

La mobilité sociale et son évolution : le rôle des anticipations réexaminé

Dominique GOUX, Éric MAURIN *

RÉSUMÉ. – Le niveau de formation atteint par chacun dépend en partie de ce qu'il anticipe en retour de ses efforts dans le système éducatif alors qu'il est encore étudiant. Nous montrons qu'en négligeant le rôle tenu par les anticipations des étudiants, les mesures statistiques traditionnelles des inégalités devant l'enseignement surestiment la part de l'école dans les mécanismes de reproduction des inégalités. Elles exposent également à fournir une évaluation erronée de l'évolution de l'inégalité des chances avec le temps.

Social Mobility and its Evolution: The Role of Expectations Reexamined

ABSTRACT. – The number of years spent in the educational system depends on the expected returns to education. We show that neglecting the role of expectations, usual statistical measures of inequalities in educational opportunities strongly overestimate social determinism within the educational system. Usual measures also provide biased estimates of the evolution of inequalities in educational opportunities over time.

* D. GOUX : INSEE, Division Emploi ; É. MAURIN : CREST-INSEE, Département de la Recherche de l'INSEE.

Une version antérieure de cette étude a été présentée au *Séminaire Recherche de l'INSEE* (mars 1999). Nous remercions les deux rapporteurs et le rédacteur en chef de la revue pour leur relecture attentive et les améliorations qu'ils ont permis d'apporter à ce travail.

1 Introduction

Il n'est pas facile de réussir à l'Université ou d'obtenir un diplôme d'une Grande École. Il s'agit d'objectifs de longue haleine, demandant efforts et sacrifices. Les étudiants sont d'autant moins hésitants à les consentir qu'ils envisagent une bonne insertion professionnelle et une pleine reconnaissance de leurs titres scolaires. De ce point de vue, une origine sociale modeste peut constituer un handicap et freiner l'ambition scolaire, puis universitaire, des étudiants. Comment accepter de sacrifier plusieurs années de sa vie à l'école si l'on anticipe ne pas avoir les moyens (en l'occurrence le capital social) de faire plus tard une carrière à la hauteur de son diplôme ?

Dans la mesure où ils ne peuvent pas disposer des atouts permettant de faire fructifier plus tard un diplôme du supérieur, beaucoup d'élèves d'origine modeste ont objectivement moins intérêt à se lancer dans des études longues et à persévérer à l'école, même quand ils y réussissent aussi bien que les autres. En d'autres termes, une partie des inégalités de diplôme entre personnes d'origine aisée et personnes d'origine modeste n'est potentiellement pas liée au fonctionnement du système éducatif lui-même, mais trouve ses racines dans une forme tout à fait différente d'injustice sociale : un même niveau de formation ne conduit pas nécessairement aux mêmes carrières professionnelles les personnes d'origine aisée et celles d'origines modeste.

Comme le soulignent HAVEMAN et WOLFE [1995], l'immense majorité de la littérature sur l'inégalité des chances néglige ce processus. Dans cet article, nous montrons que ce parti-pris expose à fournir une image incomplète et difficilement interprétable de l'inégalité des chances à l'école. Il expose également à des diagnostics trompeurs sur l'évolution des inégalités sociales entre personnes d'origines différentes.

Pour illustrer cette thèse, nous développons un modèle de mobilité sociale très simple, explicitant la façon dont l'inégalité des chances sur le marché du travail affecte les anticipations et les décisions des étudiants. Le modèle est estimé en utilisant une base de données françaises nouvelle. L'identification des paramètres structurels repose sur un ensemble d'informations sur l'enfance et l'adolescence des répondants, permettant de construire des variables instrumentales originales.

Nos résultats suggèrent que l'analyse statistique habituelle surestime fortement l'impact structurel de l'origine sociale sur la scolarité. Selon nos résultats, les tables de mobilité et les mesures habituelles de l'inégalité des chances à l'école offrent une image déformée du rôle exact de l'école dans les processus de reproduction des inégalités. Il est hasardeux de se fonder sur ces mesures pour évaluer les politiques éducatives en faveur de l'égalité des chances, mais également pour émettre un diagnostic sur les tendances de la fluidité sociale.

1.1 L'inégalité des chances et son évolution

Le premier problème étudié dans cet article est le suivant : les enfants issus de milieux modestes restent moins longtemps à l'école, obtiennent des

diplômes moins élevés, puis des revenus plus faibles. Qu'est-ce qui, dans cette inégalité des chances, est dû au fonctionnement du système scolaire ? Qu'est-ce qui est dû aux autres institutions, et en particulier au marché du travail ?

La seconde interrogation porte sur l'évolution des inégalités entre individus d'origines différentes. Supposons qu'entre deux périodes, on observe une baisse des inégalités entre personnes d'origine modeste et personnes d'origine aisée. Comment l'interpréter ? Est-ce parce que les enfants d'origine aisée ont cessé d'être avantagés culturellement et matériellement dans leurs parcours scolaire et professionnel (société plus fluide) ? Ou bien est-ce parce que la valeur sociale des avantages dont ils continuent d'hériter a changé ? Est-ce parce que l'héritage culturel et social est moins inégalement réparti, ou bien est-ce parce que l'importance sociale de ce dont quelques-uns héritent a diminué par rapport à la valeur de ce dont tout le monde peut bénéficier ?

Nous voulons mettre l'accent sur un aspect de ces questions qui nous paraît à la fois essentiel et négligé : pour comprendre où et comment se créent les inégalités entre enfants d'origines différentes et pour interpréter la façon dont évoluent ces inégalités, on ne peut pas faire l'économie d'une théorie des conduites individuelles. Il est, en particulier, tout à fait crucial de modéliser la façon dont se forment les anticipations des étudiants et le rôle de ces anticipations.

