

Analyse économique des politiques éducatives : l'augmentation de la scolarisation en France de 1982 à 1993*

Thierry Magnac[†] David Thesmar[‡]

Première version: 15 septembre 2000

Révision: Avril 2001

Abstract : In this paper, we analyze the increase in education levels of young generations in France between 1980 and 1993. We insist on three factors that could have explained this increase : i) increasing returns to education in the labour market including its protective role against unemployment; ii) decreasing costs, either objective or subjective, of education that families or students have to pay; iii) decreasing "selectivity" of the education authorities. Using data from a survey on educational histories and on employment and by estimating a structural model of educational choices, we show that the third factor bears the principal burden of the increase in the education level. We however cannot distinguish in the influence of this factor what are the respective roles of opening new educational tracks, such as vocational tracks and, of rendering selection less severe by making repetitions less frequent.

Résumé : Nous analysons dans cet article l'augmentation du niveau d'éducation des jeunes générations en France entre 1980 et 1993. Nous isolons trois facteurs économiques principaux qui pourraient être responsables de cette hausse: i) la hausse des rendements de l'éducation sur le marché du travail en incluant son rôle de protection contre le chômage; ii) une baisse des coûts objectifs et subjectifs d'éducation payés par les étudiants et les familles; iii) une baisse de la sélectivité du système éducatif. En utilisant les données de l'enquête FQP en 1993, les enquêtes Emploi entre 1982 et 1993 et en développant un modèle structurel de choix éducatif que nous estimons, nous montrons que c'est le troisième facteur qui porte la responsabilité principale de l'augmentation du niveau d'éducation. Nous ne

*Les auteurs remercient Francis Kramarz, Guy Laroque, et Bernard Salanié et deux rapporteurs anonymes pour leurs critiques constructives ainsi que les audiences des séminaires donnés à l'INSEE, USaó Paulo, PUC-Rio et Tilburg. Nous demeurons néanmoins seuls responsables des erreurs et omissions qui subsistent.

[†]INRA et CREST

[‡]INSEE et CREST

savons toutefois pas distinguer dans l'influence de ce facteur le rôle de l'ouverture de nouvelles filières éducatives de la moindre prévalence des redoublements.

1 Introduction

Le niveau d'éducation des jeunes générations en France a crû de façon importante entre 1985 et 1996. Moins de 30% d'une génération en 1985 quittait le système éducatif avec un diplôme supérieur au baccalauréat. En 1996, plus de 60% de la jeune génération avait atteint ce niveau (Emin et Fallourd, 1996, Estrade et Minni, 1996). Du côté de l'offre, le moteur principal de cette augmentation du niveau d'éducation fut l'impulsion donnée aux politiques éducatives dans la seconde moitié des années 1980. En 1985, on annonça que l'objectif était d'amener 80% d'une génération au baccalauréat en l'année 2000. D'autres politiques éducatives, en particulier celles liées à l'éducation technique et à l'université, furent mises en oeuvre durant cette période (Prost, 1998).

L'analyse économique de cette augmentation peut se dériver d'un modèle simple d'investissement en capital humain (Becker, 1975). Celui-ci conduit à isoler trois causes potentielles pour expliquer les changements des comportements éducatifs. En premier lieu, l'évolution des salaires et des probabilités de chômage peuvent conduire à augmenter les rendements de l'éducation (Kodde, 1988). Deuxièmement, les coûts objectifs et subjectifs d'éducation peuvent diminuer à cause de politiques éducatives volontaristes. Finalement, il peut y avoir une augmentation des probabilités de passage à un niveau supérieur, variable sous le contrôle des autorités éducatives, mais qui conduit à une augmentation ex ante des rendements de l'éducation. L'idée poursuivie dans cet article est de désenchevêtrer l'importance respective de ces différents facteurs dans l'augmentation notable du niveau d'éducation à la fin des années 1980 et début des années 1990.

Il est à remarquer que la mesure des premier et troisième facteurs est assez aisée. En effet, d'une part on peut mesurer la hausse des inégalités dans les

salaires et dans les probabilités de chômage à partir d'estimations de fonctions de salaires sur des coupes répétées comme celles fournies par les enquêtes Emploi de l'INSEE. D'autre part, on peut mesurer l'évolution des probabilités de passage d'une classe à l'autre en utilisant des données sur les trajectoires scolaires. C'est l'enquête Formation et Qualification Professionnelle de l'INSEE qui fournira, dans cet article, de telles estimations. La mesure de l'évolution des coûts objectifs et subjectifs individuels est plus difficile et nécessite l'estimation d'un modèle structurel d'investissement en capital humain à la suite par exemple de Keane et Wolpin (1996), Belzil et Hansen (1999) ou Eckstein et Wolpin (1999). L'estimation de ces modèles de choix discret dynamique est a priori difficile: nous utilisons ici une méthode de moments construite à partir de la preuve d'identification non paramétrique développée dans un article compagnon (Magnac et Thesmar, 2000).

Pour résumer brièvement nos résultats, nous ne trouvons pas d'évolution notable, ni des rendements de l'éducation sur le marché du travail, ni des coûts objectifs et subjectifs de l'éducation. Il semble donc que le principal facteur de l'augmentation du niveau d'éducation sur cette période soit la plus grande facilité de passage entre deux niveaux d'éducation successifs. Celle-ci peut être rapportée à l'ouverture de nouvelles filières d'éducation destinées à des publics qui autrefois quittaient plus tôt le système éducatif, ou à la moindre grande sévérité des autorités éducatives pour le passage en classe supérieure. Quoiqu'il en soit, ce résultat soulève la question de l'outil et du critère les plus appropriés d'une politique éducative volontariste.

Dans la section 2, nous présentons les données et les faits stylisés sur les trajectoires éducatives et les salaires au cours de cette période. Nous y justifions l'utilisation du modèle structurel que nous présentons dans la section 3. La

méthode d'estimation et les résultats sont décrits dans la section 4. La section 5 conclut.

2 Données et faits stylisés

Dans cette section, nous décrivons d'abord brièvement le système éducatif et l'enquête Formation Qualification Professionnelle (FQP, INSEE 1993) qui nous permet de construire les parcours scolaires et les variables décrivant le contexte socio-familial. A partir de cette enquête, nous montrons que l'augmentation de la scolarisation des jeunes générations affecte surtout les niveaux d'éducation autour du baccalauréat. Ensuite nous analysons les déterminants socio-démographiques de cette augmentation et nous revenons sur les principaux facteurs économiques qui expliquent cette hausse.

Nous essayons de mettre en valeur l'inflexion marquée des politiques éducatives, qui, selon divers indices, peut se situer approximativement vers les années 1987 à 1989, comme le montre par exemple le graphique 1 représentant la part du budget public français consacré aux dépenses d'éducation¹. C'est pourquoi nous choisissons d'analyser les trajectoires scolaires des cohortes nées entre 1963 et 1973. Les premières cohortes ont fini leur scolarité avant la période de rupture alors que les dernières cohortes ont vécu leur trajectoire scolaire après 16 ans après 1987. Ce type de diagnostic qui contraste des évolutions entre les périodes avant la réforme et après la réforme (Heckman, LaLonde et Smith, 2000) reste néanmoins sujette à caution puisque la sortie de la dépression économique du milieu des années 80 se situe à la même période et ces effets du cycle macroéconomique doivent être contrôlés.

¹Les changements se sont étalés dans le temps entre les annonces des politiques et leurs mises en oeuvre.

2.1 Brève description du système éducatif et des données de l'enquête FQP en 1993

L'enquête FQP est réalisée environ tous les huit ans par l'INSEE auprès d'un échantillon représentatif de ménages. Nous exploitons la dernière vague en date: celle de 1993. Dans chaque ménage appartenant à la base de sondage, chacun des membres est interrogé séparément sur son parcours scolaire et professionnel, sa situation socio-économique actuelle et celle de ses parents, etc. En particulier, l'ensemble du parcours scolaire est détaillé d'année en année dès la première année de présence à l'école. Notre échantillon de travail extrait de cette enquête est composée de jeunes hommes nés entre 1963 et 1973. Les premières analyses nous ont en effet conduit au diagnostic de différences marquées entre jeunes hommes et jeunes femmes et nous avons donc choisi l'une de ces populations seulement. L'échantillon est donc constitué d'environ 2000 jeunes hommes âgés de 20 à 30 ans en 1993.

Comme l'information donnée sur les parcours scolaires est rétrospective, la question de l'attrition, ou endogénéité de la sortie du panel, ne se pose pas. Certains de nos parcours scolaires peuvent aussi être censurés à droite car l'individu correspondant était encore à l'école en 1993. Cette censure peut bien sûr avoir un effet non négligeable sur des statistiques descriptives en coupe transversale, mais elle n'en a aucun sur des statistiques descriptives longitudinales ou sur l'estimation de notre modèle structurel en temps discret. La présence d'erreurs de mesure ou les problèmes de sélection de l'échantillon sont plus importants. En particulier, la base de sondage étant constituée de logements, les jeunes hommes qui ont quitté le domicile familial afin de suivre des études supérieures, de faire leur service militaire ou de chercher du travail au loin, risquent de ne pas être inclus dans l'échantillon. En effet, les casernes, les cités universitaires ou foyers

collectifs ne sont pas dans le champ de l'enquête. Témoignage de l'importance potentielle de la sélection, les classes d'âge ne sont pas distribuées uniformément, décroissant de 12% de l'échantillon à 20 ans à 6% pour les individus âgés de 25 ans. Cette proportion croît à nouveau après 25 ans pour rejoindre 10% à 30 ans. Nous faisons ici le choix de négliger ce phénomène, en supposant que la décision de quitter le domicile familial est exogène aux choix entre travail et éducation².

