De l'utilité de s'éduquer : remise en question de la linéarité du rendement de l'éducation sur le salaire en France

Mattias Mano

Université Paris1 – UFR 02 Sciences Economiques

Master 2 Recherche Economie Théorique et Empirique

De l'utilité de s'éduquer

Remise en question de la linéarité du rendement de l'éducation sur le salaire en France

Abstract: The rate of return of education on wage is a well-known problem in economics since the 60's with the works of G. BECKER and J. MINCER. The present master thesis discusses one of the main assumption of Mincer equation: the linearity of the rate of return of education. Working on the FQP 2003 dataset of the INSEE, we show that with an appropriate specification, an educational year leading to a degree has a higher rate of return than another without diploma.

Key words: educational economics, human capital, non-linear rate of return, equation of Mincer,

Dirigé par Philippe GAGNEPAIN
Présenté et soutenu par Mattias MANO
mattias.mano@gmail.com

2012-2013

L'université de Paris 1 Panthéon Sorbonne n'entend donner aucune approbation, ni désapprobation aux opinions émises dans ce mémoire ; elles doivent être considérées comme propre à leur auteur.

Remerciements

Ce mémoire de Master est le résultat de mon travail de recherche effectué durant mon Master 2 Economie Théorique et Empirique. Je souhaite adresser tous mes remerciements aux personnes qui m'ont apporté leur aide et ont ainsi contribué à son élaboration.

De grands remerciements à Monsieur Philippe GAGNEPAIN, directreur de recherche de ce mémoire, pour son aide précieuse et pour le temps qu'il a bien voulu me consacrer.

Je remercie également les professeurs de l'Université Paris I-Sorbonne et de l'ENSAE qui ont su me diriger vers les personnes qualifiées et me donner des références essentielles, particulièrement Messieurs Alain TROGNON, Marc GURGAND et Antoine TERRACOL.

Enfin, j'adresse mes plus sincères remerciements à mes parents et tous mes proches et amis, qui m'ont soutenu et toujours aidé et encouragé au cours de la réalisation de ce travail.

Table des matières

1 Introduction

La figure 2 [1] (8.1 page 34) donne le salaire médian (2009) correspondant au niveau d'éducation des Français. Nous pouvons constater immédiatement une corrélation positive entre le niveau d'éducation et le salaire médian. Ce constat constitue le point de départ de notre analyse. En effet, il semble que l'acquisition d'un niveau supérieur d'études permet une augmentation du salaire. Du moins, c'est l'hypothèse de base, centrée autour du capital humain, développé par G. Becker [8] [2] dans les années 1960.

Le modèle initial permettant d'établir un lien entre le revenu - le salaire - et le niveau de formation [3] a été établi par J. Mincer [19] au cours de la même décennie. Nous aurons l'occasion d'approfondir ce modèle plus en détails mais l'une des hypothèses majeures est la linéarité du lien entre ces deux variables : le gain sur le salaire sera équivalent en fin de licence comme à l'issue d'une maîtrise. La figure 3 (8.1 page 34), reposant sur les mêmes données que le graphique précédent, donne les écarts de salaires entre deux niveaux d'éducation consécutifs [4]. Exception faite du niveau "Grandes Ecoles", nous constatons immédiatement que l'évolution financière est toujours positive. Si l'on suppose qu'un étudiant effectue le cursus baccalauréat - licence - master - doctorat, ses revenus augmenteront régulièrement à chaque étape (120€ - 280€ - 270€ – 540€), excepté donc entre licence et master, étape pour laquelle la différence de gain s'élève à 10€ de moins qu'entre licence et baccalauréat. Ce simple histogramme tend à rendre inappropriée l'hypothèse du modèle de Mincer puisque sous cette dernière, l'écart entre la licence et le baccalauréat (280€) devrait être le même que celui entre le master et le doctorat (540€) car le même nombre d'années sépare ces niveaux d'études. En outre, comme le définit C. Van de Velde [24], en France, le diplôme joue un rôle proche de la "tyrannie" pour les jeunes arrivant sur le marché du travail. Une année non diplômante (telle qu'une deuxième année de licence, une première année de DUT...) ne devrait pas induire une évolution notoire du revenu, s'il s'agit de la dernière année d'éducation. Le taux de rendement semble effectivement varier en fonction de l'année validée.

Il semblerait donc que le modèle linéaire liant le revenu au niveau d'éducation ne soit pas le plus adapté à cette réalité. Dans ces conditions, comment prendre en compte cette fluctuation du rendement de l'éducation ? Quelle spécification permettrait de modéliser ce qui semble être un état de fait en France ? Si un tel modèle existait, il remettrait en

1. Pour des raisons de clarté, l'ensemble des tableaux et graphiques seront présentés en annexes.
2. Pour plus de détails, se reporter à la bibliographie.
3. Dans la présente étude, le niveau de formation et le niveau d'éducation se réfèrent au nombre d'années d'études effectué.
4. Sans diplôme ou uniquement brevet des collèges avec un niveau CAP (2-1) ; avec le diplôme baccalauréat (3-1) ; la différence entre le baccalauréat et niveau bac+2 (4-3) ; avec licence (5-3) ; différence entre licence et master (6-5) ; grande école et licence (7-5) ; et enfin différence entre doctorat et master (8-6).

cause le modèle initiateur de Mincer dans lequel le taux de rendement est le même quel que soit le niveau d'études validé.

Même si une simple analyse graphique n'apporte pas une preuve tangible, de nombreux travaux, depuis les années 1960, comme nous le verrons par la suite, ont développé des modèles consistant à cette spécification. Le présent travail va tendre à apporter dans ce débat, de nouveaux arguments favorables à ce point de vue, en analysant les données de l'enquête Formation et Qualification Professionnelle (FQP) 2003 de l'Institut national de la statistique et des études économiques (INSEE - FQP 2003) [5].

Après un rappel des grandes théories couvrant cette question dans la section 2, la section 3 présente les modèles importants de la littérature. Les sections 4 et 5 exposent les données, les variables ainsi que les modélisations utilisées. La section 6 correspond à l'analyse des résultats pour conclure par la section 7.

5. Les données ont été obtenues auprès du Centre Maurice Halbwachs

2 Cadre théorique

Les travailleurs les plus formés et les plus expérimentés tendent à avoir des salaires plus élevés. Depuis les années 1960, l'explication principale donnée à cette corrélation est que le temps passé à s'éduquer ou à acquérir de l'expérience, augmente directement la productivité du travailleur : cette relation illustre l'idée essentielle de la *théorie du capital humain*, développé par G. Becker dans son célèbre ouvrage de 1964 [8].

Cependant cette relation n'explique pas à elle seule l'ensemble des différences salariales associées aux expériences scolaires et professionnelles des individus. Par exemple, il semblerait que les employés plus instruits aient moins tendance à s'absenter ou à abandonner leur emploi. Les employeurs préfèrent donc recruter les employés ayant ce profil. Mais, en réalité, de telles informations ne sont pas disponibles a priori. Les employeurs peuvent alors utiliser le niveau d'éducation comme critère principal pour sélectionner les candidats et tenter de limiter le risque d'abstentéisme. C'est la raison pour laquelle les étudiants vont privilégier un certain niveau de formation comme *signal* de leur sérieux à l'attention des employeurs. Ce signal permet aux employeurs d'apprécier les employés potentiels en obtenant des informations sur leurs capacités inobservables, rendues visibles par ce dernier. Ce raisonnement s'appuie sur *la théorie du signal*, seconde grande théorie que nous allons présenter.

2.1 Théorie du capital humain

La théorie du capital humain est principalement une théorie de la "demande individuelle d'éducation", comme l'explicite M. Gurgand dans son ouvrage *Economie de l'éducation* [14]. Cependant, depuis les néoclassiques, le comportement individuel optimal doit aussi l'être socialement : si l'investissement est profitable pour l'individu, il doit aussi l'être pour la société. L'idée principale de cette théorie est que des individus ont des salaires plus importants parce qu'ils ont une plus grande "productivité marginale" due à un plus grand niveau d'instruction.

Formellement, le salaire est une fonction du nombre d'années d'études : $w(S)$. Si $F(K, S)$ est la fonction de production, dans un cadre néoclassique, nous avons :

$$F'_S(K, S) = w(S)$$

Dans un marché financier parfait, hypothèse qui sera discutée par la suite, seule la richesse intertemporelle est significative pour l'utilité de l'individu et non pas sa distribution dans le temps :

$$\begin{cases} U(c_1, c_2) \\ \text{s. c. } c_1 + \frac{c_2}{1+r} = w_1 + \frac{w_2}{1+r} \end{cases}$$

Mais avec l'hypothèse de marché parfait, la contrainte est remplacée par deux nouvelles contraintes : $c_1 = w_1$ et $c_2 = w_2$. La solution du programme révèle une importante propriété : les décisions de consommation et de revenu sont dissociables.

Désormais, oublions l'individu consommateur et étudions l'agent investisseur. Pour choisir son nombre d'années d'études optimal S, l'agent sait qu'il paye des frais de scolarité f et qu'il subit un coût d'opportunité : le salaire qu'il ne gagne pas en se formant. A l'équilibre nous avons :

$$\frac{w'(S)}{r} = w(S) + f,$$

avec le taux d'intérêt r.

$w(S) + f$ représente le coût marginal (direct et d'opportunité) et $\frac{w'(S)}{r}$ le retour marginal, c'est-à-dire le montant qu'un individu va obtenir (sur l'ensemble de sa vie) s'il reste un temps défini supplémentaire dans un cursus éducatif.

2.2 Théorie du signal

Dans les années 1970, l'hypothèse de perfection des marchés financiers est remise en cause par les nouvelles théories économiques. Les modèles appliqués à l'éducation et au marché du travail n'y échappent pas. Dans la théorie du capital humain, nous supposions que la productivité des agents est observable. Ce qui est impossible dans la réalité : l'employeur ne peut pas connaître à l'avance la productivité de chaque employé potentiel. Mais si nous maintenons l'hypothèse que le nombre d'années d'études influence la productivité, alors les agents peuvent choisir de valoriser leur formation, dans leur C.V. par exemple, afin de donner une indication sur leur productivité potentielle.

L'hypothèse d'imperfection sur le marché du travail donne naissance à une nouvelle théorie : *la théorie du signal*. Développée par Spence [22], Arrow [4] notamment, cette théorie part du postulat suivant : l'éducation peut entraîner des salaires plus élevés sans impliquer une productivité individuelle plus importante.

L'objectif d'un employeur est de recruter les individus les plus productifs. S'il s'avère, que statistiquement parlant, les employés les plus productifs sont les plus diplômés, alors l'employeur va leur donner un meilleur salaire, sans se soucier de savoir si c'est bien l'école

qui détermine le niveau de productivité. Les individus sont donc incités à augmenter leur niveau d'éducation pour accéder à ces hauts salaires. La question qui en découle est : pourquoi les individus ne s'éduquent-ils pas tous autant que possible ? S'ils avaient tous le niveau de formation le plus élevé, il n'y aurait alors plus de différence repérable entre les individus selon leur productivité et le diplôme perdrait son rôle de signal. Mais il faut également prendre en compte le coût relatif - l'effort - à s'éduquer pour les individus : si les plus productifs sur le marché du travail sont aussi les meilleurs à l'école, ce coût sera plus important pour les moins productifs. Par exemple, ils peuvent mettre deux fois plus de temps pour atteindre un même niveau d'études qu'un individu avec une forte productivité initiale, ce qui augmente fortement ce coût et ne rend pas intéressant pour eux la démarche nécessaire pour s'éduquer.

C'est sur ce point que les deux théories s'opposent : alors que la théorie du capital humain suppose que l'augmentation du nombre d'années d'études augmente la productivité individuelle, la théorie du signal suppose que la productivité individuelle est exogène (un individu naît avec un niveau de productivité donné). C'est pourquoi un individu ayant une forte productivité initiale fera d'avantage d'études pour se différencier des individus moins dotés en productivité. Ainsi, l'employeur qui cherche à embaucher les plus productifs payera en fonction du nombre d'années d'études réalisé. *In fine*, le nombre d'années d'études agit comme un *signal* entre les employeurs et les salariés potentiels.