Sans une telle théorie, on confond les inégalités dues au fonctionnement du système scolaire et celles générées par le marché du travail : il est impossible d'isoler et d'évaluer le rôle de l'école dans les processus de transmission des inégalités. Sans une telle théorie, il est également impossible d'interpréter l'évolution des inégalités sociales entre personnes issues de milieux différents. Une baisse du lien statistique entre origine et destinée sociale peut aussi bien refléter le passage à une société plus fluide, moins sélective pour les personnes d'origine modeste, qu'un renforcement général (*ie*, valable pour tous) de la valeur sociale des diplômes. Tout ce qui permet d'accroître cette valeur (par exemple : une innovation technologique) contribue à attirer dans l'enseignement secondaire et supérieur des catégories d'étudiants qui ne s'y seraient pas aventurés sinon. Une évolution bénéficiant de la même façon à tout le monde peut ainsi très bien faire reculer les inégalités de diplôme et faire disparaître leurs effets sur le marché du travail. Un même ensemble de statistiques sur l'évolution des inégalités selon l'origine sociale peut ainsi prendre un sens très différent selon que l'on suppose ou non les étudiants capables d'adapter leur conduite aux changements de contextes technologique et institutionnel.

Il est clair que ces idées ne sont pas nouvelles. À notre connaissance, elles ne se sont toutefois jamais réellement trouvées traduites, ni testées, dans les modèles empiriques de la sociologie appliquée. De fait, dans la tradition sociologique, les efforts méthodologiques ont d'abord pour objectif de résumer de la façon la plus pertinente possible la complexité de l'information statistique apportée par les données d'enquêtes¹. En général, les tests mis en œuvre visent avant tout à évaluer et comparer différentes formes de *simplifi-*

1. Pour des avancées récentes de ce type, voir XIE [1992] ou WONG [1992]. Parmi les applications de l'analyse statistique, on peut citer ERIKSON et GOLDTHORPE [1992] et GANZEBOOM, LUIJKX et TREIMAN [1989] pour des comparaisons internationales et GOUX et MAURIN [1997] pour une analyse du cas français.

cation de l'information disponible. Tout le courant d'analyse log-linéaire ou log-multiplicative des tables de mobilité se situe dans cette perspective. Il est très rare de voir pris en compte les comportements individuels et leur interaction avec les institutions. Du coup, les résultats existants sont difficiles à interpréter et n'ont qu'un usage fragile en matière d'évaluation des politiques publiques ou de recommandation de politique sociale en faveur d'une plus grande égalité des chances. Toute mesure empirique des corrélations entre l'origine familiale et les destinées scolaire et sociale pose de sérieux problèmes d'interprétation dès l'instant où l'on n'a pas explicité le modèle de comportement sous-jacent.

1.2 Un modèle de mobilité sociale avec anticipations

S'agissant du modèle de comportement proposé dans cet article, notre objectif n'est pas de valider une théorie microéconomique ou microsociologique nouvelle. Pour que la démarche reste la plus claire possible, nous allons au contraire nous en tenir à un modèle simple. Il peut se résumer en une seule hypothèse fondamentale : au-delà des différentes formes de déterminisme social, les individus passent d'autant plus de temps à l'école que les bénéfices anticipés sur le marché du travail sont importants. Si, d'une génération à l'autre, l'éducation cesse d'être un atout sur le marché du travail, alors les individus passeront moins de temps dans le système scolaire. Si, en revanche, l'éducation donne un avantage salarial croissant aux diplômés, alors les individus resteront plus longtemps à l'école.

Dès l'instant où l'on prend cette hypothèse au sérieux, les inégalités de diplôme entre personnes d'origines différentes ne s'interprètent plus nécessairement en termes d'inégalité des chances dans le système scolaire. Les mécanismes en jeu sont les suivants : supposons que l'enseignement supérieur soit parfaitement méritocratique, coûte la même chose pour tout le monde (disons C) et multiplie les perspectives salariales de chaque bachelier par 1.5. Sous ces hypothèses, l'Université ne génère aucune inégalité particulière entre personnes d'origines différentes. Mais supposons que – du fait d'une meilleure information sur le fonctionnement du marché du travail par exemple – les personnes d'origine aisée gagnent en moyenne 20 % de plus que les autres, à diplôme égal. En d'autres termes, la richesse que multiplie l'Université par 1.5 est en moyenne 1.2 fois plus élevée pour les personnes d'origine aisée. Sous ces hypothèses, il y a plus de chance que l'accroissement de richesse généré par l'Université soit supérieur à C pour les étudiants d'origine aisée que pour les étudiants d'origine modeste. À la limite seules les personnes d'origine aisée trouveront intérêt à entreprendre des études supérieures, alors même que l'Université fonctionne de la même façon pour tous.

En poursuivant sur cet exemple, on constate également que si du jour au lendemain (suite à une innovation institutionnelle ou technologique très favorable aux diplômés) l'Université multiplie les perspectives salariales non plus par 1.5, mais par 3, tout le monde peut soudain trouver intérêt à poursuivre ses études dans le supérieur, y compris les personnes d'origine modeste. Les inégalités devant le diplôme et le revenu s'effondrent, alors même que rien n'aura changé dans le fonctionnement du système scolaire ou dans la sélectivité du marché du travail.

L'article s'organise de la façon suivante : nous commençons par un survol de la littérature sur les liens entre mobilité sociale et anticipations (section 2). Ensuite, nous spécifions notre modèle de détermination simultanée de la durée des études et des revenus du travail (section 3). Nous testons la cohérence et le réalisme empirique de ce modèle à l'aide de données longitudinales françaises récentes (section 4). L'estimation du modèle confirme que l'approche statistique habituelle donne une vision incomplète et difficilement interprétable des processus de mobilité sociale et de leur évolution (section 5). Ce travail invite à poursuivre un programme de recherche dont, pour finir, nous essayons de donner quelques-uns des principaux éléments (section 6).

2 Survol de la littérature

Dans la littérature économique, l'idée selon laquelle le diplôme et le statut social (ou le revenu) se déterminent simultanément est assez ancienne. Elle a surtout été développée pour l'analyse de la valeur de l'éducation sur le marché du travail. Il est clair qu'en négligeant ce qui détermine simultanément les diplômes et la qualité de la carrière professionnelle, on risque de mal identifier l'impact réel du diplôme sur les salaires. Les premières justifications formelles de ce type d'arguments remontent à WILLIS et ROSEN [1979]. On retrouve également ce type de modèle sous différentes formes dans les travaux de MANSKI et WISE [1983], GAREN [1984], HARTOG *et alii* [1989] ou HAVEMAN *et alii* [1995].