La présence dans le système éducatif est obligatoire entre 6 et 16 ans depuis 1959 (Prost, 1998). Le choix de quitter l'école n'est donc donné qu'à partir de 16 ans. Néanmoins, c'est la structure du système éducatif avant 16 ans qui détermine la façon dont nous construisons notre échantillon car le parcours scolaire précédant 16 ans constitue un déterminant important de la trajectoire à partir de cet âge. A partir de maintenant pour simplifier l'exposition, nous notons chaque classe par l'âge qu'aurait un élève qui y participerait sans avoir jamais redoublé ou sauté une classe auparavant. Avant d'atteindre un niveau quatrième ou équivalent (14 ans), les élèves suivent des cursus très homogènes. Seuls 2.2% des individus dans nos données ont participé à des programmes d'éducation spécialisée, qui concernent des élèves handicapés ou en situation d'échec scolaire grave. L'éducation "générale" est donc la norme de 6 à 14 ans, puis, à 14 ans, les trajectoires peuvent bifurquer pour la première fois. La plupart des élèves (les deux tiers environ) poursuivent le programme général. Les autres entrent dans des filières technologiques "courtes" (CAP,BEP), ou des programmes d'apprentissage incluant des cours jusqu'à l'âge de 16 ans.

Mais c'est à partir de 16 ans que les trajectoires se diversifient réellement.

Les années de lycée, entre 16 et 18 ans, peuvent se passer dans deux filières

²L'hypothèse est forte. Une façon possible de corriger ce biais de sélection est d'utiliser des pondérations dans la procédure d'estimation (Coslett [1994]), en calculant les poids à partir d'une autre source dont la base de sondage n'est pas le ménage, comme le recensement de 1990.

différentes: générale, ou technique “longue”, représentant à peu près le même nombre d’élèves. Ceux qui ont amorcé des études techniques courtes au collège les finissent en général entre 16 et 18 ans. Après l’obtention du baccalauréat, qui clôt les années de lycée, les élèves peuvent poursuivre leurs études dans l’enseignement supérieur.

Notre échantillon final exclut les élèves qui ont suivi une éducation spécialisée, ou l’apprentissage. Il ne contient pas non plus les individus qui n’ont pas atteint la sixième à 16 ans, car nous estimons que ces histoires quasiment irréversibles ne sont pas prises en compte par le modèle que nous développerons dans la section 3. Finalement, il nous reste 1519 individus, pour lesquels nous disposons des variables suivantes:

- **Parcours scolaire complet:** à chaque âge, le niveau de la classe dans laquelle se trouve l’élève. Nous définissons le niveau comme l’âge qu’un élève aurait dans cette classe s’il n’avait pas redoublé ou sauté une classe.
- **Contexte socio-familial:** niveau d’étude atteint par les parents (sans le baccalauréat, baccalauréat ou diplôme universitaire), l’âge, la profession des parents suivant la nomenclature socioprofessionnelle à deux chiffres³, le lieu de résidence (commune rurale ou non, habitant Paris ou non), la taille de la fratrie, et le fait que la mère soit une femme au foyer.

D’abord et de manière peu surprenante, la probabilité de quitter le système éducatif s’accroît tout au long de la scolarité. Pour fixer les idées, les taux de départ ont été calculés pour chaque niveau scolaire. Ils sont représentés dans le tableau 1, pour l’ensemble des individus (dernière colonne), mais aussi séparé-

³Nous n’utilisons pas cette variable directement mais nous profiterons de sa forte corrélation avec les revenus pour construire, par imputation, des revenus familiaux. Pour cela, nous nous servons des enquêtes Emploi décrites plus loin pour estimer des fonctions de revenus des parents en fonction des CSP et d’autres variables comme l’éducation et l’âge.

Niveau	Ensemble	Echantillon 1	Echantillon 2
16	6.2 (0.6)	3.3 (0.6)	9.2 (1.0)
17	24.4 (1.0)	19.2 (1.3)	30.0 (1.6)
18	14.6 (1.1)	10.8 (1.2)	20.9 (2.1)
19	7.1 (1.0)	5.8 (1.0)	10.3 (2.1)
20	23.2 (1.7)	18.4 (1.8)	39.3 (4.1)
21	19.0 (2.6)	18.5 (2.8)	20.9 (6.2)
22	24.2 (3.8)	23.5 (4.2)	26.9 (8.7)
23	47.9 (5.9)	50.8 (6.5)	33.3 (13.6)
Observations	1519	793	726

Note: Ecart types entre parenthèses. Echantillon 1: Elèves n'ayant pas redoublé avant l'âge de 16 ans. Echantillon 2: Elèves ayant redoublé au moins une fois avant l'âge de 16 ans. Parts d'individus quittant l'école à un niveau donné. Lecture: "Dans l'ensemble, 6.2% des élèves du niveau seconde (16 ans) ne sont pas à l'école l'année d'après".

Tableau 1: Taux de sortie de l'école

ment pour les individus qui n'ont pas redoublé avant l'âge de 16 ans, et pour ceux qui ont redoublé au moins une fois. Entre la quatrième et la seconde, le taux de sortie est faible (entre 3 et 7%). Il s'élève ensuite aux alentours de 20% en première et en terminale. Puis il chute à 7% en première année d'études supérieures, certainement parce que la première année d'études supérieures ne donne pas droit, en général, à l'obtention d'un diplôme. A partir du niveau BAC+2, le taux de sortie dépasse 20%. Ce tableau illustre aussi combien les taux de départ sont différents entre les individus qui ont redoublé avant la seconde et les autres. Cette très forte corrélation peut traduire (1) la présence d'hétérogénéité inobservée, révélée en partie par le filtre de la sélection et/ou (2) une "stigmatisation" de l'élève par le système éducatif. Quoiqu'il en soit, il est nécessaire d'inclure cette variable dans notre analyse des choix d'éducation.

La montée de la scolarisation est attestée par notre échantillon issu de FQP. Le graphique 2 retrace, pour chaque niveau d'éducation (du niveau quatrième ou niveau 14, au niveau DEA, ou niveau 23), l'évolution de la fréquence empirique des individus qui restent dans le système scolaire à chaque niveau d'éducation et selon leur année de naissance. Ces fréquences sont construites cohorte de naissance par cohorte de naissance et les courbes inférieures et supérieures fournissent les intervalles de confiance à 95%. Entre les niveaux quatrième (14 ans) et seconde (16 ans), la probabilité de demeurer dans le système éducatif ne s'est pas significativement accrue pour les cohortes récentes. Ce résultat n'est pas très surprenant dans la mesure où celle-ci est déjà très élevée (entre 95 et 100%). C'est pour les niveaux première (17) et terminale (18) que l'accroissement de la probabilité de continuer ses études a été le plus important ces dernières années. Dans notre échantillon, un élève né en 1973 a une chance de passer de première à terminale qui est environ 15% plus importante qu'un élève né en 1963. Cette augmentation de la probabilité de continuer ses études est égale à 10% entre la terminale et le niveau supérieur. Ces deux écarts sont statistiquement significatifs. Au delà du niveau de la terminale, les accroissements sont moins forts, et non significatifs, principalement car le nombre d'observations devient trop faible.

Dans le tableau 2, nous avons étendu l'analyse des déterminants de la sortie de système éducatif à d'autres variables du contexte social et familial. Pour cela, nous estimons la moyenne des variables socio-démographiques conditionnellement à l'obtention d'un certain niveau d'éducation. On calcule la moyenne du revenu des parents dans les échantillons des personnes qui ont atteint au moins le niveau 14 (respectivement les niveaux 15 à 23). C'est une des manières de présenter la sélection dans le système scolaire en fonction des variables socio-démographiques (Mare, 1980). Si la moyenne conditionnelle d'une variable augmente de façon

Niveau	16	17	18	19	20	23
Taille de la fratrie	2.23 (0.05)	2.06 (0.05)	1.87 (0.05)	1.91 (0.06)	1.87 (0.07)	1.97 (0.20)
Habitat rural (%)	21 (1)	20 (1)	14 (1)	12 (1)	08 (1)	05 (4)
Parents: études supérieures (%)	31 (1)	34 (1)	45 (2)	51 (2)	57 (2)	73 (7)
Parents: lycée (%)	55 (1)	54 (1)	46 (2)	41 (2)	35 (2)	24 (7)
Parents: log(salaire)	5.81 (0.01)	5.85 (0.01)	5.91 (0.01)	5.93 (0.01)	5.96 (0.01)	6.00 (0.04)
Mère au foyer (%)	18 (1)	15 (1)	14 (1)	14 (1)	15 (2)	11 (5)

Note: Écarts types entre parenthèses. Moyenne de la variable considérée pour toutes les observations correspondant à un élève dans la classe du niveau considéré. Lecture " parmi les élèves qui sont en seconde, 31% ont des parents ayant fait des études supérieures".

Table 2: Déterminants des choix éducatifs

monotone avec le niveau scolaire, ceci traduit le fait qu'elle exerce une influence positive sur la probabilité de continuer dans le système éducatif.

Jusqu'au niveau terminale, le nombre de frères et soeurs apparaît comme un facteur favorisant la sortie du système éducatif puisque sa moyenne conditionnelle en fonction du niveau d'éducation diminue. Les parents des élèves restant longtemps à l'école ont plus de chance d'avoir fait des études supérieures, sont mieux payés et habitent plus souvent en milieu urbain. Même si le fait que les parents aient été au lycée eux mêmes est aussi un facteur favorisant l'allongement de la durée des études des enfants, cette influence devient de moins en moins importante avec la durée de la scolarité. Enfin, la part d'étudiants dont la mère ne travaille pas diminue avec le niveau scolaire, quoique non significativement.