Notons qu'une augmentation du niveau d'éducation de la société peut ne pas être socialement optimale. En effet, comme le démontre Baudelot et Glaude [7], il existe en France un déclassement des diplômes : "Le déclassement du niveau Licence-Maîtrise est ici très net : alors que près d'un jeune sur deux titulaires de ce diplôme figurait, en 1970, parmi les 10 % les mieux payés, moins d'un sur trois se trouve dans ce cas en 1985." Ce déclassement s'accompagne d'une augmentation de l'âge moyen de sortie de l'école ce qui implique directement une augmentation des effectifs scolaires et donc une augmentation des diplômés. Il existe ainsi un réel problème quant à la démocratisation de l'éducation : d'une part, elle permet à tous d'acquérir une connaissance commune, ce qui améliore la productivité totale d'un pays ; d'autre part, elle crée une plus grande concurrence au niveau de la demande de travail ce qui entraîne, après un ajustement possible, une augmentation du taux de chômage puisque l'offre d'emploi ne change pas.

3 Modélisation du rendement de l'éducation

Bien que les deux théories de l'éducation paraissent cohérentes avec les observations empiriques (elles diffèrent seulement dans l'interprétation des résultats), nous sommes amenés à choisir un modèle de référence. Rappelons que dans sa version la plus extrême,

la théorie du signal considère que les individus ont une dotation de productivité innée et ce indépendamment de leur parcours scolaire. Ainsi, l'école n'ajouterait rien à leurs capacités. Ce qui signifie que l'éducation n'est pas un investissement rentable pour la société. Cette vision peu optimiste du système éducatif nous mène à nous orienter vers une approche intégrant les hypothèses de la théorie du capital humain. De plus, une certaine incohérence de la théorie du signal avec la réalité nous interpelle : de nombreux économistes comme E. Maurin ou D. Goux ont démontré qu'il existe une certaine reproduction sociale, *i.e.* qu'un enfant dont les parents sont éduqués a une plus grande chance d'atteindre un niveau élevé d'éducation et bien entendu, indépendamment de sa productivité initiale. A moins de supposer que les moins riches, les PCS les moins élevées impliquent une productivité initiale moins élevée - ce qui nous semble impossible à penser - la théorie du signal ne semble pas en adéquation à la réalité.

Cependant, la modélisation du rendement de l'éducation pose de nombreux problèmes. Ashenfelter, Harmon et Oosterbeek [5] mettent en avant l'existence de biais de déclaration - "reporting bias" - et de "file drawer". De plus, ils définissent assez clairement les différentes méthodes usuelles d'estimation du rendement de l'éducation par la régression des moindres carrés ordinaires (MCO). La présente section détaille ces méthodes en les complétant par d'autres modèles récurrents dans la littérature.

3.1 L'équation de Mincer et le problème de l'estimation par MCO

3.1.1 Fonction simple de rendements

Dans ses articles fondateurs de 1958 et de 1974, J. Mincer [19] établit les bases économétriques du lien existant entre le salaire et le nombre d'années d'études. Ancré dans la théorie du capital humain, Mincer considère qu'il existe deux moyens d'en acquérir : l'éducation (mesurer par le nombre d'années d'études) et l'expérience professionnelle (mesurer par le nombre d'années d'activité professionnelle).

En effet, Mincer considère que le cycle de vie est divisé en deux temps : une période durant laquelle l'individu s'éduque (et y consacre tout son temps) et une période durant laquelle il travaille qui commence à la fin de la précédente et finit à l'âge de la retraite.

Concernant la période d'éducation, Mincer pose certaines hypothèses : elle est homogène - c'est-à-dire que toutes les années d'études apportent le même niveau de connaissances (en qualité et en quantité). De plus, il n'existe pas de dépréciation du capital humain.

Ce modèle implique que les individus acquièrent un savoir négociable sur le marché du travail quand ils terminent leurs études. Conscient de cette incohérence, Mincer introduit une nouvelle variable pour prendre en compte le fait que les individus continuent

d'investir dans leur capital humain pendant qu'ils travaillent : il s'agit de l'expérience professionnelle.

Alors qu'avec l'unique variable "éducation", le salaire est constant dans le temps, l'ajout de la notion d'expérience professionnelle donne un salaire concave (cf. Figure 1). La croissance positive du salaire s'explique d'abord par la relation positive qui le lie avec l'expérience. Ensuite sa concavité est due au rendement décroissant de l'expérience sur le salaire.

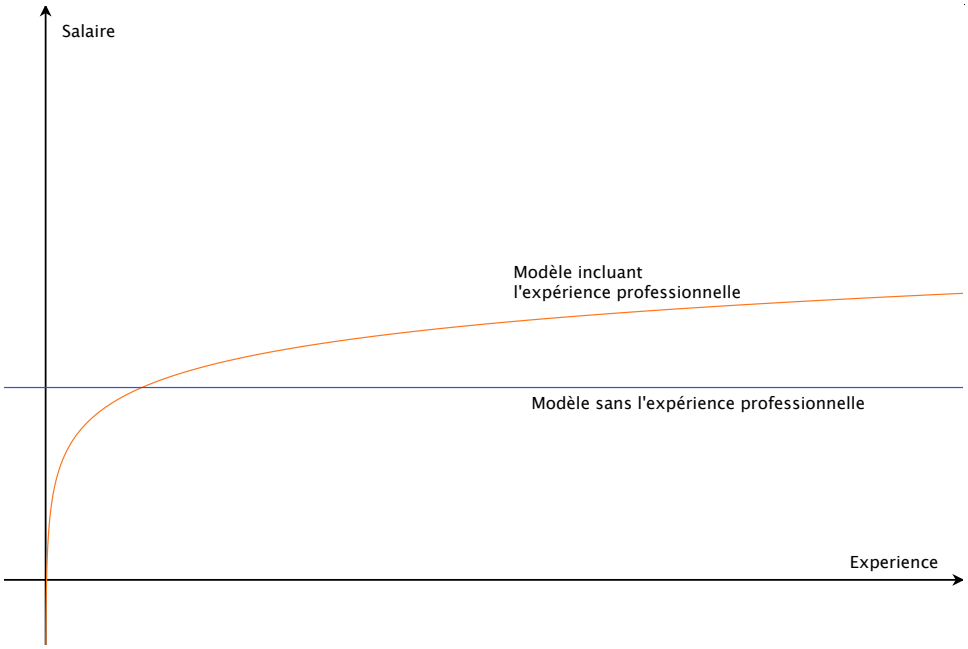


FIGURE 1 – Modèle d'éducation et profil âge-salaire prenant en compte l'acquisition du capital humain en entreprise

Mincer utilise l'équation (1) suivante pour lier le salaire réel observé avec le capital d'éducation, l'expérience professionnelle et l'expérience professionnelle au carré. L'ajout du terme quadratique permet de prendre en compte le fait que l'impact de ces variables sur le salaire est certes positif, mais leur effet marginal décroit dans le temps.

$$\log Sal = E_0 + a.E + c.exp + d.exp^2 + z.X + u \tag{1}$$

Le salaire est étudié au niveau logarithmique, E correspond au nombre d'années d'étude et exp au nombre d'années de travail, X correspond à la matrice des caractéristiques et u correspond au terme d'erreur.

Ainsi le coefficient a correspond au taux de rendement d'une année d'étude supplémentaire sur le salaire.

3.1.2 L'habilité

L'équation (1) précédente (3.1.1 page 11) constitue le fondement du travail des économistes de l'éducation. La méthode la plus courante pour calculer le rendement de l'éducation a est la régression des moindres carrés ordinaires (MCO). Cette technique est fondée sur l'hypothèse très forte que les variables explicatives ne sont pas corrélées avec le terme d'erreur u, ce qui est discutable.

Supposons que chaque individu possède une "habilité", un talent propre, qui n'est pas observable, et donc non mesurable *per se* et que ce talent a un effet sur le niveau de salaire de l'individu mais aussi sur son niveau d'éducation (nous pouvons facilement imaginer qu'un individu talentueux réussira, d'une part, mieux à l'école et aura, d'autre part, un meilleur salaire - ici, l'un n'étant pas lié à l'autre). Si une telle variable existe, l'estimation par MCO de a est alors biaisée puisqu'elle ne tient pas compte de cette habilité.

La première méthode pour corriger ce biais est d'inclure une variable proxis qui mesurerait l'habilité, par exemple un test de QI - Griliches et Mason [13], Griliches [12]. Cette méthode conclut à un biais vers le haut de a. Cependant, étant donné qu'aucune variable ne mesure l'habilité sans se référer à des normes éducatives (comme les tests de QI), l'usage de proxis influencé par l'éducation entraîne un biais vers le bas. Comment neutraliser alors l'influence propre de l'habilité ?

3.1.3 Les fratries et jumeaux

Dans leur article, Ashenfelter et Krueger [6] tentent de résoudre le problème de l'habilité par l'étude de jumeaux. En effet, cette approche part de l'hypothèse que les membres d'une fratrie - et a fortiori des jumeaux - sont quasi identiques puisqu'ils partagent le même contexte familial et social. En estimant le taux de rendement de l'éducation par différence entre les frères - ou les jumeaux - au niveau de l'éducation et du salaire, la variable "habilité", omise jusque là, est alors prise en compte et le biais qui en découle supprimé.

Le problème de cette approche réside dans le fait que si l'habilité est une composante individuelle ou familiale, ce qui ne la rend pas indépendante de l'éducation, la régression "within-family" ne permet pas de produire un estimateur moins biaisé que celui des

MCO. De plus, la mesure de l'éducation comprend encore plus d'erreurs entre deux frères qu'entre deux individus de la population. Ces erreurs impliquent que l'estimation du rendement de l'éducation qui en découle sera biaisée vers le bas. Pour résoudre ce problème, la méthode des variables instrumentales peut être adaptée.

3.1.4 Les variables instrumentales

L'introduction de déterminants à l'éducation non corrélés aux résidus du salaire permet d'obtenir une estimation du rendement de l'éducation consistante. Le principe des régressions avec variables instrumentales (VI), qui permet de faire une expérience naturelle et de comparer différents groupes entre eux, se déroule en deux étapes. D'abord, nous estimons l'effet des instruments sur la variable qui mesure l'éducation, puis nous estimons leur effet sur le salaire. La plus grande difficulté étant la sélection de bonnes variables instrumentales, qui doivent bien être indépendantes des résidus du salaire, sans quoi l'estimation du rendement de l'éducation sera biaisée : une corrélation positive implique une estimation biaisée vers le haut.

Dans ce domaine, le travail d'Angrist et Krueger [1] est précurseur. Ils utilisent le trimestre de naissance, a priori non corrélé aux résidus du revenu, comme instrument de l'éducation. C'est ainsi qu'ils arrivent à conclure que les enfants nés en début d'année étudient moins que ceux nés en fin d'année et vont ainsi avoir des salaires moins élevés.

De plus, de nombreux articles - comme celui d'Angrist et Krueger mais aussi Staiger et Stock [23] - tendent à établir que la régression des MCO est biaisée vers le bas et que l'introduction de variables instrumentales permettrait de corriger ce biais.

A partir du modèle du capital humain de Becker, en modélisant le lien entre salaire et éducation par l'équation de Mincer usuelle, Card [9] démontre que l'estimateur des MCO du rendement de l'éducation est biaisé par ce que la littérature appelle l' "habilité" des individus, développée dans la précédente section. Le fait de ne pas prendre en compte ce biais amènerait un estimateur biaisé vers le haut [6]. Auparavant, les économistes affirmaient que ce biais vers le haut dû à la variable omise "habilité" était compensé par un biais vers le bas dû à des erreurs de mesures sur la variable "éducation" (Griliches [12]). Angrist et Krueger [2] ont d'ailleurs démontré que la fiabilité du nombre d'années d'études déclaré est de 85-90% aux Etats-Unis et amène donc un biais vers le bas de 10 à 15%, suffisant pour compenser un biais d' "habilité" vers le haut selon eux. Mais Card démontre que ce biais de mesure peut ne pas suffire à compenser le biais d' "habilité" et conclut que l'estimation par VI est encore plus biaisée vers le haut que celle des MCO. Cette étude illustre le débat sur la méthodologie à employer pour modéliser le rendement de l'éducation entre la méthode des MCO et celle des VI.