En revanche, l'immense majorité de la littérature sur la mobilité sociale néglige le rôle des anticipations et la simultanéité de la détermination de l'éducation et du revenu. De fait, au-delà des débats sur la bonne mesure du statut social², on peut considérer sans trop simplifier que le paradigme dominant reste celui de BLAU et DUNCAN [1967]. Dans sa version originale, la structure causale du modèle de BLAU et DUNCAN se confond avec la structure chronologique des processus : le statut atteint par le père détermine le niveau d'éducation atteint par le répondant, lequel détermine le niveau d'entrée sur le marché du travail, lequel détermine à son tour le statut social atteint en cours de carrière (BLAU et DUNCAN, chapitre 5). Les résultats obtenus à chaque étape sont conditionnés directement et univoquement par les résultats des étapes précédentes. Il n'y a aucune place pour les processus d'anticipation, ni pour la détermination simultanée des différents processus.

L'idée d'une structure causale plus riche n'est pourtant pas absente du livre de BLAU et DUNCAN. Lorsque vient le moment non plus de décrire, mais d'expliquer la mobilité des migrants, les deux auteurs se détachent clairement d'une simple vision chronologique (BLAU et DUNCAN, chapitre 12, page 427) :

« Occupational opportunities will unquestionably continue to vary substantially in different places and they will continue to change as new

2. Dans cet article, nous évaluons le statut social à l'aide du revenu. Les arguments que nous mettons en place et testons vaudraient tout autant si nous mesurions le statut social à l'aide d'une variable discrète de classe sociale ou d'une variable continue de prestige social.

industries develop in some urban centers and technological advances make the industrial activities in another obsolete. These variations give men incentive to migrate from areas with lesser to areas with better opportunities and the flow of migrants from disadvantaged environments acts as a catalyst for occupational mobility ».

C'est bien la structure anticipée des opportunités qui détermine les migrants à se déplacer et qui au bout du compte façonne les processus observés de mobilité sociale.

Quelques années plus tard, cette idée est également au cœur des réflexions de R. BOUDON sur l'école (BOUDON [1974]). Selon lui, les conduites scolaires découlent non seulement de l'origine sociale, mais également de la perception qu'ont les individus de la valeur future de leurs diplômes. BOUDON ne teste jamais réellement ce schéma théorique. Il le mobilise néanmoins très souvent, qu'il s'agisse d'expliquer les défaillances de la politique éducative de masse ou d'interpréter la stabilité des processus d'hérédité sociale comme équilibre macrosocial entre l'ensemble des comportements individuels (BOUDON [1981]).

Les représentations de l'avenir et les aspirations des étudiants sont également au cœur du fameux « *Wisconsin model* » (modèle du Wisconsin), développé par W. SEWELL et ses collègues aux États-Unis. À partir d'une étude approfondie d'une cohorte de lycéens du Wisconsin, ces travaux révèlent l'importance qu'ont les ambitions sociales et professionnelles des lycéens sur leurs trajectoires professionnelles ultérieures effectives (voir, par exemple, SEWELL et HAUSER [1975], SEWELL, HAUSER et WOLF [1980]).

En prenant en compte des variables psychologiques et en mesurant les aspirations des étudiants, les travaux réalisés à partir du panel du Wisconsin livrent une image nettement plus sophistiquée des processus de mobilité sociale que le modèle traditionnel de BLAU et DUNCAN. La structure causale des modèles développés à partir du panel du Wisconsin reste toutefois purement chronologique : l'origine sociale détermine les aspirations des lycéens, lesquelles contribuent ensuite à déterminer la destinée sociale des étudiants. Les variables d'aspirations ne représentent ainsi qu'un canal supplémentaire par lequel transite l'effet de l'origine sociale sur la réussite professionnelle.

Dans cet article, nous allons développer et tester un modèle de nature différente. Le facteur supplémentaire que nous voulons introduire dans le modèle standard de *status attainment* n'est pas le statut social auquel aspirent les étudiants, mais leurs anticipations concernant l'impact d'un surplus de formation sur leurs revenus futurs.

Dans le modèle du Wisconsin, les aspirations déterminent la réussite scolaire et professionnelle comme n'importe quelle autre facteur exogène. On peut identifier leurs effets à l'aide des techniques usuelles, telle que la méthode des moindres carrés ordinaires (MCO). De même, ces variables d'aspirations peuvent être omises de l'analyse sans que l'on ait à craindre d'obtenir une vision biaisée de l'effet total de l'origine sociale sur la réussite scolaire et professionnelle.

En revanche, le facteur d'anticipation que nous cherchons à prendre en compte correspond à une nouvelle variable endogène : par construction, les anticipations sur le revenu ne peuvent se déterminer indépendamment de la formation. Il est impossible d'estimer leur impact sur les scolarités à l'aide des méthodes de MCO utilisées habituellement pour estimer les modèles de *status*

attainment. Pour identifier leurs effets, il est nécessaire (1) de trouver des variables affectant les anticipations sans avoir d'effet propre sur les scolarités, (2) d'appliquer des techniques de variables instrumentales appropriées. Dans ce cadre d'analyse, on ne peut pas omettre les variables d'anticipations sans s'exposer à une vision biaisée de l'ensemble du processus de mobilité sociale.