Tous ces effets sont bien évidemment liés entre eux et seule une analyse toutes choses égales par ailleurs permettra de désenchevêtrer ces divers effets. Nous reporterons dans la section 4 les résultats de l'estimation de modèles de choix discrets qui permettent de mener à bien une telle analyse. Enfin, ces déterminants pourraient aussi avoir évolués au cours de la période et leur évolution pourrait

être en mesure d'expliquer la montée de la scolarisation. Nous verrons en section 4 qu'il n'en est rien.

2.2 Les causes potentielles

L'analyse descriptive qui suit est consacrée au recensement et à l'évaluation du pouvoir explicatif des divers facteurs économiques de la montée de la scolarisation.

2.2.1 Facteurs de demande : salaires et chômage

Les facteurs ayant pu influencer la demande d'éducation sont liés aux rendements de l'éducation sur le marché du travail en termes de salaires ou l'évolution de la capacité protectrice de l'éducation contre le chômage au cours de la période (Micklewright, Pearson et Smith, 1989). Selon Goux et Maurin (1994) ou Wasmer (1998), le différentiel de salaires selon les niveaux d'éducation est resté à peu près stable entre les enquêtes FQP de 1970, 1977, 1985 et 1993. Néanmoins, la forte montée du chômage des jeunes au cours des années 1980 pourrait avoir encouragé les jeunes hommes à demeurer dans le système éducatif car le différentiel de chômage par niveau de diplôme et par tranche d'âge s'est dramatiquement creusé pendant la période considérée. Ainsi, entre 1982 et 1992, dans l'échantillon de l'enquête Emploi que nous allons présenter plus bas, le taux de chômage des jeunes hommes de niveau scolaire inférieur au baccalauréat est passé de 13.4% à 19.8%. Celui des diplômés du baccalauréat, au moins, ne s'est accru que de 10 à 12%.

L'enquête FQP, qui, depuis 1980, n'a été menée qu'en 1985 et 1993, n'est pas à même de nous fournir suffisamment d'informations sur l'évolution des rendements de l'éducation au cours de la période. C'est pourquoi nous utilisons l'enquête Emploi. Celle-ci est réalisée chaque année auprès d'un panel d'environ 100.000

ménages tournant par tiers. Comme l'enquête FQP, sa base de sondage est le ménage. Nous exploitons les vagues 1982-1993 de cette enquête. Pour chaque personne interrogée possédant un emploi, l'enquête Emploi nous fournit le salaire et les variables habituelles apparaissant dans les fonctions de salaire comme le niveau de diplôme et l'expérience professionnelle. Un des inconvénients de cette enquête vient du codage du diplôme qui n'est pas aussi détaillé que dans l'enquête FQP. D'autre part, jusqu'en 1989 inclus, le salaire est donné par tranches de valeurs et nous lui imputerons la moyenne des tranches. A partir de 1990, le montant du salaire est renseigné. Cette rupture de codage de la variable, ainsi qu'un changement de la méthode de collecte des données expliquent les variations brutales des rendements de l'éducation entre 1989 et 1990, comme nous le verrons plus bas. Cette source nous permettra toutefois d'établir la remarquable stabilité des rendements de l'éducation et de l'expérience au cours de la période pour notre échantillon composé par les hommes de 16 à 60 ans qui ne sont pas dans le système scolaire.

Nous estimons l'évolution des espérances de revenu entre 1982 et 1993 à partir de l'enquête Emploi, en retenant les deux facteurs de capital humain que sont l'expérience et de l'éducation. Nous ne cherchons pas à estimer une équation de salaire potentiel en corrigeant pour le biais de sélection dû aux chômeurs mais à estimer une équation prédictive des revenus salariaux. Cette estimation permet de mesurer l'évolution des rendements du diplôme d , en tenant compte - d'une manière très fruste - du risque de chômage. Soit w_{it} le salaire perçu, en Francs de 1980 par l'individu i à l'année t si l'individu travaille. Si l'individu ne travaille pas (inactivité ou chômage), nous posons $w_{it} = 0$.⁴ Nous estimons le modèle

⁴Cette hypothèse définit implicitement une certaine aversion pour le risque de chômage. On a aussi estimé pour tenir compte d'une aversion pour le risque plus plausible, des équations en logarithme du salaire, en posant $\log w_{it} = \log 500$ pour les non employés. Nos conclusions sur l'évolution du rendement de l'éducation sont robustes aux changements dans ces deux hy-

suivant:

$$\begin{aligned}
w_{it} = & \alpha_t + \sum_{d=1}^D \beta_t^d \mathbf{1}_{\{\text{dipl}_{it}=d\}} + \delta_t \cdot \mathbf{1}_{\{\text{technique court}\}} + \nu_t \cdot \mathbf{1}_{\{\text{technique long}\}} \quad (1) \\
& + \sum_{a=16}^{25} \gamma_t^a \mathbf{1}_{\{\text{age}_{it}=a\}} + \mu_{1t} \cdot \mathbf{1}_{\{\text{age}_{it}>25\}} \cdot \text{age}_{it} + \mu_{2t} \mathbf{1}_{\{\text{age}_{it}>25\}} \cdot \text{age}_{it}^2 + \varepsilon_{it}
\end{aligned}$$

où dipl_{it} représente le niveau de diplôme (en années d'études), et où l'on inclut deux indicatrices indiquant si l'individu a suivi un cursus technique court ou long (ou les deux). Les rendements de l'éducation β_t^d sont donc estimés pour chaque date séparément pour 6 niveaux de diplôme, correspondant aux grandes étapes du parcours scolaire (niveaux quatrième, troisième, BAC, BAC+2, BAC+3, BAC+5). Nous consacrons par ailleurs une attention particulière à modéliser les rendements de l'expérience de manière aussi flexible que possible: nous utilisons des indicatrices pour chaque année avant 25 ans, puis la forme quadratique classique après. Les coefficients, γ_t^a , mesurent l'évolution des rendements de l'expérience pour chaque année en début de vie professionnelle.

Le graphique 3 illustre l'évolution des rendements de l'éducation (coefficients β_t^d) sur la période 1982-1996 pour 6 niveaux de diplômes, en francs de 1980: quatrième, troisième, terminale, BAC+2, BAC+3 et BAC+5. Les rendements sont évalués en référence aux diplômes obtenus à 17 ans. Le décrochage de 1990, qui n'est pas de même ampleur selon les niveaux de diplôme est dû au changement de codage de la variable salaire et à une modification de la méthode de collecte. Si l'on excepte cette rupture, les rendements, qui tiennent compte de l'inactivité et du chômage, sont très stables dans le temps. L'évolution du différentiel de chômage qualifié/non qualifié ou de salaires ne semble donc pas être un bon candidat pour expliquer l'accroissement de la demande d'éducation. Pour donner également une évaluation quantitative de la montée du chômage des jeunes sur

pothèses: le niveau de revenu hors travail et l'aversion au risque.

leur revenu moyen, le graphique 4 retrace quant à lui l'évolution du rendement de l'expérience (coefficients γ_t^a) en francs de 1980 au cours de la période, à partir des indicatrices d'âge, estimées par l'équation (1). Ici encore, l'équation de l'espérance de salaire pour les jeunes hommes témoigne d'une remarquable stabilité au cours de la période.

Au total, il apparaît donc légitime de considérer que les facteurs de demande d'éducation sont restés relativement stables au cours de la période (dans notre échantillon de jeunes hommes), malgré la montée du chômage des jeunes non qualifiés dans les années 1980. Ces résultats confortent ceux qui ont été obtenus par Goux et Maurin [1994] ou Bayet et Cases [1994]. Bien qu'ils utilisent une méthodologie différente pour prendre en compte le risque de chômage, ils arrivent à la conclusion que les rendements de l'éducation sont à peu près stables, voire déclinants sur la période 1980-93 ce qui est en accord avec nos résultats.

2.2.2 Facteurs d'offre.

La hausse de la scolarisation peut aussi avoir été déclenchée par un accroissement de l'offre, suite à une politique volontariste menée à partir du milieu des années 1980: l'objectif affiché en 1985 par le gouvernement était d'amener 80% d'une génération au baccalauréat. Du point de vue des élèves, cette politique d'offre peut avoir revêtu deux formes: (1) un abaissement des coûts objectifs et subjectifs de la scolarisation pour les familles ou les étudiants (construction d'écoles ou d'universités, classes moins nombreuses, création de filières techniques) et (2) une moindre sélectivité du système éducatif, le système facilitant ainsi l'obtention plus aisée de nouveaux ou d'anciens diplômes.

L'abaissement des coûts implicites de l'éducation a constitué une priorité de l'action gouvernementale au cours des années 1980. Par la promotion de

l'enseignement technique dans le secondaire tout d'abord. Le baccalauréat professionnel est créé en 1985, et avec lui, la filière technique, jusque là limitée au CAP et BEP, peut se prolonger au lycée. Cet effort a ensuite porté sur le troisième cycle. En 1988, le gouvernement lance le plan université 2000, qui consiste à accroître la capacité d'accueil des universités. Elle s'est traduite par la création de nouveaux établissements universitaires⁵, mais aussi par la création de filières de troisième cycle plus en prise avec le marché du travail (Instituts Universitaires Techniques par exemple). La mise en place du plan Université 2000 s'est traduite par un rebond de la part des dépenses d'éducation dans le PIB à partir de la fin des années 1980 (de 6.4 à 7.4 points: graphique 1). Selon les filières la dépense moyenne par étudiant s'est accrue de 20 à 30% sur la période 1981-1996 (graphique 5).