6. *Ibid.,* p.1134

3.2 Linéarité du rendement de l'éducation

Le point de départ de la présente analyse est l'observation développée en introduction (1 page 6). En effet, l'une des principales hypothèses de Mincer [19] est la linéarité qui lie le salaire à l'éducation. Or nous avons vu que ce postulat ne semble pas être vérifié empiriquement. De même Denny et Harmon [10] apportent une sérieuse contribution à ce débat par de simples tests (coefficient pour chaque niveau d'éducation possible, introduction de l'éducation au carré et au cube, coefficient pour des niveaux d'éducation) en comparant cinq pays différents. En effet, ils démontrent la non-linéarité du taux de rendement de l'éducation pour la majorité des pays étudiés définissant ainsi l' "effet en peau de mouton".

Ainsi une large partie de la littérature s'est penchée sur la mise en place de nouveaux modèles permettant de faire varier le rendement de l'éducation en fonction de l'année supplémentaire, i.e. permettre au rendement de la $x^{ème}$ année d'éducation d'être différent de celui de la $x+1^{ème}$ année et ainsi se rapprocher au mieux de la réalité des faits.

3.2.1 Délinéariser le taux de rendement

Tout comme le démontrent Heckman et Polachek [16], il existerait un *effet diplôme* : une année d'éducation n'aboutissant pas à l'obtention d'un diplôme semble engendrer un rendement plus faible qu'une année sanctionnée par un diplôme. L'introduction du terme quadratique de l'éducation permettrait de savoir s'il existe bien une variation du taux de rendement en fonction de l'année supplémentaire concernée, comme c'est le cas dans la spécification d'Arestoff [3]. L'auteur observe en effet un "rendement au carré" positif et statistiquement différent de zéro, ce qui laisse penser que le rendement de l'éducation est une fonction positive du nombre d'années d'études.

Comment peut-on alors modéliser cette non linéarité apparente du taux de rendement de l'éducation ? Bien sûr, il est possible d'attribuer une variable dichotomique pour chaque année d'étude. Mais ce genre de modèles trouve rapidement sa limite puisqu'il entrainera des estimateurs très biaisés d'autant plus si l'échantillon d'analyse n'est pas très grand. Hungerford et Solon [17] ainsi que Schady [20] développent un modèle basé sur les années diplômantes : la modélisation prend la forme d'une fonction de Spline et rentre dans le cadre de la littérature sur le "sheepskin effects" aux Etats-Unis :

$$\log W_i = a + bX_i + cX_i^2 + dS_i + eD6_i + f\left[(S_i - 6) * D6_i\right] + gD10_i + h\left[(S - 10) * D10_i\right] + iD15_i + u_i \quad (2)$$

où nous retrouvons à gauche le calcul du revenu - du salaire - S et X sont respecti-

vement le nombre d'années d'études et le nombre d'années d'expériences et $D6, D10$ et $D15$ sont des variables dichotomiques pour les individus qui ont finalisé des cursus d'une durée minimale de six, dix et quinze années d'études correspondant à des années diplômantes en l'occurrence dans le système philippin - sujet de l'article de Schady. Le taux de rendement moyen des cinq premières années est donné par d. Celui pour la sixième année d'éducation est donné par la somme de d et e. La somme des coefficients d et f indique le taux de rendement moyen pour les trois premières années du secondaire. Nous pouvons ainsi calculer les taux de rendement moyens pour la suite du cursus éducatif.

Il n'est pas surprenant de constater que les rendements diffèrent en fonction du niveau de l'éducation et sont de plus en plus importants avec l'augmentation de ce dernier : 0.094 pour le *primary school* (équivalent au collège), 0.0100 pour le *secondary school* (lycée) et 0.167 pour l'université à bac+5.

3.2.2 Introduction du probit

L'un des derniers problèmes soulevés par ces modèles est celui de l'endogénéité des variables. Bien que cette question soit présente dans les travaux précédents, ces spécifications ont finalement du mal à résoudre ce problème. Tout comme nous l'avons montré, bien que le sens de l'erreur diffère en fonction des études, la méthode des MCO entraîne un biais certain.

J. Garen [11] démontre que l'utilisation de la variable "éducation" en terme discret implique des biais tels que l'utilisation d'un cadre d'équations simultanées serait préférable à l'approche usuelle. L'utilisation de ce type de modèles permet au rendement de l'éducation de l'année x d'être différent de celui de l'année $x + 1$: le taux de rendement de l'éducation n'est pas linéaire. A. Skalli [21] démontre qu'en termes de cadre théorique, cette non-linéarité s'inscrit directement dans la théorie du signal puisque l'effet "peau de mouton" est souvent interprété comme la récompense pour un individu de "signaler" son habilité (dans le sens définit précédemment) par rapport à ceux n'ayant pas réussi à atteindre ce même niveau d'éducation ou en un temps plus long (par des doublements de classes, reprise d'études après un temps sur le marché du travail). Le choix du cadre théorique dépasse donc le choix idéologique. Les théories du capital humain et du signal n'expliquent pas à elles seules l'intégralité des évènements concernant le processus scolaire. Alors que la théorie du capital humain n'arrive pas à intégrer la force des diplômes en France, celle du signal ignore les connaissances apportées par le système éducatif, directement applicables sur le marché du travail et qui permettent donc d'augmenter la productivité de l'individu. Il semble donc que nous devrons prendre en compte ces deux cadres au moment de l'analyse de nos résultats.

Harmon et Walker [15] sont les premiers à introduire un modèle probit ordonné dans

lequel la variable "éducation" est considérée comme un nombre entier ordonné. Cependant il conserve une fonction de gain linéaire à l'éducation. Vella et Gregory [25] développent, sur les bases de Willis et Rosen [26], un modèle dans lequel on calcule le salaire d'un individu - avec des caractéristiques et un niveau d'éducation connus - s'il avait choisi un niveau d'éducation différent. En le comparant avec le salaire actuel, nous pouvons ainsi obtenir le taux de rendement de l'éducation.

Notre travail se fonde sur la modélisation de A. Skalli [21]. L'auteur s'appuie sur la modélisation d'Heckman en deux étapes mais introduit un modèle probit ordonné. On obtient ainsi :

$$y_i^j = X_i'\beta^j + u_i^j,\ j = 1, 2, ..., J \quad (3)$$

$$S_i^* = Z_i'\gamma + v_i \quad (4)$$

$$S_i = j \text{ if } \mu^{j-1} < S_i^* \leq \mu^j \quad (5)$$

où les individus sont représentés par i, $S_i = j$ est le nombre d'années d'études, y_i^j est le log-salaire, X et Z sont des vecteurs de variables caractéristiques, S_i^* est la variable latente correspondante à S_i, les termes d'erreurs sont u_i^j et v_i et supposés normaux. Enfin, μ est un paramètre inconnu qui doit être estimé et correspond aux seuils nécessaires pour prendre la décision de s'éduquer une année supplémentaire.

4 Données et variables

Tous les résultats que nous serons amenés à exploiter proviennent de l'enquête Formation et Qualification Professionnelle de 2003 effectuée par l'INSEE. Cette base de données donne des informations précises sur la formation initiale d'une part et sur le salaire et la profession des individus d'autre part. De plus, et c'est sur ce point que réside son originalité, elle fournit des données sur les positions occupées à différents moments de la carrière, la formation postscolaire ainsi que sur la formation et la position sociale des parents[7] . Après avoir nettoyé la base, nous obtenons un échantillon de 23 665 individus interrogés sur leur activité de l'année 2003.

4.1 Variables clés

L'enquête fournit le salaire en 2002 en clair redressé ainsi que le nombre de mois travaillés. Afin de prendre en compte un nombre plus important d'individus, nous avons

7. Source : http ://www.insee.fr

construit la variable *lsalredm* comme le revenu mensuel des individus, sous sa forme logarithmique pour suivre la littérature - Heckman&Polachek [16].

Concernant le nombre d'années d'études *educan*, nous avons encore utilisé le même procédé que dans la littérature sur le sujet. C'est-à-dire que nous retranchons à l'année de fin d'études initiales, l'année de naissance moins six - six ans correspondant à l'âge minimal obligatoire en France pour l'entrée en scolarité. Ainsi, nous obtenons un nombre d'années d'études. Pour des raisons de cohérence, nous n'avons conservé que les nombres d'années d'études positifs et n'excédant pas 20 ans, ce qui correspond au plus grand nombre d'années permettant d'obtenir un doctorat en France. Cependant cet intervalle ne prend pas en compte un problème français essentiel : le doublement. Ne disposant pas de l'information, nous ne pouvons pas savoir si l'individu a doublé une de ses classes dans son parcours scolaire, ce qui apporte inévitablement un biais à cette variable.

Nous utiliserons souvent la variable *ddipl* correspondant au classement de l'INSEE suivant :

ddipl	Niveau INSEE
1	Diplôme supérieur
3	Baccalauréat + 2 ans
4	Baccalauréat ou brevet professionnel ou autre diplôme de ce niveau
5	CAP, BEP ou autre diplôme de ce niveau
6	BEPC seul
7	Aucun diplôme ou CEP

TABLE 1 – Classement du niveau d'éducation, INSEE

Cette variable se révèlera intéressante au moment du modèle multinomial. Elle donne une catégorisation claire en terme de diplôme, entre les individus. Nous ne raisonnerons plus alors en années d'études mais en niveau de diplôme.

A propos de l'expérience professionnelle, nous ne connaissons pas le temps exact de travail effectué sur l'ensemble de la carrière professionnelle. Mincer introduit l'expérience professionnelle comme l'ensemble de la période qui suit la fin de la scolarité. Nous pourrions être séduits par cette idée. Cependant, ce serait créer une variable qui ne tient pas compte de tous les arrêts qui peuvent exister durant la carrière professionnelle - chômage ou congé parental. Nous préférons donc utiliser en variable proxy l'âge de l'individu, qui comporte les mêmes biais mais qui a l'honnêteté de ne pas prétendre être exacte.

4.2 Variables caractéristiques et instrumentales

Afin d'améliorer nos spécifications, nous introduisons les variables caractéristiques classiques que sont le statut matrimonial, la nationalité et le sexe [8] ainsi qu'un ensemble de variables binaires par régions françaises. Avec les données dont nous disposons, nous avons jugé pertinent d'introduire deux autres variables. D'abord, le fait de vivre ou non dans une Zone Urbaine Sensible [9] (ZUS), qui sembe avoir indéniablement un impact sur le salaire. Ensuite, nous avons créé la variable *demenage* qui est binaire : si l'individu vit dans la même région que sa région de naissance, *demenage* prend la valeur 0, 1 si non. Bien que nous ne sachions pas quand a eu lieu le déménagement de l'individu, nous pourrons savoir s'il existe, non pas une causalité, mais au moins une corrélation entre le déplacement géographique et le salaire et l'éducation.

Concernant les variables instrumentales de l'éducation, nous avons tenus à reprendre celles d'Angrist et Krueger [1] : le trimestre de naissance. A ces variables nous ajoutons le *ddipl* de la mère et du père (respectivement *ddiplm* et *ddiplp*), ainsi que deux variables binaires : $frat$ qui prend la valeur 1 si l'individu a une fratrie et *div* qui est égale à 1 si les parents de l'individu se sont séparés ou ont divorcés au moment de l'éducation de ce dernier.

5 Modélisation

La présente section présente les modèles que nous allons utiliser pour effectuer notre analyse à partir des modèles présentés en section 3 (3 page 10).

Notre étude va tenter de comparer ces différents modèles entre eux, afin de tirer une conclusion sur la modélisation la plus adaptée à la réalité et qui prendrait en compte les divers biais existants pour chacun des modèles.

Nous étudierons d'abord l'équation de Mincer définit par la formule 6 suivante, sous les hypothèses de bases de la régression MCO :

$$lsalredm = lsalredm_0 + a.educan + b.age + z.X + u \tag{6}$$

Nous effectuerons de même avec *ddipl* à la place d'*educan*. Nous définissons le vecteur des variables caractéristiques suivant : $X' = (matri1, matri2, matri3, matri4, natio2, natio3, natio4, natio5, sexe, zus, demenage)$.

8. Pour plus de détails sur les modalités, voir annexe.

9. Les ZUS "sont des territoires infra-urbains définis par les pouvoirs publics pour être la cible prioritaire de la politique de la ville, en fonction des considérations locales liées aux difficultés que connaissent les habitants de ces territoires" - définition INSEE.