3 Le modèle

Nous allons supposer que les choix éducatifs se font pour l'essentiel au même âge a , par exemple au moment de la fin de la scolarité obligatoire. Chaque étudiant compare les revenus que permettent d'anticiper les différentes trajectoires scolaires avec la difficulté de réussir chacune d'elles. Il choisit ensuite celle des trajectoires qui lui semble correspondre au meilleur compromis entre revenus et difficultés anticipés. Dans la suite, nous noterons $c_i(s)$ le coût que représente pour lui de réussir au niveau s du système scolaire. Dans ce modèle, mesurer l'inégalité des chances revient à mesurer l'inégale difficulté de réussir les différentes scolarités possibles, c'est-à-dire l'intensité du lien entre $c_i(s)$ et l'origine sociale. Si l'on note $C_i(S)$ le coût actualisé d'un total de S années d'études tel que l'étudiant i peut l'anticiper à l'âge a , on a,

$$(1) \quad C_i(S) = \int_a^S c_i(s) \exp(-r(s-a)) ds$$

où r est un paramètre d'actualisation. Si l'on note maintenant $W_i(S)$ le revenu actualisé anticipé correspondant à ce même total de S années d'études, on a,

$$(2) \quad W_i(S) = \int_{d_i+S}^{\infty} \bar{R}_i(t, S) \exp(-r(t - (d_i + a))) dt$$

où d_i est la date de naissance de l'étudiant i et $\bar{R}_i(t, S)$ le revenu qu'il anticipe recevoir à la date t s'il entre sur le marché du travail après S années de formation. Le choix éducatif optimal S_i de l'étudiant i correspond à la maximisation de l'objectif actualisé net ($W_i(S) - C_i(S)$). Cette décision optimale vérifie la condition du premier ordre,

$$(3) \quad \frac{\partial C_i(S)}{\partial S} = \frac{\partial W_i(S)}{\partial S}$$

ce qui – après quelques manipulations – se réécrit³,

$$(4) \quad c_i(S) + \bar{R}_i(S) - \int_{d_i+S}^{\infty} \frac{\partial \bar{R}_i(t, S)}{\partial S} \exp(-r(t - (d_i + S))) dt = 0$$

3. Il suffit d'utiliser la formule de dérivation suivante,

$$\frac{d}{ds} \left(\int_A^{A(s)} F(s, t) dt \right) = \frac{dA(s)}{ds} F(s, A(s)) + \int_A^{A(s)} \frac{\partial F(s, t)}{\partial s} dt$$

où $\overline{R}_i(S)$ est une notation simplifiée pour $\overline{R}_i(d_i + S, S)$. $\overline{R}_i(S)$ représente le revenu que i anticipe en début de vie active.

Supposons maintenant que (a) le revenu anticipé pour la date t (ie, $\overline{R}_i(t, S)$) puisse se décomposer de façon log-linéaire en une composante fixe pour l'ensemble de la carrière et une composante mesurant l'impact de l'expérience sur le marché du travail et (b) le rendement anticipé de l'éducation soit le même pour tous les individus (ie, $\frac{\partial \ln \overline{R}_i(S)}{\partial S} = \alpha$). Sous cette hypothèse, on peut écrire,

$$(5) \quad \overline{R}_i(t, S) = \overline{R}_i(S) F(t - (d_i + S)),$$

avec $F(0) = 1$ et, après de nouvelles manipulations, l'équation (4) se réécrit,

$$(6) \quad \ln c_i(S) = \ln \overline{R}_i(S) + c_0$$

où c_0 est une constante⁴. En notant Z_{1i} le milieu social dans lequel l'individu i a effectué sa scolarité obligatoire et en supposant que $c_i(S)$ peut se décomposer de façon log-linéaire selon $\ln c_i(S, Z_{1i}) = \phi S + \gamma Z_{1i} + v_i$, on obtient finalement

$$(7) \quad S = S_0 + \theta \ln \overline{R}_i(S) + \gamma_1 Z_{1i} + \varepsilon_i$$

où $\theta = \frac{1}{\phi}$, $\gamma_1 = \frac{-\gamma}{\phi}$ tandis que $S_0 = \frac{c_0}{\phi}$ est une constante et $\varepsilon_i = \frac{-v_i}{\phi}$ un résidu. C'est de ce modèle dont nous allons chercher à identifier les paramètres structurels. Le paramètre γ_1 représente une mesure du déterminisme social au sein du système éducatif et le principal problème est d'estimer ce paramètre et son évolution. D'un strict point de vue économétrique, le modèle de BLAU et DUNCAN correspond au modèle (7) dans lequel on a omis la variable $\ln \overline{R}_i(S)$. Dès l'instant où l'éducation a réellement un impact sur le revenu et où cet impact est au moins partiellement anticipé et pris en compte par les étudiants, cette variable est corrélée avec l'éducation. L'omettre conduit à une estimation biaisée de γ_1 et donc à une vision déformée de l'inégalité des chances. En annexe, nous détaillons les différents types de biais entraînés par l'oubli de $\ln \overline{R}_i(S)$. Nous allons maintenant détailler la stratégie que nous avons suivie dans cet article pour estimer à la fois θ et les paramètres structurels γ_1 décrivant le lien entre origine sociale et scolarité.

3.1 Identification des paramètres structurels

En premier lieu, il est nécessaire de spécifier la façon dont se forme le revenu en début de vie active ainsi que la manière dont se forment les anticipations des étudiants à propos de ce revenu. Conformément à l'usage, nous supposons que le revenu $R_i(t, S)$ d'un individu entrant sur le marché du

4. Avec $c_0 = \ln \left(\int_0^\infty \left(-\frac{\partial F(u)}{\partial u} + \alpha F(u) - r \right) \exp(-ru) du \right)$.

travail à la date t après S années de formation peut se décomposer de façon log-linéaire selon,

$$(8) \quad \ln R_i(t, S) = \alpha S + \gamma_2 Z_{2i} + v_{it},$$

où Z_{2i} représente les ressources héritées des parents et influant directement sur la réussite professionnelle⁵. La variable v_{it} représente, quant à elle, un aléa supposé *i.i.d.* captant l'effet des facteurs conjoncturels sur le début de carrière. Le paramètre α représente le rendement de l'éducation.

S'agissant des anticipations, nous les supposons rationnelles⁶, ce qui permet d'écrire,

$$(9) \quad \ln \bar{R}_i(S, Z_{2i}) = \alpha S + \gamma_2 Z_{2i} + Cste.$$

Dans ce cadre d'analyse, l'identification du paramètre γ_1 demande simplement de trouver (a) des variables Z_{2i} expliquant les revenus, mais n'affectant pas les durées de scolarité, (b) des facteurs Z_{1i} pesant sur les décisions scolaires, mais n'affectant pas les revenus.