Cependant, le second facteur d'offre (baisse de la sélectivité du système éducatif) ne peut être négligé, comme en atteste le graphique 6: nous avons tracé, à partir de notre échantillon, l'évolution de la probabilité de passer dans la classe supérieure en fonction de la classe de départ et de l'année de naissance. Cette mesure ne contient pas les élèves qui ont eu l'opportunité de passer mais qui ont préféré quitter l'école. Il s'agit donc d'une borne inférieure de la sélection. Celle-ci est stable aux niveaux troisième et seconde; elle a la forme d'un U au niveau première. C'est à partir du niveau terminale que la probabilité d'atteindre le niveau supérieur s'est accrue considérablement entre les cohortes 1963 et 1973. En particulier, la probabilité de passer à BAC+2 à partir de BAC+1 a augmenté de 20% sur la période.

Au total, le système éducatif est devenu moins sélectif au cours de la période.

⁵Dans la banlieue parisienne, quatre nouvelles universités ont été créées (Saint Quentin-Versaille, Evry, Marne La Vallée, Cergy-Pontoise). En province, le plan université 2000 a notamment permis la création des Universités du Littoral et d'Artois.

Cette évolution s'est concentrée dans l'enseignement supérieur. Ainsi, la montée de la scolarisation enregistrée aux niveaux première et terminale pourrait avoir été déclenchée par la perspective d'études supérieures plus longues.

3 Le modèle théorique

L'explication la plus raisonnable de la montée de la scolarisation semble donc à trouver dans les facteurs d'offre: une sélectivité moindre et un abaissement du coût implicite de s'éduquer. Pour aller plus loin et départager ces deux effets, il nous faut un modèle qui relie les choix d'éducation (1) à la sélectivité et (2) aux coûts implicites. Comme nous observons dans FQP les décisions de scolarisation et l'évolution de la sélectivité, ce modèle nous permettra d'en déduire l'évolution des coûts implicites.

L'objet de cette section est de proposer un tel modèle économique, et de montrer comment on peut l'estimer. Ce modèle est un modèle d'accumulation de capital humain. A chaque période le choix est discret: rester dans le système éducatif et tenter d'augmenter son capital humain d'une année d'étude (si on ne redouble pas), ou quitter l'école et travailler. Afin de simplifier l'exposition, nous reviendrons à l'interprétation de ce modèle en termes d'offre et de demande en fin de section.

3.1 Modèle de comportement

Les agents sont en activité T périodes et prennent leur retraite ensuite. Au début de chaque période $t < T$, l'agent peut être dans l'une des deux situations suivantes: (1) étudier ou (2) ne pas étudier, état que nous assimilerons au travail même s'il inclut des périodes de chômage ou d'inactivité. On note d_t la décision de l'agent et $d_t = W$ s'il travaille, et $d_t = S$ s'il étudie. Il s'agit d'un modèle de choix discret dynamique. L'allongement des études implique en effet le sacrifice

d'un salaire aujourd'hui, contre le paiement (espéré) de salaires plus élevés plus tard. Nous supposons que quitter l'école est une décision irrévocable si bien que le problème de choix discret ne concerne que les individus scolarisés. Cette hypothèse, qui n'est pas nécessaire, mais qui simplifie l'application de notre méthode d'estimation, est validée par les données, puisqu'en France, selon l'enquête Emploi, les retours vers le système éducatif ne concernaient en 1993 que 2% de la population active.

Les études, comme le travail, apportent leurs lots de pleurs et de consolations sous forme d'une utilité u . L'agent i à la période t reçoit un flux d'utilité $u(x_{it}, \tilde{\varepsilon}_{it}, d_{it}) = u_S(x_{it}) + \varepsilon_{S,it}$ s'il est à l'école (coûts objectifs et subjectifs de l'éducation). Il reçoit $u(x_{it}, \tilde{\varepsilon}_{it}, d_{it}) = u_W(x_{it}) + \varepsilon_{W,it}$ s'il n'est pas à l'école: ceci représente les salaires ou indemnités perçus. On note les variables d'état $(x_{it}, \tilde{\varepsilon}_{it})$ affectant le bien être dans l'un et l'autre état: x_{it} correspond à l'ensemble des variables d'état observables par l'économètre (l'âge, le niveau d'études atteint et donc les décisions de scolarisation précédentes, mais aussi le métier des parents, leur niveau d'éducation, etc.). Les chocs $\tilde{\varepsilon}_{it} = (\varepsilon_{S,it}, \varepsilon_{W,it})$ sont les variables inobservables (aptitude ou goût inobservés à l'école, opportunités sur le marché du travail). Les fonctions d'utilité u_S et u_W ne sont pas supposées stationnaires. C'est précisément l'objet de cette étude que de considérer leurs non-stationarités: par exemple, l'évolution du chômage et des salaires sur la période fait baisser u_W , ou l'abaissement des coûts implicites de l'éducation fait croître u_S . L'indice de la période, t , est donc inclus parmi les variables observables x_{it} .

Au sein d'une période, la séquence des événements est la suivante: au début de la période l'agent est à l'école. Il est doté de l'histoire passée x_{it} . Les valeurs des deux chocs $\tilde{\varepsilon}_{it}$ sont tirées selon une loi de fonction de répartition G constante

au cours du temps et indépendante du passé⁶. La réalisation de ces chocs traduit un état de la nature pouvant affecter les préférences de l'individu observées par celui-ci, mais inconnues de l'économètre. Par exemple, il découvre alors qu'il dispose d'une bonne opportunité sur le marché du travail, que les classes sont trop nombreuses, ou que les professeurs dans sa classe sont de bons pédagogues. En fonction de ces chocs, et des variables d'états observables x_{it} , il décide alors s'il quitte l'école ($d_{it} = W$), ou s'il demeure dans le système éducatif ($d_{it} = S$). Il reçoit alors l'utilité instantanée correspondante. Finalement, une fois x_{it} , $\tilde{\varepsilon}_{it}$, et d_{it} connus, x_{it+1} est tirée selon une loi conditionnelle à (x_{it}, d_{it}) .⁷ En particulier, le passage dans la classe supérieure, ou l'obtention du diplôme sanctionnant les études sont gouvernés par cette loi. La séquence des évènements au sein d'une période est résumée figure 7.

La décision de l'agent est donc le résultat d'un arbitrage entre (1) le coût d'opportunité de ne pas travailler et donc de ne pas toucher de salaire et (2) le gain en terme de salaires futurs que représente la possibilité d'avoir un niveau scolaire accru d'un an. Il y a, dans le modèle, deux sources d'incertitude. La première concerne les chocs affectant les utilités instantanées de travailler et de demeurer à l'école: elle précède la décision. La deuxième provient du processus de transition de x_{it} à x_{it+1} et se réalise après la décision de scolarisation. En particulier, c'est la probabilité de passer dans la classe supérieure qui représente le degré de sélectivité du système éducatif dans notre modèle.

⁶ Cette hypothèse est forte. Elle signifie qu'il n'y a pas d'hétérogénéité inobservable persistante. On peut justifier cette hypothèse en supposant que l'individu ne dispose pas de plus d'information que l'économètre lorsqu'il forme ses anticipations sur les futurs $\tilde{\varepsilon}_{it}$. On peut aussi étendre ce modèle pour prendre en compte l'hétérogénéité inobservable mais la méthode d'estimation est alors beaucoup plus lourde à mettre en oeuvre (Magnac et Thesmar, 2000).

⁷ En toute généralité, il faudrait supposer que la loi de x_{it+1} dépend de x_{it} , d_{it} et $\tilde{\varepsilon}_{it}$. Mais pour des raisons de cohérence interne du modèle explicitées par Manski [1993], il est valide de se restreindre aux cas où la loi de x_{it+1} ne dépend que de x_{it} et d_{it} . On peut établir que cette hypothèse revient à supposer l'absence d'hétérogénéité inobservable. La présence de celle-ci n'est pas identifiée en l'absence de restrictions supplémentaires (Magnac et Thesmar [2000]).

Nous supposons - et c'est là notre principale restriction identifiante - que les agents sont rationnels. L'implication de cette hypothèse est que ceux-ci choisissent leur parcours scolaire en résolvant le programme de maximisation suivant, pour tout t :

$$\max_{\{d_{it}, \dots, d_{it+\tau}, \dots\}} E \left\{ \sum_{\tau=1}^T \beta^\tau u(x_{it+\tau}, \tilde{\varepsilon}_{it+\tau}, d_{it+\tau}) + V_{T+1}(x_{iT+1}) \mid x_{it}, \tilde{\varepsilon}_{it} \right\}$$

sous les contraintes:

$$x_{it+1} \text{ tirée dans une loi conditionnelle à } x_{it} \text{ et } d_{it+\tau-1}$$

$$\tilde{\varepsilon}_{it+\tau} \text{ tirée dans une loi de fonction de répartition } G$$

$$d_{it+\tau} = W \text{ si } d_{it+\tau-1} = W$$

et où $V_{T+1}(\cdot)$ est une fonction valeur quelconque pour la période $T + 1$ après la cessation d'activité et où β est le facteur d'actualisation.

3.2 Equations de Bellman

Notre objectif est de montrer comment estimer u_S , en connaissant les décisions de scolarisation des agents d_{it} , leurs caractéristiques observables x_{it} et le degré de sélectivité du système $P(H'|H, d)$. Nous allons donc résoudre, partiellement, le modèle afin d'en déduire une forme réduite qui permette d'estimer facilement u_S . Nous allons le voir, cette forme réduite se dérive des équations de Bellman, qui sont un autre moyen d'écrire le programme ci-dessus.