Notre second modèle consiste à effectuer une régression avec VI par la méthode des doubles moindres carrés ordinaires (*2SLS*) :

$$educan = educan_0 + \alpha.age + \beta.age2 + \delta.V + \eta.X + u \quad (7)$$

$$lsalredm = lsalredm_0 + \lambda.educan + \mu.age + \pi.age2 + \tau.X + v \quad (8)$$

Ensuite, nous reproduirons le modèle de Schady [20] en utilisant les variables binaires suivantes $D12, D15, D17$ prenant la valeur 1 si l'individu a effectué au moins 12, 15 et 17 années d'études correspondant aux étapes du cycle scolaire français : baccalauréat, licence et master. Nous introduisons également les variables croisées entre *educan* et ces binaires :

$$\begin{aligned} lsalredm = \quad & lsalredm_0 + a_1.educan + a_2.D12 + a_3.D15 + a_4.D17 + \\ & + a_5.educan * D12 + a_6.educan * D15 \\ & + b_1.age + b_2.age2 + b.X + w \end{aligned}$$

Enfin, nous utiliserons un probit multinomial comme première étape avec *ddipl* avant de l'introduire dans la régression du salaire et d'appliquer un MCO dans le but de vérifier si la spécification multinomiale explique mieux la réalité que le simple *2SLS*. Le modèle prend la forme suivante :

$$ddipl = \alpha_0.age + \alpha_1.age2 + \alpha_3.X + \alpha_4.V + u_1 \quad (9)$$

$$lsalredm = lsalredm_0 + \beta_1.ddipl + \beta_2.age + \beta_3.age2 + \beta_4.X + u_2 \quad (10)$$

6 Résultats

6.1 Description de l'échantillon

Notre échantillon se compose de près de 24 000 individus. Avant d'étudier les différentes régressions, nous tenions à développer le contenu de nos données.

La table 3 (8.3 page 36) présente les statistiques descriptives. En moyenne, les individus gagnent 1737€ par mois en ayant effectué près de treize années d'études et sont âgés de 40 ans. Cependant nous pouvons observer que la dispersion des salaires est très importante (avec un écart-type de 2791) - c'est pourquoi nous effectuons la transformation logarithmique - ainsi que celle de l'âge (dont l'écart-type est de 10).

Concernant le niveau de diplôme, notons que près de 25% des individus n'en ont aucun, qu'un autre quart ont un niveau CAP ou BEP, que seulement 16% ont le baccalauréat. Ainsi les 23% restant possèdent un niveau de formation supérieur au bac.

Au niveau des variables caractéristiques, les hommes sont légèrement majoritaires puisqu'ils représentent un peu plus de 51% de l'échantillon, une très faible proportion des individus vit en ZUS (5,9%), presqu'un tiers ne vit pas dans sa région de naissance, nous avons plus de 90% des individus qui sont Français de naissance et 57% de l'échantillon sont mariés contre un tiers de célibataires alors que 1,5% est veuf. Il est intéressant de noter que 8,4% des individus sont divorcés ce qui est légèrement plus faible que le nombre d'individus ayant effectué sa scolarité avec des parents séparés ou divorcés (8,81%). En outre, plus de 90% des individus ont au moins un frère ou une soeur. Pour conclure, concernant le mois de naissance, la population est bien répartie avec 25,32% pour le premier trimestre, 25,80% pour le deuxième, 24,80% et 24,08% pour le troisième et le quatrième respectivement.

6.2 Analyses des modèles

6.2.1 MCO

Avant de répondre à la question de la linéarité du taux de rendement de l'éducation, nous tenions à vérifier les résultats standards sur le sujet par la modélisation de Mincer. Le tableau 4 (8.4.1 page 37) présente ce modèle avec différentes spécifications. La première colonne présente les résultats avec seulement *educan*. Le rendement s'élève alors à 5,78% de salaire par année d'études supplémentaire. Les colonnes 2 et 3 introduisent les variables *age*, *age*2 et *educan*2. Le rendement diminue de 8,01% à 7,58% pour la spécification avec les variables au carré. En outre nous observons que l'âge a un effet important sur le salaire puisque que les individus perçoivent 5,76% de plus en salaire par année supplémentaire. Notons cependant que les variables au carré ont des coefficients négatifs : bien que peu élevés (-0,01% pour l'éducation et -0,05% pour l'âge), ce résultat montre que ces variables ont un rendement positif dans le temps, mais que ce dernier décroît. Cela prouve, par ailleurs que les rendements de l'éducation et de l'âge ne sont pas constants - dans le cas contraire, le coefficient des variables au carré serait nul. Les colonnes 4 et 5 introduisent les variables caractéristiques ainsi que les binaires régionales (colonne 5). Si nous nous concentrons sur la dernière colonne, le rendement de l'éducation est alors plus élevé que ce nous pouvions penser jusqu'à présent puisque qu'il atteint 8,22% par année d'éducation supplémentaire. L'impact de l'âge diminue un peu puisque qu'il a un effet de 5,21% et les variables au carré restent faibles mais toujours négatives - notons que *educan*2 n'est plus significative.

En nous penchant sur les variables caractéristiques, nous obtenons un résultat classique concernant le sexe. En effet, dans notre échantillon, le fait d'être un homme apporte un gain de près de 40% sur le salaire. Cette observation reflète la discrimination salariale qu'il existe en France entre les sexes. Mais cet écart peut aussi s'expliquer par le phénomène de "plafond de verre" : à niveau de formation égal, les femmes ont moins souvent accès à des postes supérieurs, qui sont généralement mieux rémunérés.

Par ailleurs, le fait de vivre dans une ZUS aurait un impact négatif sur le salaire : les individus vivant dans ces quartiers subissent une perte de salaire de près de 10%. Ceci pourrait s'expliquer par le fait que la population de ces zones urbaines subit aussi une discrimination du fait de leur localisation dans ces secteurs. Cet impact conséquent sur le salaire ne peut être ignoré, même s'il reste complexe à analyser. De la même manière, nous constatons que le fait d'avoir déménagé à un moment donné de sa vie semble apporter un gain de salaire de 5,61%. Cela pourrait s'expliquer par une plus grande adaptation de la population au marché du travail car les perspectives d'emploi sont différentes selon les territoires. Concernant le statut matrimonial et la nationalité des individus, nous avons choisi les modalités les plus représentées dans l'échantillon comme référent : être marié d'une part et être français par la naissance de l'autre. Le célibat aurait un impact négatif sur le salaire puisqu'il semble induire une diminution de salaire de 5,70%. De même, nous constatons que le veuvage a un effet dépréciatif sur le salaire de plus de 10%. A contrario, les divorcés perçoivent un salaire augmenté de 2,53%. Mais il semble difficile d'expliquer ces résultats sans informations supplémentaires sur cette variable. Par ailleurs nous observons que quelque soit la provenance de l'individu dans le monde, son salaire chute de 7 à 37 % comparé à un Français de souche - les individus ayant acquis la nationalité gagnent 12% de moins que les Français de naissance, ce qui est plus qu'un individu provenant du reste de l'Europe. Cependant, la discrimination la plus importante concerne les individus d'origine africaine qui subissent plus d'un tiers de perte de salaire.

Nous avons effectué les mêmes régressions que précédemment en remplaçant la variable *educan* par *ddipl* (8.4.2 page 38). Rappelons que cette dernière correspond au niveau de diplôme le plus élevé obtenu par l'indidu. Nous observons que plus le niveau du diplôme est élevé, plus le salaire est important - le coefficient est négatif mais rappelons que le niveau 1 correspond aux diplômes supérieurs alors que le niveau 7 correspond aux individus sans diplôme ou avec un CEP. En comparant à la colonne 5 du modèle avec *educan* nous constatons que les variables ont le même effet sur le salaire en termes de signe avec toutefois, de légères différences en termes de coefficient, bien que dans cette spécification, l'ensemble des variables sont significatives au seuil de 1%, ce qui n'est pas le cas avec *educan*.

6.2.2 Spécification de Schady

Ensuite nous avons voulu reproduire le modèle développé par Schady (5 page 19) pour étudier plus précisément la linéarité du rendement de l'éducation. Ici encore, nous ne constatons pas de changement pour les variables caractéristiques. Cependant, concernant le rendement de l'éducation, nous obtenons des résultats étayant la non linéarité de ce dernier. En effet, nous pouvons le calculer pour les différentes étapes du parcours éducatif - la table 6 présente les résultats (8.4.3 page 39) : le taux de rendement moyen pour les onze premières années d'éducation est de 6,32% alors que le rendement de la douzième année s'élève à 12,36% - dans la colonne 2, prenant en compte le biais régional. Rappelons que le nombre d'années moyen pour l'échantillon est légèrement supérieur à 12. Cette spécification, plus précise, donne donc un rendement encore plus élevé que pour le modèle avec *educan*. Cette année est particulièrement importante puisqu'elle correspond à l'année de terminale et donc à l'obtention du bac - en faisant l'hypothèse qu'il n'y a pas de doublement. De plus, le rendement moyen des trois premières années d'université est de 8,33%, ce qui est bien moindre que celui du baccalauréat. Le taux de rendement moyen de la quinzième année d'éducation - correspondant à un niveau licence - atteint 13,45%, rendement supérieur à celui du baccalauréat. Celui des seize premières années s'élève à 2,02%. En effet, en moyenne, le rendement est négatif pour les seize premières années. Enfin, le gain moyen pour la dix-septième année - correspondant à un niveau master - est de 17,38%. Ces résultats sont très intéressants puisqu'ils indiquent que dans le système français, il est préférable d'achever un cycle d'éducation - baccalauréat, licence, master - plutôt que d'en entamer un sans l'achever. La théorie du signal prend alors pleinement sens dans ce modèle : une année non diplômante n'apporte pas de gain. En effet, si nous comparons les rendements du baccalauréat, de la licence et du master - respectivement 12,36%, 13,45% et 17,38% - nous observons d'abord qu'ils diffèrent, même en les divisant par le nombre d'années : respectivement 1,0300%, 0,8967% et 1,0224%. Nous constatons alors que le rendement annuel décroit entre la licence et le baccalauréat ; puis croît pour le master, qui reste néanmoins inférieur à celui du baccalauréat. Nous retrouvons ici le résultat donné par la variable *educan*2 qui montrait déjà que le rendement est d'abord non linéaire puis décroît avec les années d'éducation. Ce modèle apporte donc une preuve sérieuse de la non-linéarité du rendement de l'éducation dans le système français.

6.2.3 Variables Instrumentales

La méthode des VI corrige le biais de l'estimateur des MCO comme nous l'avons vu durant la deuxième partie (3 page 10). Nous avons ensuite appliqué la méthode des doubles moindres carrés. Ainsi, nous étudions les déterminants de l'éducation pour

améliorer son implication dans l'équation du salaire. Le tableau 7 (8.4.4 page 40) présente les résultats des estimations. La première colonne présente les résultats avec seulement $trimnais, ddiplm$ et $ddiplp$ comme variables instrumentales. Nous introduisons ensuite $frat$ et div, puis les binaires de régions. Nous constatons d'abord que l'introduction de VI supplémentaires réduit l'impact de l'âge sur l'éducation, même si cet effet est toujours positif. Par ailleurs, $age2$ ne subit pas de modification quelque soit la spécification de $educan$. Concernant le statut matrimonial, les célibataires effectuent entre 0,13 et 0,15 année d'éducation en plus par rapport aux indivus mariés, les divorcés suivent près de 0,30 année de moins, et les veufs environ 0,70 année de moins. En partant de l'hypothèse que les veufs sont plus âgés que les divorcés, que les indidus mariés et les célibataires sont les plus jeunes et en notant que le nombre d'années d'éducation a augmenté en moyenne durant le 20ème siècle, ces résultats s'expliquent donc simplement par un effet de génération [10].

Les résultats concernant la nationalité sont encore plus frappants que ceux relatifs au salaire. Alors que les Français d'adoption poursuivent leurs études 0,66 année de moins que les Français de naissance, les étrangers issus du reste de l'Europe perdent plus de deux années d'études, et ceux provenant d'Afrique 1,3 année ; ceux du reste du monde : 1,8. Ces résultats, bien qu'importants, peuvent s'expliquer par plusieurs facteurs. L'explication simple serait de considérer que les étrangers s'éduquent moins que les Français d'origine. Mais cette interprétation ne prend pas en compte les différents systèmes scolaires du monde entier. En effet, pour atteindre le même niveau d'éducation, certains pays proposent des cursus plus ou moins longs que ceux proposés en France. Nous pouvons donc nous interroger sur l'impact de ce système, en terme de qualité d'éducation, mais le modèle ne peut pas la mesurer. Il est donc possible que, pour une durée de formation différente, deux pays décernent le même diplôme. Cela peut expliquer les écarts que nous constatons pour les différentes nationalités.