Ce type de problème est déjà ancien et cette étude n'est pas la première à chercher à le résoudre. Dans WILLIS et ROSEN [1979], la décision d'entrer à l'Université est supposée dépendre uniquement de l'origine sociale des étudiants (telle que la mesure le métier, l'éducation ou la religion du père) tandis que les revenus sont présumés varier avec les seules capacités individuelles (telles que les mesurent les résultats aux différents tests d'aptitudes).

Dans HAVEMAN *et alii* [1995] les variables expliquant le revenu et la scolarité sont choisies dans la même liste. Les auteurs proposent une série d'estimations correspondant chacune à des conditions d'exclusion particulières. Ils ne comparent pas la qualité de leurs différentes estimations. De façon générale, ni WILLIS et ROSEN [1979], ni HAVEMAN *et alii* [1995] ne proposent de tests de validité de leurs instruments.

Dans HARTOG *et alii* [1989], les mêmes variables expliquent le revenu et l'éducation. L'identification n'est possible qu'à travers les formes fonctionnelles, c'est-à-dire en postulant *a priori* que le revenu et l'éducation ne dépendent pas de la même façon des facteurs exogènes mesurés dans l'enquête.

Au-delà de leurs hypothèses d'exclusion, WILLIS et ROSEN [1979] supposent que les facteurs résiduels suivent une loi normale multivariée, ce qui est également l'hypothèse retenue par HAVEMAN *et alii* [1995] ou HARTOG *et alii*

5. Soulignons que nous ne retenons comme facteurs explicatifs du revenu aucune des variables de taille d'entreprise ou de secteur d'activité qu'il est d'usage de considérer comme des déterminants du revenu. Le problème est que le choix du secteur et de l'entreprise sont potentiellement tout aussi endogènes que la durée de la scolarité. Si nous introduisions des variables caractérisant le secteur et la taille de l'entreprise comme régresseurs, il faudrait en même temps modéliser la façon dont se déterminent le choix du secteur et de l'entreprise et trouver un ensemble de variables instrumentales permettant d'identifier leur effet sur le revenu. On n'aurait fait que compliquer le problème.

6. Dans leur analyse du choix de l'Université, MANSKI et WISE [1983] préfèrent ne pas supposer un tel degré de rationalité chez les étudiants. Chez eux, les choix des étudiants se fondent non pas sur l'espérance des revenus futurs, mais sur les résultats moyens des différentes Universités. De son côté, dans son analyse des choix de filières, FREEMAN [1971] suppose que les étudiants sont « *myopes* » et se fondent non pas sur une évaluation raisonnable de l'avenir, mais sur les revenus reçus par les générations anciennes ayant faits les mêmes choix qu'eux. Sur les mérites respectifs des différents modèles d'anticipation, voir MANSKI [1993].

[1989]. En donnant une forme paramétrique simple à la distribution des résidus, ils peuvent recourir à des estimateurs du maximum de vraisemblance assez faciles à mettre en œuvre. La robustesse des résultats obtenus sous ces hypothèses n'est pas établie.

De façon générale, il n'est pas aisé de trouver des instruments crédibles pour l'identification de l'effet des variables anticipées, ni très facile d'estimer cet effet sans faire d'hypothèses restrictives sur la structure de variance-covariance des résidus. De ce point de vue, notre approche a plusieurs caractéristiques intéressantes :

a) nous utilisons des données qui ont, en partie, été construites précisément en vue de traiter les problèmes posés par l'endogénéité des choix éducatifs. Nous disposons, en particulier, d'informations sur les accidents susceptibles de perturber l'enfance (décès et maladie de proches, par exemple) sans pour autant avoir de lien avec les aptitudes individuelles. Par ailleurs, nous avons des informations sur le contexte familial à la fois au début et à la fin des scolarités ;

b) pour le travail d'estimation, nous utilisons la méthode des moments généralisée, laquelle ne demande pas d'hypothèse *a priori* sur la structure de variance-covariance des résidus ;

c) nous proposons des tests de validité pour les conditions d'exclusion dont nous nous servons pour identifier les paramètres structurels.

À notre connaissance, c'est la première fois que l'effet de l'origine sociale sur le revenu est estimé à l'aide de données aussi riches et de techniques d'estimation adaptées aux problèmes d'endogénéité. Dans les sections suivantes, nous présentons plus en détail nos données et les stratégies d'instrumentation sur lesquelles nous allons nous appuyer.

4 Les données

Nos investigations empiriques vont s'appuyer sur l'enquête méthodologique « *Éducation et Carrières Professionnelles* » réalisée en 1997 par l'Insee. Un échantillon de 1099 hommes âgés de 20 à 45 ans a été interrogé, dans les régions du Poitou-Charentes, de l'Alsace et du Centre, en novembre 1997.

L'enquête fournit pour chaque individu les renseignements habituellement collectés dans les enquêtes menées par l'Insee : âge, sexe, nationalité, situation matrimoniale légale et de fait.

L'enquête comporte, par ailleurs, un descriptif du déroulement de la scolarité et notamment les renseignements habituellement collectés comme la durée effective des études, la date de fin des études et le diplôme le plus élevé obtenu.

L'enquête contient également un descriptif assez précis de l'origine sociale des individus. On connaît ainsi les diplômes les plus élevés du père et de la mère de chaque enquêté. Une originalité de l'enquête tient au fait que le métier du père est connu à deux moments différents de la scolarité de l'enfant : d'une part à l'entrée au cours préparatoire (CP) de l'enfant (quand il a

eu 6 ans, s'il n'a pas été scolarisé), d'autre part au moment où il a quitté l'école (quand il a eu 16 ans, s'il n'a pas été scolarisé).