Soit $V_S(x_{it}, \tilde{\varepsilon}_{it})$ la somme actualisée des utilités futures d'un agent i choisissant à la date t de rester dans le système éducatif. Soit $V_W(x_{it}, \tilde{\varepsilon}_{it})$ la somme actualisée des utilités futures d'un agent qui n'est pas dans le système éducatif à la période t . Les agents décident donc de rester dans le système éducatif si et seulement si $V_S(x_{it}, \tilde{\varepsilon}_{it}) > V_W(x_{it}, \tilde{\varepsilon}_{it})$. Comme l'horizon de ce problème dynamique est fini, les fonctions de valeur $V_W(x_{it}, \tilde{\varepsilon}_{it})$ et $V_S(x_{it}, \tilde{\varepsilon}_{it})$ satisfont alors

les équations de Bellman :

$$V_S(x_{it}, \tilde{\varepsilon}_{it}) = u_S(x_{it}) + \tilde{\varepsilon}_{S,it} + \beta E(\max\{V_S(x_{it+1}, \tilde{\varepsilon}_{it+1}), V_W(x_{it+1}, \tilde{\varepsilon}_{it+1})\} | x_{it}, d_{it} = S) \quad (2)$$

$$V_W(x_{it}, \tilde{\varepsilon}_{it}) = u_W(x_{it}) + \tilde{\varepsilon}_{W,it} + \beta E(V_W(x_{it+1}, \tilde{\varepsilon}_{it+1}) | x_{it}, d_{it} = W) \quad (3)$$

L'équation régissant V_W ne contient pas de terme en maximum car nous avons supposé que la décision de quitter l'école était irrévocable. Notons $\Delta(x_{it}, \tilde{\varepsilon}_{it}) = V_S(x_{it}, \tilde{\varepsilon}_{it}) - V_W(x_{it}, \tilde{\varepsilon}_{it})$: i choisit donc de rester dans le système éducatif si $\Delta(x_{it}, \tilde{\varepsilon}_{it}) > 0$. En soustrayant les deux équations précédentes membres à membres, on obtient :

$$\begin{aligned} \Delta(x_{it}, \tilde{\varepsilon}_{it}) &= \underbrace{[u_S(x_{it}) - u_W(x_{it})]}_{\text{Gain Instantané}} + (\varepsilon_{S,it} - \varepsilon_{W,it}) \\ &+ \underbrace{\beta E(\max\{\Delta(x_{it+1}, \tilde{\varepsilon}_{it+1}), 0\} | x_{it}, d_{it} = S)}_{\text{Valeur d'option}} \\ &+ \underbrace{\beta(E(V_W(x_{it+1}, \tilde{\varepsilon}_{it+1}) | x_{it}, d_{it} = S) - E(V_W(x_{it+1}, \tilde{\varepsilon}_{it+1}) | x_{it}, d_{it} = W))}_{\text{Rendement futur direct de l'éducation}} \end{aligned} \quad (4)$$

Comme le fait apparaître cette équation, la décision de poursuivre les études est influencée par trois effets. Tout d'abord il y a le rendement instantané de demeurer dans le système scolaire, qui inclut les plaisirs et déplaisirs liés aux études, les coûts monétaires de l'éducation, et le coût d'opportunité de ne pas être sur le marché du travail. Le second terme représente la valeur d'option de rester à l'école (Conway, Melnik et Pollatschek, 1973). En effet, comme le retrait est une décision irréversible, rester aujourd'hui permet à l'agent de rester demain si cela en vaut la peine. Le dernier terme traduit les gains potentiels en terme de salaires futurs de rester à l'école aujourd'hui.

Soit $\Delta(x_{it}, \tilde{\varepsilon}_{it}) = \Delta^*(x_{it}) + (\varepsilon_{S,it} - \varepsilon_{W,it})$. Notons que Δ présente l'avantage d'être linéaire en $v = \varepsilon_{S,it} - \varepsilon_{W,it}$, si bien que Δ^* représente bien l'espérance

de la valeur nette de rester dans le système scolaire. Ceci provient du fait que nous avons supposé les chocs $\tilde{\varepsilon}$ i.i.d.: la valeur actuelle de ceux-ci n'apporte donc aucune information, ni sur les futurs chocs, ni sur les futures décisions qui en dépendront. Ainsi, les termes en espérance des équations (2) et (3) ne dépendent pas du choc courant $\tilde{\varepsilon}_{it}$. Du fait de cette linéarité, on écrit la valeur d'option (second terme de l'équation (4)) de la manière suivante:

$$\beta E(E[\max\{\Delta^*(x_{it+1}) + (\varepsilon_{S,it+1} - \varepsilon_{W,it+1}), 0\} | x_{it+1}] | x_{it}, d_{it} = S)$$

On notera $R(x) = E_v(\max\{x + v, 0\})$, où $v = \varepsilon_{S,it} - \varepsilon_{W,it}$. Comme les chocs sont i.i.d., on peut alors réécrire (4) comme :

$$\begin{aligned} \Delta^*(x_{it}) &= u_S(x_{it}) - u_W(x_{it}) + \beta E(R(\Delta^*(x_{it+1})) | x_{it}, d_{it} = S) \\ &+ \beta (E(V_W(x_{it+1}, \tilde{\varepsilon}_{it+1}) | x_{it}, d_{it} = S) - E(V_W(x_{it+1}, \tilde{\varepsilon}_{it+1}) | x_{it}, d_{it} = W)) \end{aligned} \quad (5)$$

La décision de choisir l'état de valeur maximale donne la probabilité que l'agent choisisse de rester à l'école:

$$p(x_{it}) = F(\Delta^*(x_{it}))$$

où F se déduit de la fonction de répartition de $-v = -(\varepsilon_{S,it} - \varepsilon_{W,it})$. Si F est supposée strictement croissante, elle est inversible. Si bien qu'on peut écrire $\Delta^*(x_{it}) = F^{-1}(p(x_{it}))$. En remplaçant $\Delta^*(x_{it})$ par sa valeur dans l'équation (5), et en manipulant cette dernière, on obtient :

$$u_S(x_{it}) = U_S(x_{it}) + C_S(x_{it}) \quad (6)$$

où:

$$U_S = F^{-1}(p(x_{it})) - \beta E[R(F^{-1}(p(x_{it+1}))) | x_{it}, d_{it} = S] \quad (7)$$

$$C_S = u_W(x_{it}) \quad (8)$$

$$+ \beta (E(V_W(x_{it+1}, \tilde{\varepsilon}_{it+1}) | x_{it}, d_{it} = W) - E(V_W(x_{it+1}, \tilde{\varepsilon}_{it+1}) | x_{it}, d_{it} = S))$$

Cette expression décompose u_S en deux termes U_S et C_S , qui peuvent s'interpréter simplement. U_S est la différence entre l'utilité intertemporelle d'un agent choisissant d'être scolarisé pour la période courante, puis de travailler après, et celle d'un agent choisissant de travailler dès la période courante. U_S est donc *le gain net actualisé de rester à l'école une seule année de plus*. Le second terme, C_S , est somme de deux termes. Le premier représente l'utilité de travailler pendant la période courante. Le second correspond au gain futur de commencer à travailler dès la période présente au lieu d'attendre la période suivante (en terme d'expérience par exemple). En d'autres termes, C_S traduit le *coût d'opportunité actualisé de rester à l'école une seule année de plus*.

3.3 Méthode d'estimation

Le gain actualisé de continuer les études une seule année de plus, U_S , est juste identifié non paramétriquement, pourvu que l'on connaisse la distribution $F(\cdot)$ et le facteur d'escompte β . Ce résultat est établi dans un article compagnon (Magnac et Thesmar, 2000), qui propose aussi une méthode d'estimation simple. Celle-ci est basée sur le fait que (7) peut se réécrire:

$$U_S(x_{it}) = E \{ F^{-1}(p(x_{it})) - \beta R(F^{-1}(p(x_{it+1}))) | x_{it}, d_{it} = S \} \quad (9)$$

qui est une équation de moment. Dans un premier temps, on estime de la manière la plus flexible possible - ou même non paramétriquement - la probabilité $p(x_{it})$ de demeurer à l'école pour chaque agent. Si l'on connaît F et β , il est alors possible de calculer, pour les agents qui choisissent de rester scolarisés, la quantité:

$$Y_{it} = F^{-1}(\widehat{p}(x_{it})) - \beta R(F^{-1}(\widehat{p}(x_{it+1}))) \quad (10)$$

Pour estimer (9), il suffit alors de regresser de manière aussi flexible que possible la quantité Y_{it} sur x_{it} dans cet échantillon (c'est à dire tels que $d_{it} = S$) pour

obtenir $U_S(\cdot)$. Cette méthode suppose de connaître la loi des $\varepsilon_S - \varepsilon_W$ et le paramètre β mais n'impose aucune restriction sur la fonction U_S .

Une fois U_S estimée, il reste à estimer C_S pour obtenir u_S (équation(6)). Or, le coût d'opportunité C_S s'obtient à partir des rendements du travail u_W (equation (8)). Pour simplifier, nous supposons plus bas que u_W ne dépend que du salaire espéré (c'est à dire pondéré de la probabilité de travailler), qui lui même est une fonction du niveau d'éducation E_{it} et de l'âge A_{it} de l'agent:

$$u_W(x_{it}) = E(w_{it}|E_{it}, A_{it})$$

que l'on peut estimer à partir des données de l'enquête Emploi (voir section 2.2.1).

3.4 Facteurs d'offre et facteurs de demande

Ce modèle permet de mettre à plat les déterminants des choix de scolarisation en incorporant les divers éléments introduits dans l'analyse descriptive. L'impact des salaires et des probabilités de chômage est mesuré par le facteur de demande u_W . Le coût implicite d'éducation (premier facteur d'offre) est contenu dans l'utilité u_S et la sélectivité du système (second facteur d'offre) est représenté par la probabilité de transition $\Pr(x_{it+1} | x_{it}, d_{it} = S)$. La méthode d'estimation ci-dessus nous permet de déduire l'évolution du coût de l'éducation u_S de (1) la variation des décisions d_{it} (2) de la sélectivité, et (3) de l'évolution des salaires (u_W).