Un autre résultat intéressant concerne la relation éducation-sexe. Alors que les hommes perçoivent des salaires supérieurs comme nous l'avons vu précédemment (6.2.1 page 21), ils s'éduquent moins que les femmes - plus d'une demi-année de différence. Et les femmes, donc, se forment plus en moyenne que les hommes, pour des salaires inférieurs. Ce résultat confirme la discrimination salariale selon le sexe.

Nous constatons les mêmes effets pour zus et $demenage$ sur l'éducation que sur le salaire : le fait d'habiter dans une ZUS diminue le nombre d'années d'études de plus d'une demi-année alors que le fait d'avoir déménagé l'augmente. Nous pouvons penser que les résultats sur le salaire sont des conséquences de l'éducation.

10. "En un siècle, la durée de scolarisation moyenne s'est accrue de plus onze ans, passant de moins de sept ans à plus de dix-sept ans." Analyse sur les enjeux de l'éducation et plus précisément de l'école pour tous, dans les années à venir Claude Lelièvre.

Nous nous intéressons désormais aux variables instrumentales. Tout d'abord nous observons que plus le trimestre de naissance est tardif dans l'année, plus l'impact sur l'éducation est important et positif. Nous retrouvons ici les résultats d'Angrist et Krueger [1] à un détail près : alors que le deuxième trimestre perd en significativité avec l'introduction de $frat$ et div pour ne plus être significatif en présence des binaires régionales, le troisième trimestre n'est pas significatif et ce quel que soit la spécification. Seul le quatrième trimestre reste significatif au seuil de 1% pour tous les modèles. Le résultat d'Angrist et Krueger [1] est donc moins important que dans notre échantillon et est à questionner, comme l'ont fait d'autres auteurs - bien que nous observons la même tendance qu'eux. De plus, le fait d'avoir des frères et soeurs diminue de plus d'une demi-année l'éducation et de 0,40 an le fait d'avoir des parents séparés. Nous pouvons donc en conclure que le contexte familial dans lequel évolue l'individu a une influence importance sur son cursus scolaire. Au delà de l'impact psychologique que peut avoir le divorce des parents sur les jeunes enfants, cela peut avoir des conséquences sur les moyens financiers qu'a le parent élevant l'enfant et donc entrainer des études plus courtes qu'avec la présence des deux parents dans le ménage. En outre, si les parents ont à charge plusieurs enfants, ils n'ont pas nécessairement les ressources nécessaires à de longues études pour l'ensemble de leurs enfants. Enfin, les individus avec des parents ayant un faible niveau de diplôme - correspondant aux variables *ddiplm* et *ddiplp* - vont eux-mêmes avoir un plus faible niveau. Ce résultat appuie la fameuse *reproduction sociale* que présentent de nombreux sociologues (P. Bourdieu et J.-C. Passeron [18]).

Quel est l'impact de cette méthodologie sur la spécification du salaire ? Nous constatons d'abord que quel que soit la spécification, le rendement de l'éducation VI est beaucoup plus important que celui des MCO. En effet, alors que les MCO attribuent un rendement de 8,22% de salaire supplémentaire, la même spécification par la méthode des VI observe un rendement de 11,05% - notons que l'introduction de $frat$ et div augmente le rendement et que ce dernier diminue après la présence des binaires régionales. Ces résultats coïncident avec la littérature affirmant que la méthode des MCO entraine un biais vers le bas du rendement de l'éducation - comme nous l'avons vu au cours de la section 2 (3.1.4 page 14). Le rendement est donc plus important que nous pouvions le penser initialement. Au sujet des variables explicatives, alors que cette spécification n'influe pas l'impact de l'âge ni du sexe sur le salaire et que le signe des autres variables ne change pas, nous constatons que leur impact est grandement modifié par la méthode des VI. En effet, le résultat relatif à la situation de célibat ou de mariage est plus important qu'initialement observé - à raison de près du double pour les divorcés, l'effet d'être veuf est moins important. Par ailleurs, le fait d'être étranger provenant du reste de l'Europe perd sa significativité - ce qui laisse penser qu'il n'y a donc pas de différence sur le marché

du travail entre un Français et un Européen. De même, le fait d'acquérir la nationalité française ou de provenir d'Afrique ou d'une autre région du monde a moins d'impact que prévu par la méthode des MCO. Enfin, les rendements des variables *zus* et *demenage* sont aussi biaisés mais vers le haut.

Pour conclure sur ces résultats, nous tenions à préciser que nous avions effectué les tests appropriés confirmant pour chaque spécification présentée que *educan* est endogène - justifiant l'utilisation des VI - et que les instruments sont pertinents. Notons enfin que seule la spécification avec *div*, $frat$ et sans les binaires régionales n'est pas suridentifiée. Nous pouvons en conclure que la présence des variables $trimenais_i$ entraîne la suridentification des autres modèles.

6.2.4 Probit

Les variables instrumentales atténuent donc l'impact des variables caractéristiques mais elles augmentent grandement le rendement de l'éducation. Cette méthode offre une analyse plus précise de l'impact de l'éducation sur le salaire - bien que ces résultats ne fassent pas consensus dans la littérature. Cependant, nous savons que cette spécification ne résout pas l'ensemble des problèmes de l'endogénéité dû à la variable *educan*. C'est pourquoi nous introduisons un modèle en deux étapes dans lequel l'éducation n'est pas estimée par un MCO mais par un modèle probit ordonné. Cependant, nous ne pouvons pas utiliser *educan* sans introduire un problème d'identification puisque cela entraîne l'analyse d'une variable ordonnée à vingt-deux modalités. Nous utilisons donc *ddipl* - rappelons que cette variable à six modalités est contrainte de telle sorte que le niveau 1 correspond au plus haut niveau de diplômes et que le niveau 7 correspond au niveau sans diplôme. La table 8 (8.4.5 page 41) présente les résultats des régressions.

Concentrons-nous d'abord sur la spécification de *ddipl*. Nous observons immédiatement que *age* a certes un impact positif sur le niveau de diplôme mais est significatif à 10% sans la présence des binaires régionales et plus du tout significatif avec l'introduction de ces binaires, alors qu'*age*2 conserve le même effet que dans les modélisations avec *educan*.

Concernant le statut matrimonial, alors que le fait d'être célibataire a un effet non significatif par rapport au fait d'être marié, les veufs et les divorcés ont un niveau de diplôme moins élevé que les individus mariés. De même qu'avec *educan*, nous observons que les étrangers et ceux qui ont acquis la nationalité française ont un niveau de diplôme moins élevé que les Français de naissance. Nous retrouvons ici encore le même résultat qu'avec *educan*. Nous obtenons les mêmes résultats que la spécification d'*educan* dans la méthode des VI pour *sexe*, *zus* et *demenage* : les femmes ont un niveau de diplôme plus élevé que les hommes ; les individus vivant en ZUS sont moins éduqués et le fait d'avoir

déménagé induit un haut niveau de diplôme.

Dans cette spécification, le trimestre de naissance n'est plus du tout significatif et ce, quelque soit le trimestre. Ici encore, cette variable en tant qu'instrument est donc remise en question. Pour les variables $ddiplp, ddiplm, frat$ et div nous retrouvons à nouveau les résultats des VI. Ceci confirme l'influence du contexte familial sur le niveau d'éducation des individus.

Considérons, à présent, la deuxième étape de la modélisation : nous avons reconstruit une variable *ddiplpro* à partir des seuils du probit, puis effectué un MCO du salaire avec cette nouvelle variable. Nous observons que cela diminue le coefficient associé à *ddipl* par rapport aux MCO : il passe de -0,1372 à -0,1266 sans les binaires régionales et de -0,1314 à -0,1128 avec. Les MCO introduisent donc un biais vers le haut sur cette variable - en valeur absolue. Le reste des variables caractéristiques conserve à peu près le même impact sur le salaire avec cette spécification, comparativement à la méthode des MCO. L'introduction du probit permet donc de résoudre le problème d'endogénéité lié à l'éducation mais cela ne modifie pas de manière significative les résultats des autres variables.

La difficulté de cette spécification réside dans le fait de modéliser l'éducation. Bien que *ddipl* nous donne un indicateur, nous ne pouvons comparer les rendements de l'éducation entre eux. Etant donné qu'il est impossible de poursuivre notre analyse à ce sujet, la dernière partie va exploiter les résultats supplémentaires donnés par le probit : comment les variables caractéristiques et instrumentales influencent la probabilité d'atteindre les niveaux définis par *ddipl* ?

6.2.5 Les déterminants de l'éducation

Le tableau 9 (8.4.6 page 42) présente les effets marginaux des variables sur les différents niveaux de *ddipl*. Nous constatons que l'effet de l'âge n'est pas très important ; toutefois, les signes pour chaque niveau indiquent que les plus âgés atteignent les niveaux de diplôme les plus élevés. La question est de savoir s'il s'agit d'un effet de génération par lequel les précédentes générations s'éduquent plus que les nouvelles - ce qui irait à l'encontre des résultats sur l'évolution de la scolarisation - ou si les individus les plus jeunes n'ont pas fini leurs études ce qui expliquerait alors qu'ils aient automatiquement un niveau moins élevé que leurs aînés.

Le statut matrimonial a un faible impact surtout pour les célibataires. Les statuts "veuf" et "divorcé" impactent plus le niveau puisqu'ils ont un effet entre 1% et 8% - le fait d'être veuf a un effet marginal de 7,98% sur le niveau 7, i.e. le plus bas alors qu'il a un effet de -4,22% sur le niveau 1. Selon l'hypothèse du cycle de vie - célibataire, marié, divorcé, veuf - ces résultats vont à l'encontre de l'effet de l'âge. Peut être que nous pouvons attribuer ces impacts au hasard seul.

Concernant la nationalité, les résultats confirment ceux des MCO : quelque soit la nationalité, comparativement au fait d'être Français d'origine, le fait d'être étranger a un rendement marginal positif uniquement pour les niveau 5,6 et 7 correspondant à un niveau inférieur au baccalauréat. Mais ici encore, l'individu étranger peut avoir évolué dans un système éducatif différent de celui en France, ce qui pourrait expliquer ces résultats. Ou alors, les étrangers venant en France sont moins éduqués et ont donc objectivement des niveaux d'éducation moins élevés que les Français, indépendamment des différences de systèmes éducatifs.

Les rendements de $sexe, zus$ et $demenage$ confirment nos précédentes intuitions : les hommes ont des rendements négatifs entre 1 et 2% sur les plus hauts niveaux d'éducation, alors que nous constatons l'inverse pour les niveaux les plus faibles - entre 0,5 et 5% pour le niveau 7. De même, le fait d'être en ZUS a un rendement négatif entre 3 et 4,5% pour les niveaux 1,2 et 3, alors que $demenage$ a un rendement positif entre 1,6 et 2,4% pour ces mêmes niveaux. Notons que ces variables ont un impact plus important sur les niveaux 1 et 7.

Les trimestres de naissance impactent de moins de 1% dans chaque cas, ce qui conforte le manque de pertinence de cette variable. En revanche, $ddiplm, ddiplp, frat$ et div influent entre 1 et 7% selon les cas. Ici encore, les résultats nous confortent dans nos précédentes observations. Le contexte familial a une forte influence sur l'individu puisque le rendement des parents avec un haut niveau de $ddipl$: un rendement de 4,5% pour le père et 5,3% pour la mère, si l'individu est au niveau 7. Cela conforte l'idée de reproduction sociale. De même, la fratrie et le divorce tendent à pousser les individus vers les niveaux 6 et 7 : le fait de s'inscrire dans une fratrie a un rendement de 6,7% au niveau 7 contre -3,6% pour le niveau 1 ; le fait d'avoir des parents divorcés génèrerait un gain de 7,2% au niveau 7 contre -3,8% au niveau 1.

Les rendements apportent une nouvelle preuve aux intuitions que nous avons eues pour expliquer l'impact des variables sur l'éducation. Cependant, il nous semble intéressant de comparer la probabilité qu'ont les individus d'atteindre les différents niveaux. Etant donné le grand nombre de variables et de modalités, nous ne pouvons présenter l'ensemble des probabilités. Nous ne présenterons que les cas les plus probants.