Nous avons utilisé cette information sur le métier du père pour coder le prestige de sa profession en utilisant l'échelle construite récemment par CHAMBAZ, MAURIN et TORELLI [1998]. Cette échelle mesure les différences dans la façon dont sont évalués les métiers en France⁷. Comme le montrent CHAMBAZ, MAURIN et TORELLI [1998], les différences de prestige ainsi définies sont assez étroitement corrélées avec les différences de revenus entre professions. Quand on adopte l'échelle CMT, l'élasticité prestige/revenu d'un métier est voisine de 1. En introduisant le prestige de la profession du père dans l'analyse économétrique, on contrôle ainsi pour les différences de revenu d'activité entre les familles dont sont issues les personnes enquêtées.

Un volet du questionnaire passe en revue les événements importants qui ont pu se produire au cours de la scolarité : décès du père, décès de la mère, décès d'un frère ou d'une sœur, problèmes graves de santé du père de la mère ou de l'enfant lui-même, séparation ou divorce des parents, licenciement ou faillite d'un des deux parents, déménagement(s). C'est la troisième originalité de cette enquête. L'ensemble de ces variables est susceptible de contribuer à expliquer l'échec ou la réussite scolaire. Nous les supposons, en outre, sans lien avec les aptitudes individuelles expliquant le statut social.

L'enquête permet, par ailleurs, de reconstituer la situation professionnelle en début de carrière, la date d'obtention du premier contrat de travail (mois et année), le revenu d'activité dans ce premier emploi ainsi que le nombre d'heures hebdomadaires habituelles de travail dans cet emploi. Sur l'ensemble des 1 099 questionnaires, 80 sont tels qu'une ou plusieurs informations relatives au premier emploi sont absentes. Dans la suite du papier, l'échantillon utilisable comporte donc 1 019 questionnaires.

Sur les graphiques 1 et 2, nous avons reporté la distribution de la durée des études et du revenu d'activité au début de carrière. Le tableau 1 fournit quelques statistiques sur les variables indépendantes et instrumentales de nos modèles.

5 Spécifications empiriques et résultats

L'estimation du modèle demande de formuler un ensemble d'hypothèses pour l'identification de l'effet de la durée de la scolarité sur le revenu d'une part, et de l'impact du revenu anticipé sur la durée de la scolarité d'autre part.

S'agissant d'identifier l'effet de la scolarité sur les revenus, nous supposons que le niveau de formation des parents, le fait d'avoir un de ses parents enseignants et le statut socio-économique du père n'affectent les revenus

7. L'une des rares applications du modèle standard de BLAU et DUNCAN sur données françaises utilisait également une échelle de prestige social pour mesurer l'origine sociale (voir l'article de TACHIBANAKI [1977] et le commentaire de THÉLOT [1977]).

TABLEAU 1

*Quelques statistiques descriptives***1a. Variables continues**

Variable	Moyenne	Écart-type
<i>Variables dépendantes :</i>		
Nombre d'années effectives de scolarité	13.80	3.18
Log (revenu horaire au premier emploi en F97)	1.49	0.22
Revenu au premier emploi (en francs 97)	6175	2934
Horaire hebdomadaire au premier emploi	42.2	10.6
Revenu horaire au premier emploi (en francs 97)	34.98	17.15
<i>Variables indépendantes et instrumentales :</i>		
Niveau de scolarité du père (ensemble)	11.27	3.52
génération 1968-1977	11.20	10.02
génération 1953-1967	11.39	15.34
Niveau de formation de la mère (ensemble)	11.06	2.96
génération 1968-1977	10.88	9.41
génération 1953-1967	11.34	15.09
Log (SMIC horaire brut au début de premier contrat de travail en francs 97)	1.52	0.06
SMIC horaire brut au début du premier contrat de travail en francs 97	33.07	34.92
Prestige du métier du père (au CP)	-0.25	0.88
Prestige du métier du père (fin de scolarité)	-0.11	0.89

Note: Le prestige du métier du père est mesuré en utilisant l'échelle CHAMBAZ, MAURIN et TORELLI [1998]. Le niveau de formation des parents est évalué par le nombre d'années nécessaires en moyenne pour atteindre le diplôme le plus élevé qu'ils possèdent.

Source : Enquête Méthodologique sur l'Éducation et les Carrières Professionnelles, 1997, Insee.

1b. Variables qualitatives

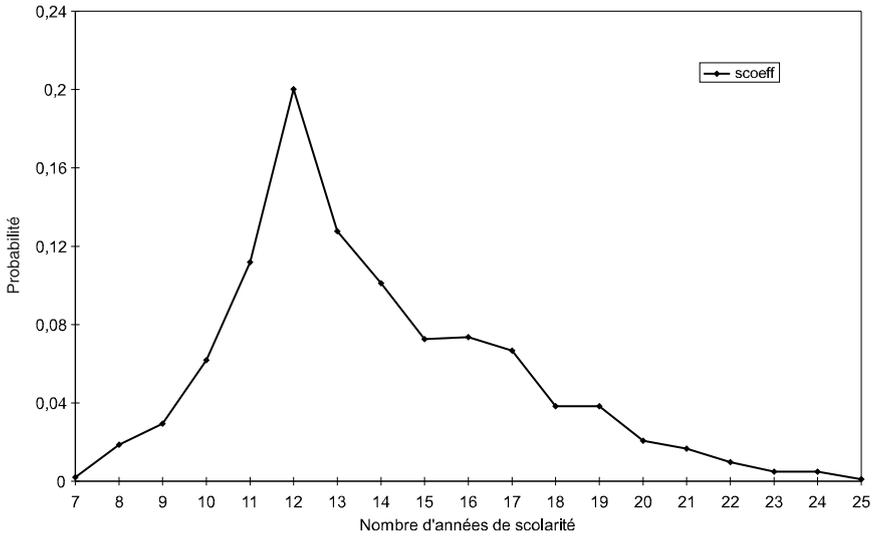
en %

Variable	Occurrence de l'événement
<i>Variables indépendantes et instrumentales :</i>	
Problèmes de santé de la mère	10.7
Faillite ou licenciement d'un parent	10.7
Décès d'un frère ou d'une soeur	3.2
Décès de la mère	1.5
Décès du père	5.6
Un des parents enseignant (au CP)	8.4
Un des parents enseignant (fin de scolarité)	9.0

Source : Enquête Méthodologique sur l'Éducation et les Carrières Professionnelles, 1997, Insee.