Remarquons toutefois que même si u_W est stable au cours de la période (section 2.2.1), on ne peut se contenter de regarder l'évolution de U_S pour en déduire celle de u_S . En effet, le coût actualisé d'opportunité C_S dépend non seulement des salaires u_w mais aussi de la sélectivité $P(x_{it+1}|x_{it}, d_{it})$. Pour le voir, supposons que $u_W(E_{it}, A_{it})$ est stationnaire (ce que nous avons démontré à

la section précédente). C_S s'écrit alors:

$$C_S(x_{it}) = u_W^*(E_{it}, A_{it}) + \beta [E(V_W(E_{it+1}, A_{it} + 1)|x_{it}, d_{it} = W) - E(V_W(E_{it+1}, A_{it} + 1)|x_{it}, d_{it} = S)]$$

Soit $\pi(x_{it}) = P(E_{it+1} = E_{it} + 1|x_{it}, d_{it} = S)$ la probabilité, exogène pour l'agent, d'acquérir un niveau d'éducation supérieur s'il reste dans le système scolaire.

Alors, si cette incertitude est la seule qui affecte les salaires futurs:

$$C_S(x_{it}) = u_W^*(E_{it}, A_{it}) + \pi(x_{it}) \cdot \beta [V_W(E_{it}, A_{it} + 1) - V_W(E_{it} + 1, A_{it} + 1)]$$

où $\pi(x_{it})$ représente la sélectivité du système éducatif qui peut varier au cours du temps. Ainsi, C_S entremêle facteurs d'offre et de demande. Ce constat est important, dans la mesure où il rappelle que l'on ne peut se contenter d'observer l'évolution de U_S pour en déduire celle de u_S . En effet, même si les salaires sont stationnaires, la fonction C_S variera avec la sélectivité $\pi(\cdot)$, qui, nous l'avons montré, a changé au cours de la période.

4 Estimation et résultats

Telle que nous l'avons décrite, notre méthode d'estimation se décompose en trois étapes. En premier lieu, on estime un modèle dichotomique pour décrire les probabilités de continuer ses études ($p(x_{it})$ dans les termes du modèle). En deuxième lieu, on utilise les équations de Bellman pour estimer le gain de rester une année supplémentaire dans le système scolaire (U_S). Finalement, on utilise des estimations de fonctions de salaire obtenues par l'enquête Emploi pour estimer les coûts directs de l'éducation (u_S).

4.1 Déterminants de la probabilité de continuer dans le système scolaire

La première étape de notre méthode d'estimation consiste en l'estimation du modèle dichotomique :

$$d_{it} = S \iff Z_{it}\gamma + u > 0$$

où Z_{it} est un ensemble de variables constituées par les variables de x_{it} et par des interactions entre celles-ci. L'histoire x_{it} est constituée de l'année de naissance, du niveau scolaire, du nombre de redoublements et des variables de contexte socio-familial comme la taille de la fratrie, l'éducation et le revenu des parents et le travail de la mère. Bien que nous supposions la distribution de u logistique et que cette première étape soit réalisée par maximum de vraisemblance, cette estimation peut s'effectuer de manière très flexible. On interagira en particulier toutes les variables sociodémographiques et de cohorte avec les variables indicatrices des différents niveaux d'éducation⁸ et on utilisera des variables indicatrices pour chaque cohorte. Seuls les coefficients estimés des variables socio-démographiques par niveau d'éducation, sont reportés dans le tableau 3. Les estimations des effets-cohortes seront décrites plus bas.

Comme nous l'avions remarqué dans la section 2, ce sont les redoublements qui constituent le déterminant de la scolarisation le plus important. Le fait d'avoir redoublé au moins une fois accroît la probabilité de quitter l'école à tous les niveaux, sauf le plus élevé (à partir de la licence - probablement par manque d'observations). Le nombre de redoublements est aussi un facteur important, puisque le fait d'avoir redoublé plus d'une fois handicape la poursuite des études plus qu'un seul redoublement (quoique non significativement). Le salaire

⁸sauf aux niveaux les plus élevés au dessus de 21 ans puisque le nombre d'observations devient trop faible pour que des différences entre les effets des variables explicatives à différents niveaux d'éducation soient significatives

Niveau	15	16	17	18	19	20	>20
Mère au foyer	0.48 (0.47)	0.14 (0.32)	0.33* (0.19)	0.63* (0.33)	0.37 (0.49)	1.47** (0.40)	-0.74* (0.42)
Taille de fratrie	-0.12* (0.07)	-0.05 (0.05)	-0.05 (0.03)	-0.14** (0.06)	0.04 (0.10)	-0.04 (0.07)	0.01 (0.08)
Parents: Etudes supérieures	1.62** (0.71)	1.45** (0.39)	1.31** (0.23)	-0.42 (0.42)	-0.48 (0.68)	0.23 (0.40)	0.11 (0.64)
Parents: Lycée	1.02** (0.38)	0.83** (0.26)	0.21 (0.17)	-0.99** (0.40)	-0.76 (0.64)	-0.28 (0.37)	-0.24 (0.57)
Parents: log(salaire)	1.00 (0.64)	0.20 (0.43)	1.23** (0.24)	1.12** (0.38)	1.26** (0.65)	1.36** (0.49)	-0.12 (0.57)
1 Redoublement	-1.14** (0.46)	-0.60** (0.28)	-0.30** (0.15)	-0.76** (0.27)	-1.48** (0.58)	-0.99** (0.32)	0.10 (0.33)
>1 Redoublement	-1.94** (0.47)	-1.36** (0.27)	-0.68** (0.15)	-1.31** (0.26)	-2.07** (0.56)	-1.09** (0.31)	-0.37 (0.29)

Note: Les effets-cohortes estimés ne sont pas reportés ici. * signifie "significatif à 10%", et ** "significatif à 5%".

Table 3: Déterminants de la probabilité de continuer ses études

des parents a un effet positif sur la poursuite des études. Cet effet prend de l'ampleur avec le niveau scolaire comme dans les données américaines exploitées par Mare [1980]. Cet effet n'est significatif chez nous qu'aux niveaux les plus élevés (première, terminale et premier cycle universitaire). Aux niveaux les plus bas (troisième, seconde et première), c'est plutôt le niveau d'éducation que le revenu des parents qui semble importer. Ceci suggère que le soutien scolaire par les parents pourrait être déterminant dans les toutes premières années du lycée. La présence de la mère au foyer semble avoir un effet positif sur la poursuite des études, bien qu'il ne soit jamais significatif, sauf au niveau BAC+2. La taille de la fratrie semble agir négativement, mais le coefficient n'est significatif qu'au niveau 18 (terminale), comme si les parents arbitraient entre quantité et qualité des enfants. Au total, nos résultats sont très similaires à ceux obtenus par Mare [1980] et Cameron et Heckman [1998] sur données américaines⁹.

⁹Toutefois, comme l'ont montré Cameron et Heckman, 1998, ces estimations dépendent beaucoup de l'hypothèse d'absence d'hétérogénéité persistante entre individus et donc d'éventuels effets de sélection.

L'objet principal de cette étude consiste à comprendre la montée de la scolarisation. Afin d'isoler l'effet cohorte sur la scolarisation, nous avons calculé la probabilité de rester à l'école :

$$\widehat{p}(x_{it}) = \frac{1}{1 + e^{-Z_{it}\widehat{\gamma}}} \quad (11)$$

en fixant les variables sociofamiliales à des valeurs égales à leur moyenne dans la population sur tout l'échantillon. On prend ainsi en compte l'éventuelle évolution des autres déterminants de l'éducation sur la période, comme par exemple l'accroissement du niveau d'éducation des parents. Puis, nous avons retracé dans le graphique 8 l'évolution de la probabilité prédite par l'équation (11) pour les cohortes d'individus nés entre 1963 et 1973, et pour chaque niveau scolaire. On voit que le taux de scolarisation s'accroît de manière significative pour les niveaux terminale (18 ans) et surtout première (17 ans). L'évolution des déterminants socio-familiaux du taux de scolarisation n'a donc que peu de pouvoir explicatif de l'évolution sur la période. C'est donc dans la structure du modèle (coût de l'éducation, sélectivité ou salaires) qu'il faut chercher la raison de la montée de la scolarisation

4.2 Le gain actualisé de rester dans le système scolaire

Nous estimons maintenant le gain actualisé à étudier une seule année de plus U_S en construisant la variable donnée par l'équation (10). Pour ce faire, nous supposons que $\beta = 0.95$ et nous adopterons l'hypothèse que la fonction F est dérivée d'une loi logistique. On aura alors :

$$F^{-1}(p) = \log\left(\frac{p}{1-p}\right) \quad R(x) = \log(x) - \log(1-x)$$

Pour estimer U_S de manière flexible en utilisant (10), il suffit de régresser Y_{it} sur des variables Z_{it} construites à partir des variables de x_{it} , par la méthode des

moindres carrés ordinaires, comme on l’a expliqué dans la section précédente. On a:

$$Y_{it} = Z_{it}\hat{\gamma} - \beta \log [1 + \exp(Z_{it+1}\hat{\gamma})] = Z_{it}\delta + \nu_{it} \quad (12)$$

Le calcul de la variance de $\hat{\delta}$ doit néanmoins tenir compte du fait que γ est estimé en première étape, et d’une possible hétéroscédasticité (voir calculs de l’écart type en annexe). Les coefficients estimés $\hat{\delta}$ des variables liées au contexte socio-familial sont présentés dans le tableau 4. On remarquera, de façon générale, que les coefficients sont peu souvent significatifs. Le revenu des parents agit en général positivement et le niveau scolaire atteint par les parents accroît le rendement moyen de la poursuite des études, mais, de façon significative, aux niveaux d’éducation les plus bas seulement (16 et 17). La variable dénotant la présence de la mère au foyer est peu significative et change de signe très souvent. Le fait d’avoir redoublé, et le nombre de redoublements abaisse le rendement privé de l’éducation, suggérant que les redoublements sont un contrôle au moins partiel de l’hétérogénéité inobservée. Le nombre de frères et soeurs affecte négativement le gain d’une année supplémentaire, surtout aux niveaux d’éducation les plus bas.