Nous avons étudié le profil des Français de naissance, mariés, ayant déménagé, sans fratrie, ne vivant pas en ZUS et dont les parents n'ont pas divorcés (tableau 10, p.41). L'impact du sexe est indéniable sur la probabilité d'accéder aux différents niveaux. En effet, les femmes ont plus de 22% de chance d'accéder au niveau 1, le plus élevé alors que les hommes avec les mêmes caractéristiques n'ont que 17,70% de chance d'y accéder. L'écart pour les niveaux 3 et 4 diminue, et la tendance s'inverse pour les niveaux les plus bas : les hommes ont 16,31% de chance d'être de niveau 7 contre seulement 11,06% pour

les femmes. Ce constat confirme la meilleure réussite des femmes à accéder à un haut niveau d'éducation.

Ensuite, nous observons l'incidence d'avoir déménagé sur les hommes français d'origine, mariés, avec des frères et soeurs, ne vivant pas en ZUS et dont les parents n'ont pas divorcé au moment de l'éducation (tableau 11, p.41). Ceux qui ont déménagé ont 5% de plus d'accéder au niveau 1 alors qu'ils ont 8% de chance en moins de ne pas être diplômés (niveau 7). La difficulté de cette variable demeure dans le fait que nous ne savons pas à quelle époque de la vie de l'individu a eu lieu le déménagement. Nous pouvons toutefois supposer que les étudiants sont amenés à se déplacer pour accéder aux meilleurs universités et donc aux plus hauts niveaux d'éducation.

Nous étudions la même population en introduisant l'impact de la fratrie parmi les individus qui ont déménagé (tableau 12, p.41). Les probabilités nous prouvent encore que le fait d'avoir une fratrie diminue grandement les chances d'atteindre un haut niveau d'éducation. En effet, un individu sans frère aura 4% de plus d'être de niveau 1 qu'un individu avec frère ; la tendance s'inverse pour le niveau 7, pour lequel l'écart est de 5%. Notons, simplement, qu'il n'y a pas de différence significative pour la catégorie 5, représentant la probabilité la plus importante pour tout notre échantillon.

Enfin nous examinons l'impact d'être Français d'origine sur les hommes mariés ne vivant pas en ZUS, ayant déménagé, ayant des frères et soeurs et dont les parents n'ont pas divorcé (tableau 13, p.41). Ici encore, les étrangers ont beaucoup moins de chance d'atteindre un niveau élevé que les Français de naissance. En effet, alors que les derniers ont 13,81% de chance d'arriver au niveau 1, les étrangers en ont seulement 7,60%. Et inversement, les Français ont 21,14% de chance d'être au niveau 7 contre 29,25% pour les non-Français. Nous pouvons penser que ce sont les moins éduqués qui arrivent en France d'où ce pourcentage de près de 30% pour la catégorie 7.

En regardant l'ensemble de ces résultats, nous observons que les plus grandes probabilités correspondent aux niveaux 5 et 7. La population que nous avons ciblée semble donc, dans son ensemble, peu éduquée puisque seul un faible pourcentage atteint un niveau égal ou supérieur au baccalauréat. Cette observation peut s'expliquer par l'effet de génération puisque la majorité de nos individus sont nés dans les années 60 et que le "marché des diplômes" a grandement évolué depuis. Aujourd'hui, les individus sont incités à s'éduquer d'avantage, dans un contexte économique très tendu et dans lequel seule une partie des plus instruits arrivent à s'introduire sur le marché de l'emploi.

7 Conclusions

Le présent travail tend à apporter de nouvelles preuves concernant la modélisation de l'éducation comme variable explicative du salaire des individus en France, étayées par l'enquête FQP 2003 de l'INSEE. Nous avons d'abord repris les travaux initiateurs de Mincer [19] dans les années 60. Cependant, comme nous l'avons vu, cette modélisation est sujette à de nombreux biais économétriques. Bien que la méthode des variables instrumentales améliore la spécification du salaire et de l'éducation, nous observons des résultats contradictoires dans la littérature. En effet, nous observons dans notre base que l'estimateur des MCO est biaisé vers le bas. Mais nous ne pouvons affirmés que ce résultat est définitif puisque d'autres auteurs arrivent à la conclusion inverse. Ce débat relatifs aux VI nous prouve que cette méthodologie ne parvient pas à corriger l'ensemble des biais existants dans la modélisation de l'éducation.

Bien que nous observions des différences entre les MCO et les VI, aucune des deux spécifications ne répond à la question de la linéarité du rendement de l'éducation. Une piste nous est fournie par la variable *educan*2 - nombre d'années d'études au carré - puisque le coefficient associé n'est pas nul ce qui prouve que le rendement est concave. La modélisation de Schady que nous avons appliquée aux données françaises permet d'apporter une preuve tangible de la non linéarité du rendement de l'éducation : les années consécutives du baccalauréat, de la licence et du master constituent des années charnières dans le parcours universitaire des étudiants français. D'une part, ces années n'apportent pas le même rendement sur le salaire - respectivement 1,0300%, 0,8967% et 1,0224% par année d'études. Ce gain est décroissant selon le niveau d'éducation atteint. D'autre part, nous constatons qu'un cycle universitaire non achevé porte préjudice à l'individu. Par exemple, le rendement de la première année de Master est négatif : -6,31%. Ce résultat apporte un réel soutient à la théorie du signal dans le système éducatif français dans lequel seul les années diplômantes sont valorisées sur le marché du travail. Cependant, au même titre que la modélisation par MCO, cette spécification ne résout pas les problèmes d'endogénéité de l'éducation. Bien que le résultat est intéressant, nous devons le considérer comme étant biaisé sans savoir comment le corriger.

Afin de résoudre ce problème d'endogénéité, les modélisations multinomiales apportent de sérieuses réponses. C'est pourquoi nous avons finalement spécifié l'éducation par un probit ordonné en utilisant une variable "niveau d'éducation" définie par l'INSEE - *ddipl*. La comparaison avec les MCO conclut à une estimation biaisée vers le haut - en valeur absolue. Cependant, l'analyse du coefficient de cette variable est plus compliquée que celle de l'éducation calculée en nombre d'années. C'est pourquoi nous avons terminé notre étude par l'analyse des déterminants de l'éducation grâce aux effets marginaux et aux

probabilités estimés par le probit. Contrairement au salaire, les femmes ont un meilleur parcours scolaire que les hommes. Par ailleurs, il existe de sérieuses différences entre un Français d'origine et un étranger : ce dernier a beaucoup moins de chance d'accéder à un haut niveau d'éducation. En outre, l'étude démontre que le contexte familial demeure un facteur très important et déterminant sur la formation des individus. En effet, le fait d'avoir une fratrie et des parents qui ont divorcé durant la scolarité, tend à diminuer la longueur des études. De même, la reproduction sociale est forte : des parents avec un faible niveau d'éducation diminue les chances pour les individus d'obtenir un diplôme du supérieur.

En somme, cette analyse apporte un éclairage nouveau sur la population française. Cependant, nous n'avons pu produire de résultats sur le rendement des différentes années d'éducation par une modélisation économétrique diminuant au maximum les biais existant - notamment celui de l'endogénéité de l'éducation. Le manque de données précises sur les parcours professionnels et sur les diplômes exacts obtenus - rappelons que la variable de l'INSEE ne différencie pas un licencié d'un doctorant - nous semble être le premier obstacle à une modélisation précise du rendement de l'éducation. Une enquête plus approfondie permettrait d'obtenir ces données et de construire un modèle supprimant toute endogénéité de l'éducation.

Il demeure que dans ce travail, nous n'avons pas travaillé sur la question de l'homogénéité des individus concernant leur "habilité à s'éduquer". En effet, toutes les études présentées ici posent l'hypothèse que la non-linéarité est due à l'enseignement reçu au cours des différentes années d'éducation. Nous pourrions supposer qu'en réalité, le rendement est le même pour chaque année d'éducation mais que l'habilité des individus fait varier le rendement obtenu. Une modélisation plus théorique se confrontant aux données réelles pourrait être la prochaine étape des travaux sur la question du gain de l'éducation.

Références

[1] Joshua Angrist and Alan Krueger. Does compulsory school attendance affect schooling and earnings ? Working Paper 653, Princeton University, Department of Economics, Industrial Relations Section., 1990.

[2] Joshua Angrist and Alan Krueger. Empirical strategies in labor economics. Working Paper 780, Princeton University, Department of Economics, Industrial Relations Section., 1998.

[3] Florence Arestoff. Taux de rendement de l'éducation sur le marché du travail d'un pays en développement. une analyse micro-économétrique. Open Access publications from Université Paris-Dauphine urn :hdl :123456789/4924, Université Paris-Dauphine, 2001.

[4] Kenneth J Arrow. Higher education as a filter. *Journal of Public Economics*, 2(3) :193–216, 1973.

[5] Orley Ashenfelter, Colm Harmon, and Hessel Oosterbeek. A review of estimates of the Schooling/Earnings relationship, with tests for publication bias. Working Paper 804, Princeton University, Department of Economics, Industrial Relations Section., 1999.

[6] Orley Ashenfelter and Alan Krueger. Estimates of the economic return to schooling from a new sample of twins. Working Paper 683, Princeton University, Department of Economics, Industrial Relations Section., 1992.

[7] Christian Baudelot and Michel Glaude. Les diplômes se dévaluent-ils en se multipliant ? *Economie et statistique*, 225(1) :3–16, 1989.

[8] Gary S. Becker. Human capital : A theoretical and empirical analysis, with special reference to education. In *Human Capital : A Theoretical and Empirical Analysis, with Special Reference to Education, 2nd ed.*, page 22–0. NBER, 1975.

[9] David Card. Estimating the return to schooling : Progress on some persistent econometric problems. *Econometrica*, 69(5) :1127–1160, 2001.

[10] Kevin J. Denny and Colm P. Harmon. Testing for sheepskin effects in earnings equations : evidence for five countries. *Applied Economics Letters*, 8(9) :635–637, 2001.

[11] John Garen. The returns to schooling : A selectivity bias approach with a continuous choice variable. *Econometrica*, 52(5) :1199–1218, 1984.

[12] Zvi Griliches. Estimating the returns to schooling : Some econometric problems. *Econometrica*, 45(1) :1–22, 1977.

[13] Zvi Griliches and William M. Mason. Education, income, and ability. *The Journal of Political Economy*, 80(3) :S74–S103, 1972.

[14] Marc Gurgand. *Économie de l'éducation*. La découverte, Paris, 2005.

[15] Colm Harmon and Ian Walker. Estimates of the economic return to schooling for the united kingdom. *American Economic Review*, 85(5) :1278–86, 1995.

[16] James Heckman and Solomon Polachek. Empirical evidence on the functional form of the earnings-schooling relationship. *Journal of the American Statistical Association*, 69(346) :350, June 1974.

[17] Thomas Hungerford and Gary Solon. Sheepskin effects in the returns to education. *The Review of Economics and Statistics*, 69(1) :175–77, 1987.

[18] Forquin Jean-Claude. Bourdieu (pierre), passeron (jean-claude). — ˜˜la reproduction. eléments pour une théorie du système d'enseignement˜˜. *Revue française de pédagogie*, 15(1) :39–44, 1971.

[19] Jacob A. Mincer. Schooling, experience, and earnings. NBER books, National Bureau of Economic Research, Inc, 1974.

[20] Norbert Schady. Convexity and sheepskin. 2001.

[21] Ali Skalli. Are successive investments in education equally worthwhile ? endogenous schooling decisions and non-linearities in the earnings–schooling relationship. *Economics of Education Review*, 26(2) :215–231, April 2007.

[22] A Michael Spence. Job market signaling. *The Quarterly Journal of Economics*, 87(3) :355–74, 1973.

[23] Douglas Staiger and James H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3) :557, May 1997.

