : How Much does Marriage Matter ?, *mimeo.*

Gibbons, R. and Waldman, M. (2006). Enriching a Theory of Wage and Promotion

Dynamics inside Firms, *Journal of Labor Economics*, 24 (1), 59-107.

Gittleman, M., and M. Joyce (1995). Earnings Mobility in the United States, 1967-1991, *Monthly Labor Review*, 118, 9, 3-13.

Gottschalk, P. (1997). Inequality, income growth, and mobility : the basic facts, *The Journal of Economic Perspectives*, 11, 2, 21-40.

Heckman, J.J. (1981a). Heterogeneity and state dependence, *in Studies in Labor Markets, ed. by S. Rosen*, pp. 91-139. Chicago University Press, Chicago.

Heckman, J.J. (1981b). The Incidental Parameter Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process, *in Structural Analysis of Discrete Data with Econometric Applications, ed. by C. Manski, and D.McFadden*, pp. 179-195. MIT Press, Cambridge, MA.

Heckman, J.J., and B. Singer (1984). A Method For Minimizing the Impact of Distributional Assumptions in Econometric Models For Duration Data, *Econometrica*, 52, 271-320.

Heckman, J.J., Lochner, L. and Taber, C. (1998). General Equilibrium Treatment Effects : a Study of Tuition Policy, *American Economic Review, Papers and Proceedings*, 88 (2), 381-386.

Heckman, J.J. and Navarro, S. (2007). Dynamic Discrete Choices and Dynamic Treatment Effects, *Journal of Econometrics*, 136, 341-396.

Honore, B.E. and Kyriazidou, E. (2000). Panel Data Discrete Choice Models with Lagged Dependent Variables, *Econometrica*, 68, 4, 839-874.

Hoxby and Weingarth (2005), Taking race out of the equation : school reassignment and the structure of peer effects, mimeo.

Jovanovic, B. (1979). Job Matching and the Theory of Turnover, *Journal of Political Economy*, 87, 5, 972-990.

Kahneman, D. and Tversky, A. (1979). Prospect Theory : an Analysis of Decision under Risk, *Econometrica*, 47, 263-291.

Kane, T. (1996). College Cost, Borrowing Constraints and the Timing of College Entry, *Eastern Economic Journal*, 22 (2), 181-194.

Katz, L. and Autor, D (1999). Changes in the Wage Structure and Earnings Inequality, in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, 3, North Holland, 1463-1555.

Keane, M. and Wolpin, K. (1997). The career Decisions of Young Men, *Journal of Political Economy*, 105, 473-522.

Keane, M. and Wolpin, K. (2001). The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment, *The International Economic Review*, 42 (4).

Kermer and Levy (2003), Peer effects and alcohol use among College students, mimeo.

Kopczuk, W., Saez, E. and J. Song (2007). Uncovering the American Drean : Inequality and Mobility in Social Security Earnings Data Since 1937, *NBER Working Paper* No 13345.

Lavy et al. (2007), Inside the black box of ability peer effects : evidence from variation in high and low achieversr in the classroom, mimeo.

Lazear (2001), Education production, *Quarterly Journal of Economics*, 116, 777-803.

Levy, F. and Murnane, R. (1992), U.S. Earnings Levels and Earnings Inequality : A review of Recent Trends and Proposed Explanations, *Journal of Economic Literature*, 30 (3), 1333-1381.

Magnac, T. (2000). Subsidized training and youth employment, *The Economic Journal*, 110, 805-37.

Magnac, T. and Thesmar, D. (2002). Analyse Economique des Politiques Educatives : l'Augmentation de la Scolarité en France de 1982 à 1993, *Annales d'Economie et de Statistique*, 65, 1-32.

Manski, C.F. (2004). Measuring Expectations, *Econometrica*, 72 (5), 1329-1376.

Manski (1993), Identification and endogenous social effects : the reflection problem, *Review of Economic Studies*, 60, 531-542.

Martins and Walker (2006), Student achievement and university classes : effects of attendance, size, peers and teachers, IZA Discussion Papers 2490.

Meghir, R. and Pistaferri, L. (2004) Income Variance Dynamics and Heterogeneity, *Econometrica*, 72, 1-32.

Mincer, J. (1958) Investment in Human Capital and Personal Income Distribution, *Journal of Political Economy*, 66, 281-302.

Moffit, R. and Gottshalk, P. (1998). Trends in the variance of permanent and transitory earnings in the United-States and their relation to earnings mobility, *mimeo*.

Moffit, R. and Gottshalk, P. (2002). Trends in the transitory variance of earnings in the United-States, *The Economic Journal*, 112, C68-C73.

Riley, J. (1976). Information Screening and Human Capital, *American Economic Review*, 66, 254-260.

Rosen, S. (1982). Authority, Control, and the Distribution of Earnings, *Bell Journal of Economics*, 13 (2), 311-323.

Rust, J. (1994). Structural Estimation of Markov Decision Processes, in *Handbook of Econometrics*, ed. by R. Engle and D. McFadden. Amsterdam ; Elsevier Science, North-Holland Publishers, 3081-4143.

Sacerdote (2001), Peer effects with random assignement : results for Darthmouth roommates, *Quarterly Journal of Economics*, 116, 681-704.

Schiller, B. (1977). Relative Earnings Mobility in the United States, *American Economic Review*, 67, 5, 926-941.

Schiller, B. (1994). Relative Earnings Redux : Youth Mobility in the 80's, *Review of Income and Wealth*, 40, 441-456.

Spence, M. (1973). Job Market Signalling, *Quarterly Journal of Economics*, 87, 355-374.

Stiglitz, J. (1975). The Theory of Screening, Education and the Distribution of Income, *American Economic Review*, 65, 283-300.

Stinebrickner, R. and Stinebrickner, T.R. (2007). The Effect on Credit Constraints on the College Drop-out Decision : a Direct Approach Using a New Panel Study, *NBER Working Paper* No 13340.

Stockey, N. and Lucas, R.E. (1989). Recursive Methods in Economic Dynamics, *Harvard University Press. Massachusetts.*

Tallis, G.M. (1961). The Moment Generating Function of the Truncated Multi-Normal Distribution, *Journal of the Royal Statistical Society. Series B (Methodological)*, 23 (1), 223-229.

von Neumann, J. and Morgenstern, O. (1947). Theory of Games and Economic Behavior, *Princeton University Press.*

Waldman, M. (1984). Worker Allocation, hierarchies, and the Wage Distribution, *Review of Economic Studies*, 51 (1), 95-109.

Weber, A. (2002) . State Dependence and Wage Dynamics : A Heterogeneous Markov Chain Model for Wage Mobility in Austria, *Economics Series*, 114, Institute for Advanced Studies.

Willis, R. and Rosen, S. (1979). Education and Self-Selection, *Journal of Political Economy*, 87, S7-S36.

Winston and Zimmerman (2004), Peer effects in higher education, in Hoxby (ed.) *Colleges choices : the economics of where to go, when to go and how to pay for it*, Chicago University Press.

Wooldridge, J. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity, *Journal of Applied Econometrics*, 20, 1, 39-54.

Zimmerman (2003), Peer effects in academic outcomes : evidence from a natural experiment, *Review of Economic and Statistics*, 85, 9-23.