Tel que nous l’estimons, l’effet de la cohorte sur le gain d’une année supplémentaire, U_S est difficile à interpréter: peu de tendances claires se dégagent. Pour chaque niveau scolaire, le graphique 9 en trace l’évolution moyenne pour les cohortes d’agents nés entre 1963 et 1973. Comme nous l’avons fait pour les probabilités de scolarisation, nous avons fixé la valeur des explicatives autres que la date de naissance à leur moyenne pour tout l’échantillon, afin de se concentrer sur les effets purs des cohortes. En général, l’évolution de U_S ne dévoile aucune tendance claire, à la hausse ou à la baisse au cours de la période essentiellement à cause de problèmes de significativité. Une exception est peut être le niveau ter-

Niveau	15	16	17	18	19	20	>20
Taille de Fratrie	-0.07 (0.07)	-0.02 (0.05)	0.03 (0.04)	-0.16 (0.08)	0.05 (0.09)	-0.02 (0.06)	-0.01 (0.03)
Mère au Foyer	0.30 (0.46)	-0.11 (0.30)	-0.16 (0.22)	0.29 (0.40)	-0.53 (0.43)	1.26 (0.32)	-0.24 (0.17)
Parents: Etudes Supérieures	0.32 (0.67)	0.36 (0.36)	0.90 (0.27)	-0.17 (0.56)	-0.55 (0.61)	-0.07 (0.32)	0.03 (0.25)
Parents: Etudes Secondaires	0.26 (0.37)	0.60 (0.25)	0.49 (0.21)	0.33 (0.52)	0.45 (0.57)	0.27 (0.30)	0.07 (0.24)
Log(Salaire parents)	0.72 (0.65)	-0.58 (0.39)	0.20 (0.27)	0.06 (0.49)	0.23 (0.59)	0.98 (0.45)	-0.04 (0.20)
1 redoublement	-0.57 (0.53)	-0.32 (0.31)	0.24 (0.23)	0.43 (0.48)	-0.78 (0.63)	-0.90 (0.38)	0.12 (0.22)
>2 redoublements	-0.76 (0.50)	-0.79 (0.28)	0.13 (0.20)	0.17 (0.45)	-1.32 (0.57)	-0.91 (0.30)	-0.18 (0.14)

Note: Coefficients du rendement actualisé d'une seule année d'étude supplémentaire. Estimation en deux étapes, tenant de compte de l'hétéroscédasticité. La dernière équation inclut deux indicatrices de BAC+4 et BAC+5. * signifie "significatif à 10%", et ** "singificatif à 5%".

Table 4: Rendement actualisé d'une seule année supplémentaire d'école

minale (18), qui semble connaître une hausse pour les toutes dernières cohortes (1971-1973).

Finalement, bien que le facteur β ne soit pas identifié -il faut le fixer pour calculer Y_{it} , les estimations des effets cohortes sont peu sensibles aux variations de sa valeur. Soulignons toutefois que l'effet des variables explicatives leur est sensible¹⁰.

Ainsi, la hausse de la scolarisation remarquée dans le graphique 8, qui portait sur les niveaux 17, 18 et 20 ne semble que très partiellement prise en compte par l'évolution de U_S . Il est néanmoins très difficile d'avancer une interprétation des variations de U_S en terme de u_S . En effet, malgré la stabilité des salaires sur la période, nous avons vu plus haut qu'il est nécessaire d'estimer C_S afin de tenir compte de l'évolution de la sélectivité du système éducatif au cours de la période.

¹⁰Nous ne commenterons pas ces résultats ici. Une version précédente de ce papier, disponible auprès des auteurs réalise ce "test" de sensibilité, pour β compris entre 0.80 et 0.99. Nous avons aussi évalué la robustesse de nos résultats au remplacement d'une loi logistique par une loi normale.

Finalement, l'estimation de U_S permet aussi d'avoir une idée du comportement de la valeur d'option de l'éducation au cours de la période. En effet, ce terme est directement identifié par les données:

$$\begin{aligned} E \{ \max \{ V_{S_{it+1}} - V_{W_{it+1}}, 0 \} | x_{it}, d_{it} = S \} &= E \{ \beta R(-F^{-1}(p(x_{it+1}))) | x_{it}, d_{it} = S \} \\ &= F^{-1}(p(x_{it})) + U_S \end{aligned}$$

Ainsi, la hausse de la scolarisation au niveau première pourrait avoir été déclenchée par l'accroissement de la valeur d'option liée à la participation au système éducatif à ce niveau. En effet, le terme U_S varie peu au niveau première, alors que $p(x_{it})$ s'accroît.

Au niveau terminale cependant, U_S s'accroît légèrement. Il s'agit donc de savoir si cette évolution reflète un accroissement du rendement instantané u_S , ou si elle traduit la moindre sélectivité du système éducatif à ce niveau (voir graphique 5). Pour conclure, nous avons vu qu'il est nécessaire d'estimer C_S , ce que nous faisons maintenant.

4.3 L'estimation des coûts d'éducation

Pour estimer l'évolution des coûts d'éducation sur la période, la non identification de la distribution des chocs sur les préférences, F , a une conséquence importante. En effet, il est bien connu que dans les modèles de choix discret, la variance des chocs n'est pas identifiée et est traditionnellement fixée à une valeur égale à 1. Ceci se traduit par le fait que l'échelle des fonctions valeurs, des fonctions d'utilité ou des fonctions composées, comme le gain de rester à l'école une année supplémentaire, U_S , n'est pas identifiée. Or, pour calculer les coûts de l'éducation donnés par l'équation (7), on utilise l'évaluation de U_S que l'on vient de faire et l'évaluation monétaire des gains sur le marché du travail, C_S , en utilisant les salaires perçus. Il faut donc trouver une restriction identifiante qui permette de

rendre cohérentes les échelles de ces deux termes U_S et C_S .

On supposera donc que F n'est connue qu'à un facteur d'échelle σ près. Nous supposerons en outre que σ peut dépendre de x_{it} , si bien que F s'écrit maintenant:

$$F(u|x_{it}) = G\left(\frac{u}{\sigma(x_{it})}\right) \text{ où } G \text{ est la fonction de répartition d'un logit simple}$$

L'équation de moment (9), définissant U_S est alors modifiée en pondérant ses différents termes par les termes de variance :

$$U_S(x_{it}) = \sigma(x_{it}).G^{-1}(p(x_{it})) - \beta.E [\sigma(x_{it+1}).R(-G^{-1}(p(x_{it+1})))|x_{it}, d_{it} = S] \quad (13)$$

On voit de manière claire dans l'équation (13) que σ "calibre" l'importance relative de U_S par rapport à U_W dans l'estimation de u_S . Il faut des hypothèses supplémentaires pour identifier $\sigma(\cdot)$ que nous explicitons plus bas. Nous précisons d'abord les hypothèses que nous faisons sur les gains de la sortie du système éducatif.

En suivant la littérature actuelle sur les investissements en capital humain (par exemple Keane et Wolpin, 1996), nous supposerons d'abord que u_W^* s'écrit:

$$u_W^*(x_{it}) = E(w_{it}|x_{it}, T_{it} = 1) .P(T_{it} = 1|x_{it})$$

où $T_{it} = 1$ si l'agent travaille, et 0 sinon. Cette définition de u_W^* revient à donner une utilité équivalente à zéro franc si l'agent ne travaille pas, négligeant les allocations chômage, et la possibilité de transferts familiaux (voir note 4).

Cette spécification rend aisée l'estimation de u_W^* . Nous nous appuyons sur l'extrême stabilité des rendements de l'éducation au cours de la période comme nous l'avons montré plus haut pour estimer u_W^* en n'utilisant que la vague de 1993 de l'enquête FQP. L'intérêt de cette démarche est que nous disposons dans FQP d'une nomenclature de diplôme plus détaillée, nous permettant d'estimer

le rendement salarial de l'éducation pour tous les niveaux entre 15 et 23.¹¹ Nous regressons la variable:

$$SAL_{it} = \begin{cases} w_{it} & \text{si } T_{it} = 1 \\ 0 & \text{si } T_{it} = 0 \end{cases}$$

sur des indicatrices de niveau scolaire au sens où nous l'avons défini précédemment, et sur des indicatrices d'expérience jusqu'à 25, puis sur un polynôme quadratique de 25 à 60. Ces coefficients nous permettent d'imputer à chaque individu de notre échantillon une utilité instantanée de travailler \widehat{u}_W^* , ainsi qu'une espérance de valeur actualisée \widehat{V}_W de travailler. C_S est maintenant identifiée.

Pour estimer U_S , il faut encore connaître σ . Afin de rendre les échelles de U_S et C_S comparables, nous supposons que $\sigma^2(x_{it}) = \text{var}(SAL_{it}|x_{it})$, ce qui revient à écrire:

$$\text{var}(\varepsilon_S - \varepsilon_W) = \text{var}(\varepsilon_W) \tag{14}$$

ce qui est vrai en particulier si l'on suppose qu'il n'y a pas de chocs de préférences sur les coûts d'éducation ($\varepsilon_S = 0$). Nous estimons $\sigma(\cdot)$ conjointement à $u_W(\cdot)$ par une procédure de pseudo-maximum de vraisemblance. Nous utilisons alors l'équation (6) pour estimer u_S .

Les effet-cohortes des u_S ainsi estimés sont décrits dans le graphique 10. Les rendements ainsi estimés sont faiblement significatifs (les écarts types sont importants) et très instables. La restriction (14) n'est donc pas très identifiante, et il faudrait donc peut être lui en substituer une autre, mieux fondée théoriquement.