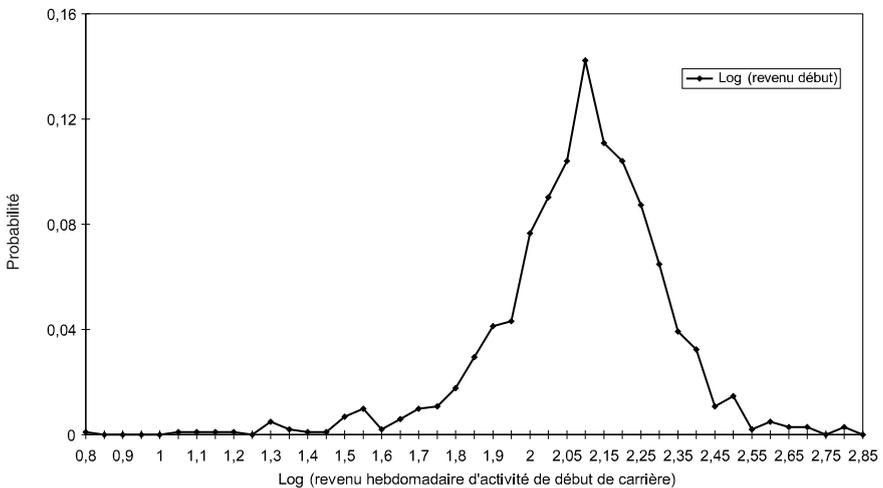
GRAPHIQUE 1

Fonction de densité des durées de scolarité



GRAPHIQUE 2

Fonction de densité des logarithmes de revenu d'activité



futurs que dans la mesure où ils déterminent la durée de scolarité. De même, toute une série d'événements susceptibles de perturber l'enfance seront supposés n'affecter les revenus futurs que dans la mesure où ils affectent les carrières scolaires : décès de la mère, du père, de l'un des frères et sœurs, problèmes de santé de la mère, faillite ou licenciement du père. Nous noterons Z_{1i} l'ensemble de ces variables instrumentales : elles permettent d'identifier l'effet de la durée des études sur le revenu.

S'agissant d'identifier l'effet du revenu anticipé sur les scolarités, notre hypothèse sera que le type de capital social et professionnel dont dispose le père (mesuré au moment où le répondant achève sa scolarité) n'affecte les scolarités que dans la mesure où il détermine les perspectives futures de revenu. Selon que le père est agriculteur, patron ou salarié, il dispose d'un capital social et professionnel très différent, plus ou moins facile à transmettre et plus ou moins facile à transformer en atout professionnel pour les enfants. Si son enfant est en échec scolaire, un patron a la possibilité de lui léguer son capital productif ou de l'intégrer dans son entreprise. Il peut ainsi lui procurer un avantage décisif sur les enfants de salariés ayant le même (faible) niveau de scolarité. De même, si son enfant est diplômé, un patron peut lui faire bénéficier de son réseau de relations professionnelles, *a priori* plus riche (d'un point de vue économique) que celui d'un salarié ou d'un agriculteur, particulièrement s'il s'agit d'un patron expérimenté.

Pour mettre ces hypothèses en forme, nous avons construit pour chaque répondant une variable croisant (a) le statut socioprofessionnel du père (*ie*, agriculteur, patron, salarié) à la fin de la scolarité du répondant et (b) la proximité de la retraite pour le père (*ie*, plus de 55 ans ou pas) mesurée au moment où le répondant a seize ans. Notre hypothèse de travail est que cette variable croisée détermine les perspectives de revenu à l'entrée sur le marché du travail sans avoir d'impact direct sur les scolarités. Nous noterons Z_{2i} ces variables instrumentales : elles permettent d'identifier l'effet du revenu anticipé sur les scolarités⁸.

Ces différentes hypothèses d'identification étant posées, la stratégie d'estimation est séquentielle. Dans un premier temps, nous régressons le revenu sur (i) la scolarité, (ii) le statut socio-économique du père à la fin de la scolarité obligatoire ainsi que sur (iii) l'ensemble des déterminants spécifiques du revenu (Z_{2i}). L'estimation est réalisée à l'aide de la méthode des moments généralisée (voir WHITE [1982]) en utilisant les variables de Z_{1i} comme instruments pour identifier l'effet de la scolarité sur le revenu⁹. Un test d'Hausman permet d'évaluer la qualité des instruments¹⁰. De cette première étape, nous obtenons une mesure du rendement de l'éducation (*ie*, $\hat{\alpha}$), une mesure de l'effet de la structure du capital professionnel paternel sur les performances économiques des enfants (*ie*, $\hat{\gamma}_2$) ainsi qu'une mesure de l'espérance du revenu (*ie*, $\hat{\alpha}S + \hat{\gamma}_2 Z_{2i}$) pour chaque répondant.

Dans une seconde étape, nous régressons la durée de scolarité sur (i) le statut socio-économique du père à la fin de la scolarité obligatoire, (ii) le revenu anticipé estimé en première étape et (iii) les différents déterminants

8. Il faut souligner que dans notre analyse, le niveau socio-économique du père au moment de la fin de la scolarité est un déterminant potentiel de la durée de la scolarité *et* du revenu anticipé. En d'autres termes, nous excluons qu'il puisse contribuer à l'identification de l'effet de l'une ou l'autre des deux variables endogènes. C'est bien la structure du capital paternel, et non pas son niveau qui permet d'identifier l'effet de l'espérance de revenu.