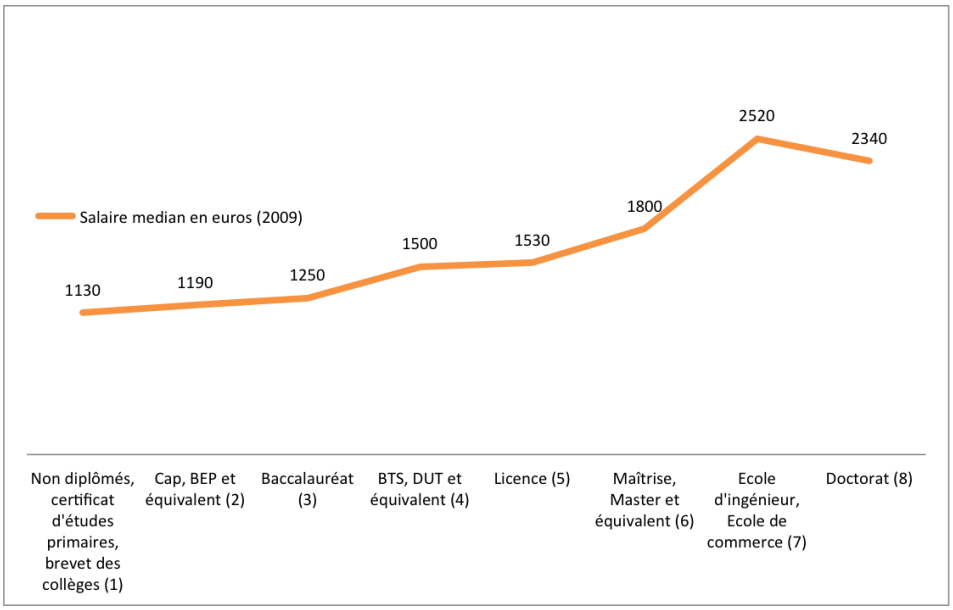
[24] Cecile Van de Velde. *Devenir adulte : sociologie comparee de la jeunesse en Europe*. Le lien social. Presses universitaires de France, Paris, 2008.

[25] Francis Vella and R. G Gregory. Selection bias and human capital investment : Estimating the rates of return to education for young males. *Labour Economics*, 3(2) :197–219, 1996.

[26] Robert J. Willis and Sherwin Rosen. Education and self-selection. NBER Working Paper 0249, National Bureau of Economic Research, Inc, 1978.

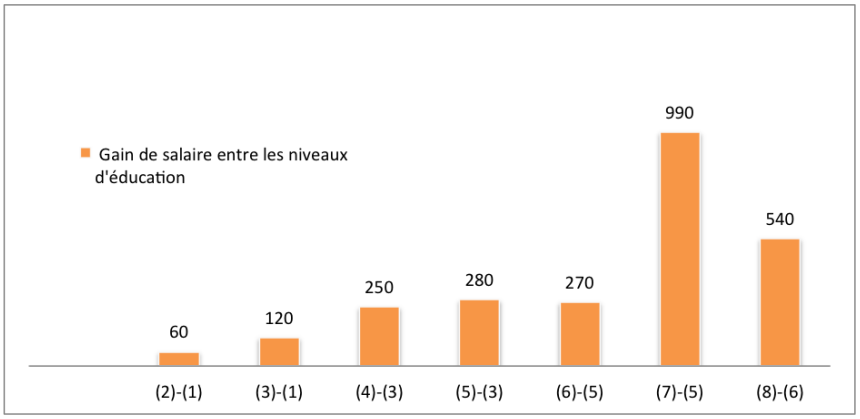
8 Annexes

8.1 Dispersion des salaires par niveau d'éducation



Source : Insee, cumul des enquêtes Emploi de 2003 à 2009.

FIGURE 2 – Salaire médian en euros (2009) en fonction du niveau d'éducation



Source : Insee, cumul des enquêtes Emploi de 2003 à 2009.

FIGURE 3 – Gain de salaire entre les niveaux d'éducation

8.2 Variables

Variables	Modalités	Détails
salredm		Montant du salaire en €
lsalredm		Logarithme du salaire
educan		Nombre d'années d'études
age		Âge
ddipl	1	Diplôme supérieur
	3	Baccalauréat + 2 ans
	4	Baccalauréat ou brevet professionnel ou autre diplôme de ce niveau
	5	CAP, BEP ou autre diplôme de ce niveau
	6	BEPC seul
	7	Aucun diplôme ou CEP
matri	1	Célibataire
	2	Marié
	3	Veuf
	4	Divorcé
nat	1	Français de naissance
	2	Acquisition de la nationalité Française
	3	Reste de l'Europe
	4	Africain
	5	Reste du monde
sexe	0	Femme
	1	Homme
zus	0	Ne vit pas en ZUS
	1	Vit en ZUS
demenage	0	Région de naissance non différente de la région de résidence actuelle
	1	Région de naissance différente de la région de résidence actuelle
trimnais	1	Janvier-Février-Mars
	2	Avril-Mai-Juin
	3	Juillet-Août-Septembre
	4	Octobre-Novembre-Décembre
frat	0	N'a pas de frère ni soeur
	1	A au moins un frère ou une soeur
div	0	Parents ensembles au moment de la scolarité
	1	Parents séparés ou divorcés durant la scolarité

TABLE 2 – Variables

8.3 Statistiques descriptives

Variables	Modalités	Moyenne	Ecart-type	Pourcentage
salredm		1737,1430	2791,3380	
lsalredm		7,1946	0,7123	
educan		12,8226	3,0782	
age		40,1713	10,474	
ddipl	1			11,39
	3			11,59
	4			16,48
	5			26,51
	6			9,19
	7			24,84
matri	1			32,98
	2			57,05
	3			1,53
	4			8,44
nat	1			92,12
	2			3,58
	3			2,37
	4			1,28
	5			0,65
sexe	0			48,65
	1			51,35
zus	0			94,09
	1			5,91
demenage	0			67,75
	1			32,25
trimnais	1			25,32
	2			25,80
	3			24,80
	4			24,08
frat	0			9,28
	1			90,72
div	0			91,19
	1			8,81

TABLE 3 – Etat de l'échantillon

8.4 Analyses

8.4.1 MCO Educan

	1	2	3	4	5
educan	0,0578***	0,0801***	0,0758***	0,0715***	0,0822***
	(0,0015)	(0,0016)	(0,0103)	(0,0099)	(0,0099)
educan2			-0,0001***	-0,0003	-0,0002
			(0,0004)	(0,0004)	(0,0004)
age		0,0181***	0,0576***	0,0531***	0,0521***
		(0,0005)	(0,0032)	(0,0032)	(0,0032)
age2			-0,0005***	-0,0004***	-0,0004***
			(0,0000)	(0,0000)	(0,0000)
matri1				-0,0451***	-0,0570***
				(0,0099)	(0,0097)
matri3				-0,1125***	-0,1072***
				(0,0391)	(0,0388)
matri4				0,0298**	0,0253*
				(0,0151)	(0,0149)
natio2				-0,0946***	-0,1242***
				(0,0237)	(0,0236)
natio3				-0,0273	-0,0735**
				(0,0311)	(0,0310)
natio4				-0,3242***	-0,3748***
				(0,0417)	(0,0411)
natio5				-0,1191**	-0,1935***
				(0,0585)	(0,0592)
sexe				0,3907***	0,3934***
				(0,0084)	(0,0083)
zus				-0,0629***	-0,0948***
				(0,0185)	(0,0184)
demenage				0,0907***	0,0561***
				(0,0097)	(0,0098)
Régions	Non	Non	Non	Non	Oui
Nb Obs	23 665	23 665	23 665	23 665	23 665
R^2	0,0623	0,1240	0,1300	0,21206	0,2339

Note : le test de significativité est précisé par le nombre d'étoiles représentant le seuil de significativité. Aucune : non significatif, * : 10%, ** : 5%, *** : 1%. L'écart-type est précisé entre parenthèses.

TABLE 4 – MCO *educan*

8.4.2 MCO Ddipl

	1	2	3	4
ddipl	-0,1339***	-0,1336***	-0,1372***	-0,1314***
	(0,0024)	(0,0024)	(0,0024)	(0,0024)
age	0,0150***	0,0588***	0,0557***	0,0540***
	(0,0004)	(0,0032)	(0,0032)	(0,0031)
age2		-0,0005***	-0,0005***	-0,0005***
		(0,0000)	(0,0000)	(0,0000)
matri1			-0,0292***	-0,0410***
			(0,0097)	(0,0096)
matri3			-0,1163***	-0,1131***
			(0,0390)	(0,0386)
matri4			0,0396***	0,0358**
			(0,0150)	(0,0148)
natio2			-0,0890***	-0,1188***
			(0,0236)	(0,0235)
natio3			-0,0833***	-0,1317**
			(0,0299)	(0,0298)
natio4			-0,3213***	-0,3749***
			(0,0428)	(0,0424)
natio5			-0,1873***	-0,2585***
			(0,0584)	(0,0592)
sexe			0,4071***	0,4088***
			(0,0083)	(0,0082)
zus			-0,0529***	-0,0842***
			(0,0182)	(0,0181)
demenage			0,0819***	0,0507***
			(0,0095)	(0,0097)
Régions	Non	Non	Non	Oui
Nb Obs	23 665	23 665	23 665	23 665
R^2	0,1363	0,1437	0,2298	0,2500

Note : le test de significativité est précisé par le nombre d'étoiles représentant le seuil de significativité. Aucune : non significatif, * : 10%, ** : 5%, *** : 1%. L'écart-type est précisé entre parenthèses.

TABLE 5 – MCO *ddipl*

8.4.3 MCO Schady

	1	2
educan	0,0612***	0,0632***
	(0,0052)	(0,0052)
D12	0,0443***	0,0404**
	(0,0168)	(0,0166)
D15	0,0591***	0,0512***
	(0,0208)	(0,0205)
D17	0,1016***	0,0905
	(0,0273)	(0,0268)
educanD12	0,0237***	0,0201**
	(0,0095)	(0,0094)
educanD15	-0,0625***	-0,0631***
	(0,0126)	(0,0124)
age	0,0554***	0,0543***
	(0,0032)	(0,0032)
age2	-0,0005***	-0,0005***
	(0,0000)	(0,0000)
matri1	-0,0444***	-0,0563***
	(0,0099)	(0,0097)
matri3	-0,1150***	-0,1096***
	(0,0390)	(0,0387)
matri4	0,0323**	0,0277*
	(0,0151)	(0,0149)
natio2	-0,0982***	-0,1278***
	(0,0237)	(0,0236)
natio3	-0,0403	-0,0867***
	(0,0311)	(0,0309)
natio4	-0,3321***	-0,3831***
	(0,0417)	(0,0411)
natio5	-0,1245**	-0,1993***
	(0,0584)	(0,0591)
sexe	0,3937***	0,3963***
	(0,0084)	(0,0083)
zus	-0,0633***	-0,0951***
	(0,0185)	(0,0184)
demenage	0,0902***	0,0546***
	(0,0097)	(0,0098)
Régions	Non	Oui
Nb Obs	23 665	23 665
R^2	0,2128	0,2359

Note : le test de significativité est précisé par le nombre d'étoiles représentant le seuil de significativité. Aucune : non significatif, * : 10%, ** : 5%, *** : 1%. L'écart-type est précisé entre parenthèses.