Afin d'améliorer la précision de nos estimateurs, nous avons aussi estimé un modèle où $u_W(x_{it}) = \log(w_{it})$ ¹². En plus de tenir compte d'une potentielle aversion pour le risque, cette méthode permet de supposer légitimement que ε_W

¹¹Ce qui n'est possible dans l'enquête Emploi que pour les niveaux 14,15,18,20,21 et 23, comme nous l'avons expliqué dans la section 2.

¹²Les paramètres β de $\log w = X\beta + \varepsilon$ sont stables sur la période 80-93.

est homoscedastique, si bien que $\sigma^2(x_{it}) = \sigma^2$. Malgré les gains de précision obtenus, les coûts subjectifs demeurent toujours stables sur la période.

Au total, quelle que soit la prise en compte de l'aversion pour le risque dans ce modèle, nos résultats suggèrent qu'une fois prise en compte l'évolution de la sélectivité au cours de la période, les rendements instantanés de l'éducation n'ont pas évolué au cours de la période. Il semble donc que c'est seulement les valeurs d'option qui aient évolué au long de cette période. Dit autrement la seule grandeur qui s'est accrue au cours de la période est la valeur d'option de demeurer dans le système éducatif en première et terminale, moins à cause de la réduction des coûts de scolarisation, que grâce à une plus grande permissivité du système, en particulier dans l'enseignement supérieur.

5 Conclusion.

Nous avons étudié les causes de la montée de la scolarisation en France. L'analyse descriptive réalisée dans la section 2.2 suggère que les facteurs liés à la demande (montée du chômage des jeunes et des moins diplômés) n'auraient pas eu un impact significatif. C'est pourquoi nous nous tournons vers les facteurs liés à l'offre d'éducation (baisse de la sélectivité du système éducatif, abaissement des coûts de l'éducation). Nous trouvons que ces facteurs pourraient avoir joué au cours de la période. L'évaluation de leur importance respective nécessite de construire un modèle où ces facteurs déterminent séparément la décision de quitter, ou non, l'école. Cette ambition requiert l'estimation d'un modèle dynamique de choix discret. Un tel modèle est estimé sur les données de l'enquête FQP, à l'aide d'une méthode originale, inspirée de Hotz et Miller[1993]. Les hypothèses identifiantes sont fortes. Conditionnellement à celles-ci, il semble néanmoins possible de conclure que la baisse de la sélectivité du système éducatif est la principale cause de

la montée de la scolarisation pour les jeunes hommes nés entre 1963 et 1973.

L'étude de la sensibilité des résultats à certaines hypothèses d'identification peut constituer un prolongement utile de ce travail. La prise en compte de l'hétérogénéité individuelle dans notre cadre d'analyse constitue un autre prolongement de notre étude, car en l'état, nos estimateurs peuvent être entâchés d'un biais de sélection important (Cameron et Heckman [1998]). Une telle estimation permettrait de plus de rendre crédible une utilisation de nos paramètres structurels pour simuler l'effet de politiques économiques (voir Keane et Wolpin [1996]).

Finalement, nous pensons que ce travail, en distinguant clairement deux facteurs d'offre (sélection et coût d'éducation), soulève la question de l'optimalité d'une politique éducative qui marierait ces deux instruments. Cette question n'est pas, à notre connaissance résolue théoriquement, et constitue à nos yeux un enjeu important pour la définition d'une politique d'éducation.

6 Bibliographie

Bayet, A., et C., Cases, 1996, "Earnings Inequality in France: evolution 1967-1991", INSEE, WP9617.

Becker, G., 1975, *Human Capital*, Chicago UP: Chicago.

Belzil, C., et J., Hansen, 1999, "Estimating the Returns to Education from a Non-Stationary Dynamic Programming Model", ISA WP, Bonn.

Cameron, S.V., et J.J., Heckman, 1998, "Life-Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males", *Journal of Political Economy*, 106:262-333.

Conway, Y., A., Melnik et M.A., Pollatschek, 1973, "The Option Value of Education and the Optimal Path for Investment in Human Capital", *International Economic Review*, 421-435;

Coslett, S.R., 1994, "Estimation from Endogeneous Stratified Samples" in eds Maddala, Rao and Vinod, *Handbook of Statistics*, North Holland: Amsterdam, 11:1-41.

Eckstein, Z., et K., Wolpin, 1989, "The Specification and Estimation of Dynamic Discrete Choice Models", *Journal of Human Resources*, 24:562-98.

Elbaum, M. et O., Marchand, 1994, "Emploi et Chômage des Jeunes dans les Pays Industrialisés: la Spécificité Française", *Travail et Emploi*, 58:111-121.

Emin, J.C., et P., Fallourd, 1996, "L'état du système éducatif français", eds INSEE, *Données Sociales*, INSEE:Paris, 40-47.

Estrade, M.A., et C., Minni, 1996, "La Hausse du Niveau de Formation", *INSEE Première*, 488:55-58.

Goux, D. et E., Maurin, 1994, "Education, Experience et Salaires: Tendances Récentes et Evolutions de Long Terme", *Economie et Prévision*, 116:155-

173.

Heckman, J.J., R. LaLonde et J. Smith, 2000, "The economics and econometrics of active labor market programs" in Ashenfelter et Card, *Handbook of Labour Economics*, IIIA, North Holland.

Heckman, J.J., L. Lochner et C. Taber, 1998, "General-Equilibrium Treatment Effects: A Study of Tuition Policy", *American Economic Review*, 88:381-385.

Hotz, V.J., et R.A. Miller, 1993, "Conditional Choice Probabilities and the Estimation of Dynamic Models", *Review of Economic Studies*, 60:497-529.

Keane, M. et K. Wolpin, 1996, "The Career Decisions of Young Men", *Journal of Political Economy*, 105:473-522.

Kodde, D.A., 1988, "Unemployment Expectations and Human Capital Formation", *European Economic Review*, 32:1645-1660.

Magnac, T., et D. Thesmar, 2000, "Identifying Dynamic Discrete Decision Processes", à paraître dans *Econometrica*, disponible sur le site

[<http://www.inra.fr/Internet/Departements/ESR/UR/lea/wp/wp0002.pdf>]

Manski, C.F., 1993, "Dynamic Choices in Social Settings", *Journal of Econometrics*, 58:121-136.

Mare, R.D., 1980, "Social Background and Social Continuation Decisions", *Journal of the American Statistical Association*, 75:295-305.

Micklewright, J., M. Pearson, S. Smith, 1989, "Unemployment and Early School Leaving", IFS London.

Prost, A.S. (1998), "L'éducation Nationale depuis la Libération", *Cahiers Français*, 285, La Documentation Française: Paris.

Rust, J., 1994, "Structural Estimation of Markov Decision Processes" in eds Engle et McFadden, *Handbook of Econometrics*, North Holland:Amsterdam, IV:

3081-143.

Wasmer, E., 1998, "Some Labor Market Consequences of Demographic Trends: US and France 1964-97, CEP, LSE.

A Calcul de l'écart type de l'estimateur

L'objet de cette annexe est de faire le calcul de l'écart type des estimateurs de U_S . Notons Z la matrice formée par les Z_{it} empilés, déterminants de l'utilité (équation (12)). Dans la regression de seconde étape, la variable dépendante vaut:

$$\hat{Y}_n = Z\hat{\gamma}_n - \beta \ln(1 + e^{Z_1\hat{\gamma}_n})$$

où Z_1 est la matrice constituée des variables d'état futures pour l'observation n (Z_{it+1} empilés, cf équation (12)) Nous recherchons δ qui satisfasse:

$$\hat{Y}_n = Z\delta + \eta \quad (15)$$

Un estimateur naturel de δ s'en déduit aisément:

$$\hat{\delta}_n = (Z'Z)^{-1}Z'\hat{Y}_n \quad (16)$$

Par ailleurs, $\hat{\gamma}_n$ converge vers γ . Un développement de Taylor nous conduit à:

$$Z'\hat{Y}_n \# Z'Y + Z'\frac{\partial Y}{\partial \gamma}(\hat{\gamma}_n - \gamma) \quad (17)$$

où:

$$Z'\frac{\partial Y}{\partial \gamma} \# Z'Z \left(1 - \frac{\beta}{1 + e^{-Z_1\hat{\gamma}_n}}\right)$$

Nous combinons (17), (15) et (16) pour obtenir:

$$\hat{\delta}_n - \delta = (Z'Z)^{-1}Z'\eta + (Z'Z)^{-1}Z'\frac{\partial Y}{\partial \gamma}(\hat{\gamma}_n - \gamma)$$

On montre facilement que η et $(\hat{\gamma}_n - \gamma)$ sont asymptotiquement indépendants. Par définition, η est une fonction de Z_{it} et Z_{it+1} , alors que $(\hat{\gamma}_n - \gamma)$ est construit de manière à être indépendant de Z_{it} et Z_{it+1} . Notant V_γ la matrice de variance

covariance asymptotique de $\hat{\gamma}_n$, la matrice de variance covariance $\hat{\delta}_n$ peut s'écrire:

$$\begin{aligned} V\hat{\delta}_n &= (Ez'_iz_i)^{-1}E(z'_i\eta_i^2z_i)(Ez'_iz_i)^{-1} + \\ &\quad (Ez'_iz_i)^{-1}E\left(z'_i\frac{\partial y_i}{\partial \gamma}\right)V_\gamma E\left(\frac{\partial y_i}{\partial \gamma'}z_i\right)(Ez'_iz_i)^{-1} \end{aligned}$$

où le premier terme corrige pour la présence d'hétéroscédasticité en suivant la méthode de White, et le second terme tient compte du fait que γ est estimé dans une première étape.