TABLE 6 – MCO Schady

8.4.4 VI

		1ère étape : *educan*			2nde étape : *lsalredm*		
		1	2	3	1	2	3
educan					0,1190***	0,1207***	0,1105***
					(0,0042)	(0,0042)	(0,0042)
age		0,0456***	0,0426***	0,0424***	0,0545***	0,0545***	0,0536***
		(0,0134)	(0,0134)	(0,0134)	(0,0033)	(0,0033)	(0,0033)
age2		-0,0015***	-0,0015***	-0,0015***	-0,0004***	-0,0004***	-0,0004***
		(0,0002)	(0,0002)	(0,0002)	(0,0000)	(0,0000)	(0,0000)
matri1		0,1463***	0,1511***	0,1303***	-0,0509***	-0,0513***	-0,0601***
		(0,0450)	(0,0440)	(0,0439)	(0,0103)	(0,0103)	(0,0101)
matri3		-0,7616***	-0,7546***	-0,7080***	-0,0772**	-0,0758**	-0,0792**
		(0,1525)	(0,1522)	(0,1515)	(0,0401)	(0,0401)	(0,0396)
matri4		-0,2945***	-0,2850***	-0,2822***	0,0466***	0,0470***	0,0421***
		(0,0628)	(0,0626)	(0,0623)	(0,0156)	(0,0156)	(0,0153)
natio2		-0,6685***	-0,6713***	-0,7241***	-0,0607**	-0,0587**	-0,0959***
		(0,1117)	(0,1116)	(0,1117)	(0,0254)	(0,0255)	(0,0252)
natio3		-2,2742***	-2,2746***	-2,3308***	0,0815**	0,0861***	0,0188
		(0,1417)	(0,1414)	(0,1437)	(0,0346)	(0,0347)	(0,0340)
natio4		-1,3177***	-1,3273***	-1,3981***	-0,2325***	-0,2289***	-0,2949***
		(0,2353)	(0,2356)	(0,2360)	(0,0459)	(0,0461)	(0,0447)
natio5		-1,8862***	-1,9015***	-1,9837***	-0,0202	-0,0170	-0,1025*
		(0,2873)	(0,2893)	(0,2896)	(0,0623)	(0,0625)	(0,0625)
sexe		-0,2271***	-0,2291***	-0,2219***	0,4001***	0,4006***	0,4010***
		(0,0348)	(0,0347)	(0,0346)	(0,0087)	(0,0087)	(0,0085)
zus		-0,5604***	-0,5557***	-0,5957***	-0,0257	-0,0244	-0,0589***
		(0,0806)	(0,0804)	(0,0804)	(0,0196)	(0,0196)	(0,0194)
demenage		0,4190***	0,4328***	0,3993***	0,0561***	0,0546***	0,0296***
		(0,0413)	(0,0412)	(0,0423)	(0,0105)	(0,0105)	(0,0105)
trimnais2		0,0869**	0,0825*	0,0759			
		(0,0483)	(0,0482)	(0,0479)			
trimnais3		0,0462	0,0457	0,0405			
		(0,0486)	(0,0498)	(0,0483)			
trimnais4		0,1765***	0,1748***	0,1691***			
		(0,0499)	(0,0498)	(0,0496)			
ddiplm		-0,4218***	-0,4237***	-0,4125***			
		(0,0150)	(0,0150)	(0,0150)			
ddiplp		-0,4029***	-0,4016***	-0,3944***			
		(0,0128)	(0,0128)	(0,0127)			
frat			-0,5519***	-0,5349***			
			(0,0647)	(0,0648)			
div			-0,4075***	-0,3945***			
			(0,0610)	(0,0609)			
Régions		Non	Non	Oui	Non	Non	Oui
Nb Obs		22 650	22 650	22 650	22 650	22 650	22 650
R^2		0,2864	0,2903	0,2979	0,1898	0,1855	0,2171
Test d'endogénéité	p-value				0,00	0,00	0,00
Test de suridentification	p-value				0,73	0,05	0,23
Efficience des VI	p-value				0,00	0,00	0,00
	F stat.				781	588	541

Note : le test de significativité est précisé par le nombre d'étoiles représentant le seuil de significativité. Aucune : non significatif, * : 10%, ** : 5%, *** : 1%. L'écart-type est précisé entre parenthèses.

Table 7 – VI

8.4.5 Probit

	1ère étape : *ddipl*		2ènde étape : *lsalredm*		MCO	
	1	2	1	2	MCO 3	MCO 4
ddipl			-0,1266***	-0,1128***	-0,1372***	-0,1314***
			(0,0051)	(0,0051)	(0,0024)	(0,0024)
age	-0,0091*	-0,0084	0,0539***	0,0532***	0,0557***	0,0540***
	(0,0054)	(0,0054)	(0,0035)	(0,0035)	(0,0032)	(0,0031)
age2	0,0003***	0,0003***	-0,0005***	-0,0005***	-0,0005***	-0,0005***
	(0,0001)	(0,0001)	(0,0000)	(0,0000)	(0,0000)	(0,0000)
matri1	0,0275	0,0389**	-0,0259***	-0,0381***	-0,0292***	-0,0410***
	(0,0179)	(0,0180)	(0,0105)	(0,0103)	(0,0097)	(0,0096)
matri3	0,2989***	0,2794***	-0,0940**	-0,0961**	-0,1163***	-0,1131***
	(0,0617)	(0,0617)	(0,0407)	(00404)	(0,0390)	(0,0386)
matri4	0,1574***	0,1591***	0,0474***	0,0436***	0,0396***	0,0358**
	(0,0265)	(0,0266)	(0,0165)	(0,0162)	(0,0150)	(0,0148)
natio2	0,2702***	0,3011***	-0,1089***	-0,1412***	-0,0890***	-0,1188***
	(0,0399)	(0,0401)	(0,0244)	(0,0244)	(0,0236)	(0,0235)
natio3	0,6991***	0,7320***	-0,0287	-0,0970***	-0,0833***	-0,1317**
	(0,0516)	(0,0518)	(0,0324)	(0,0327)	(0,0299)	(0,0298)
natio4	0,5684***	0,6123***	-0,2750***	-0,3498***	-0,3213***	-0,3749***
	(0,0685)	(0,0688)	(0,0442)	(0,0435)	(0,0428)	(0,0424)
natio5	0,4517***	0,5046***	-0,1678**	-0,2325***	-0,1873***	-0,2585***
	(0,0958)	(0,0961)	(0,0587)	(0,0601)	(0,0584)	(0,0592)
sexe	0,1690***	0,1663***	0,4033***	0,4041***	0,4071***	0,4088***
	(0,0143)	(0,0143)	(0,0090)	(0,0089)	(0,0083)	(0,0082)
zus	0,2738***	0,3003***	-0,0401**	-0,0711***	-0,0529***	-0,0842***
	(0,0317)	(0,0319)	(0,0195)	(0,0195)	(0,0182)	(0,0181)
demenage	-0,1797***	-0,1571***	0,0855***	0,0557***	0,0819***	0,0507***
	(0,0166)	(0,0171)	(0,0107)	(0,0108)	(0,0095)	(0,0097)
trimnais2	0,0179	0,0206				
	(0,0198)	(0,0198)				
trimnais3	0,0236	0,0260				
	(0,0200)	(0,0200)				
trimnais4	0,0004	-0,0031				
	(0,0202)	(0,0202)				
ddiplm	0,1892***	0,1855***				
	(0,0063)	(0,0063)				
ddiplp	0,1596***	0,1567***				
	(0,0052)	(0,0052)				
frat	0,2498***	0,2358***				
	(0,0250)	(0,0252)				
div	0,2545***	0,2526***				
	(0,0255)	(0,0255)				
Régions	Non	Oui	Non	Oui	Non	Oui
Nb Obs	22 920	22 920	22 920	22 920	23 665	23 665
(Pseudo) R^2	0,0770	0,0799	0,1375	0,1630	0,2298	0,2500

Note : le test de significativité est précisé par le nombre d'étoiles représentant le seuil de significativité. Aucune : non significatif, * : 10%, ** : 5%, *** : 1%. L'écart-type est précisé entre parenthèses.

TABLE 8 – Probit ddipl

8.4.6 Déterminants de l'éducation

ddipl =	1	3	4	5	6	7
age	0,0013	0,0011	0,0009	-0,0003	-0,0005	-0,0024
	(0,0008)	(0,0007)	(0,0006)	(0,0002)	(0,0003)	(0,0015)
age2	-0,0001	0,0000	0,0000	0,0000	0,0000	0,0001
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)
matri1	-0,0059	-0,0051	-0,0040	0,0013	0,0025	0,0111
	(0,0027)	(0,0023)	(0,0019)	(0,0006)	(0,0012)	(0,0051)
matri3	-0,0422	-0,0364	-0,0289	0,0097	0,0181	0,0798
	(0,0093)	(0,0081)	(0,0064)	(0,0022)	(0,0040)	(0,0176)
matri4	-0,0240	-0,0207	-0,0164	0,0055	0,0103	0,0454
	(0,0040)	(0,0035)	(0,0028)	(0,0010)	(0,0017)	(0,0076)
natio2	-0,0455	-0,0393	-0,0311	0,0104	0,0195	0,0860
	(0,0061)	(0,0053)	(0,0042)	(0,0015)	(0,0026)	(0,0115)
natio3	-0,1106	-0,0954	-0,0756	0,0253	0,0473	0,2091
	(0,0080)	(0,0069)	(0,0056)	(0,0024)	(0,0035)	(0,0149)
natio4	-0,0925	-0,0798	-0,0633	0,0212	0,0396	0,1749
	(0,0105)	(0,0091)	(0,0072)	(0,0027)	(0,0046)	(0,0197)
natio5	-0,0763	-0,0658	-0,0521	0,0174	0,0326	0,1441
	(0,0146)	(0,0126)	(0,0100)	(0,0035)	(0,0063)	(0,0275)
sexe	-0,0251	-0,0217	-0,0172	0,0057	0,0108	0,0475
	(0,0022)	(0,0019)	(0,0015)	(0,0006)	(0,0010)	(0,0041)
zus	-0,0454	-0,0391	-0,0310	0,0104	0,0194	0,0858
	(0,0049)	(0,0042)	(0,0034)	(0,0013)	(0,0021)	(0,0091)
demenage	0,0237	0,0205	0,0162	-0,0054	-0,0102	-0,0449
	(0,0026)	(0,0023)	(0,0018)	(0,0007)	(0,0011)	(0,0049)
trimnais2	-0,0031	-0,0027	-0,0021	0,0007	0,0013	0,0059
	(0,0030)	(0,0026)	(0,0021)	(0,0007)	(0,0013)	(0,0057)
trimnais3	-0,0039	-0,0034	-0,0027	0,0009	0,0017	0,0074
	(0,0030)	(0,0026)	(0,0021)	(0,0007)	(0,0013)	(0,0057)
trimnais4	-0,0005	-0,0004	-0,0003	0,0001	0,0002	0,0009
	(0,0031)	(0,0026)	(0,0021)	(0,0007)	(0,0013)	(0,0058)
ddiplm	-0,0280	-0,0242	-0,0192	0,0064	0,0120	0,0530
	(0,0010)	(0,0009)	(0,0008)	(0,0005)	(0,0005)	(0,0018)
ddiplp	-0,0237	-0,0204	-0,0162	0,0054	0,0101	0,0447
	(0,0009)	(0,0008)	(0,0006)	(0,0004)	(0,0004)	(0,0015)
frat	-0,0356	-0,0307	-0,0244	0,0081	0,0152	0,0673
	(0,0038)	(0,0033)	(0,0027)	(0,00010)	(0,0017)	(0,0072)
div	-0,0382	-0,0329	-0,0261	0,0087	0,0163	0,0721
	(0,0039)	(0,0034)	(0,0027)	(0,0011)	(0,0017)	(0,0073)

TABLE 9 – Taux de rendement

Effectif	ddipl =	1*	3*	4*	5*	6*	7*
130	Homme	17,70	14,47	18,64	25,16	7,72	16,31
107	Femme	22,83	16,93	19,77	23,17	6,23	11,06

Note : probabilité concernant les Français de naissance, mariés, qui ont déménagé, sans fratrie, ne vivant pas en ZUS et dont les parents n'ont pas divorcé.
* : l'hypothèse d'égalité des probabilités est rejettée.

TABLE 10 – Probabilités en % pour *sexe*

Effectif	ddipl =	1*	3*	4*	5*	6*	7*
3 762	$demenage = 0$	7,60	9,44	15,61	27,65	10,46	29,25
1 378	$demenage = 1$	13,81	12,53	17,52	26,21	8,79	21,14

Note : probabilité concernant les hommes français de naissance, mariés, avec fratrie, dont les parents n'ont pas divorcé et ne vivant pas en ZUS.

* : l'hypothèse d'égalité des probabilités est rejettée.

TABLE 11 – Probabilités en % pour *demenage*

Effectif	ddipl =	1*	3*	4*	5	6*	7*
130	$frat = 0$	17,70	14,47	18,64	25,16	7,72	16,32
1 378	$frat = 1$	13,81	12,53	17,52	26,21	8,79	21,14

Note : probabilité concernant les hommes français de naissance, mariés, dont les parents n'ont pas divorcé, ne vivant pas en ZUS et qui ont déménagé.

* : l'hypothèse d'égalité des probabilités est rejettée.

TABLE 12 – Probabilités en % pour $frat$

Effectif	ddipl =	1*	3*	4*	5*	6*	7*
130	$natio1 = 0$	7,60	9,44	15,61	27,65	10,46	29,25
107	$natio1 = 1$	13,81	12,53	17,52	26,21	8,79	21,14

Note : probabilité concernant les hommes mariés, avec fratrie, dont les parents n'ont pas divorcé, ne vivant pas en ZUS et ayant déménagé.

* : l'hypothèse d'égalité des probabilités est rejettée.

TABLE 13 – Probabilités en % pour $natio1$