9. Une présentation très didactique de cette méthode se trouve dans LEE [1996].

10. Dès l'instant où il y a davantage d'instruments que d'endogènes et où ces instruments sont valides, l'estimateur des GMM est différent de celui des doubles moindres carrés (DMC), mais ils sont tous deux convergents, l'estimateur GMM étant simplement plus efficace (lorsque le modèle est juste identifié, les deux estimateurs sont identiques). Sous l'hypothèse nulle de validité des instruments, la différence standardisée entre les estimateurs DMC et GMM correspond donc à une statistique d'Hausman et suit un χ^2 à k degrés de liberté, où k est le nombre d'instruments (voir WHITE [1982]).

spécifiques de la scolarité (Z_{1i}). De nouveau l'estimation est réalisée à l'aide de la méthode des moments généralisée (GMM) en utilisant cette fois les variables de Z_{2i} comme instruments pour identifier l'effet du revenu anticipé. La qualité des instruments fait l'objet d'un deuxième test d'Hausman. Cette seconde étape nous permet de tester notre modèle de détermination des durées des scolarités (*ie*, de tester l'hypothèse $\theta > 0$). Dans la mesure où le modèle n'est pas rejeté, elle permet de réévaluer l'effet de l'origine sociale sur la difficulté à poursuivre des études longues (*ie*, $\hat{\gamma}_1$).

Les tableaux 2 et 3 reproduisent les résultats des deux étapes. De façon générale, les tests d'Hausman ne repoussent pas l'hypothèse de validité des instruments utilisés dans cette analyse, en première comme en seconde étape.

5.1 Résultats : l'amplitude des biais de simultanéité

Considérons, tout d'abord, l'équation expliquant le revenu en fonction de la durée de scolarité et du statut dans l'emploi du père (mesuré, rappelons-le, lorsque le répondant a 16 ans). Le rendement estimé de l'éducation est d'environ 4 % : chaque année de scolarité supplémentaire est associée à un accroissement d'environ 4 % du revenu à l'entrée sur le marché du travail¹¹. Les tests de sur-identification ne rejettent pas l'hypothèse de validité des instruments utilisés pour l'identification de ce rendement. Lorsqu'on n'instrumente pas la durée de scolarité, on obtient un rendement de la scolarité sensiblement plus faible, de l'ordre de 2.5 %. Cette différence entre effet brut (non instrumenté) et effet vrai (instrumenté) suggère l'existence d'erreur dans la mesure de la durée de la scolarité (voir GRILICHES [1979]).

Quand on les considère de façon jointe, les variables de statut dans l'emploi du père (Z_{2i}) contribuent à expliquer le revenu à l'entrée sur le marché du travail. À durée de scolarité égale, les étudiants ayant les revenus les plus élevés au moment où ils entrent sur le marché du travail sont ceux dont le père était un patron de plus de 55 ans lorsqu'ils avaient 16 ans. À diplôme donné, ils reçoivent environ 20 % de plus que les autres à l'entrée sur le marché du travail, ce qui est tout de même équivalent au supplément de revenus que procurent 5 années d'études supplémentaires. Le capital social et professionnel dont peuvent espérer bénéficier les enfants de patrons expérimentés semble bel et bien plus important que celui des autres enfants.

Considérons maintenant l'équation expliquant la durée des scolarités en fonction de l'origine sociale et des revenus anticipés. Il ressort en premier lieu que la durée des études est corrélée positivement avec le revenu anticipé ($\theta > 0$). Les tests de sur-identification ne rejettent pas l'hypothèse de cohérence des instruments utilisés pour identifier cet effet du revenu anticipé. Le cadre théorique utilisé dans cet article ne semble donc pas rejeté par les données. La valeur estimée pour le coefficient θ s'interprète de la façon suivante : soient deux étudiants tels que – à niveau de formation donné – le

11. Soulignons qu'il s'agit d'un rendement à l'entrée sur le marché du travail et c'est peut-être la raison pour laquelle il est plutôt plus faible que les rendements estimés habituellement dans la littérature, rendements qui correspondent à des rendements moyens sur l'ensemble de la carrière (GOUX et MAURIN [1994]). Soulignons aussi qu'une spécification plus réaliste demanderait de prendre en compte des non-linéarités dans l'effet de la scolarité sur les salaires.

TABLEAU 2

L'effet de l'origine sociale sur la durée de scolarité : un modèle structurel

	Modèle	Estimations VI	
	MCO (Réduit)	DMC	MMG
Équation 1 (Variable dépendante : Durée de scolarité)			
Constante	11.1 (0.4)	1.05 (3.6)	-4.1 (6.8)
<i>Marché du travail</i>			
E(Ln(revenu au premier emploi))		5.0 (3.6)	7.5 (3.4)
<i>Origine sociale</i>			
Durée scolarité père	0.11 (0.04)	0.09 (0.03)	0.08 (0.03)
Durée scolarité mère	0.11 (0.03)	0.08 (0.03)	0.07 (0.03)
Prestige père (au CP)	0.51 (0.18)	0.42 (0.15)	0.40 (0.15)
Prestige père (en fin d'études)	0.39 (0.18)	0.36 (0.14)	0.34 (0.14)
Un des parents enseignant	0.80 (0.36)	0.61 (0.30)	0.52 (0.30)
<i>Événements de l'enfance</i>			
Problèmes de santé de la mère	0.58 (0.30)	0.46 (0.29)	0.43 (0.29)
Décès d'un frère ou d'une sœur	-1.17 (0.50)	-0.95 (0.34)	-0.78 (0.34)
Décès de la mère	-0.57 (0.76)	-0.44 (0.71)	-0.45 (0.70)
Décès du père	0.89 (0.39)	0.79 (0.35)	0.56 (0.34)
Licenciement père	0.52 (0.29)	0.42 (0.24)	0.30 (0.23)
R ²	0,206	0,474	0,588
Test d'Hausman			20.2 Df = 13 (P = 0.09)
	Modèle	Estimations VI	
	MCO	DMC	MMG
Équation 2 (Variable dépendante : Revenu à l'entrée)			
Constante	1.76 (0.03)	1.59 (0.09)	1.62 (0.09)
Nombre d'années effectives de scolarité	0.028 (0.002)	0.040 (0.006)	0.037 (0.006)
<i>Âge × CS Père (fin scolarité)</i> (réf : autre salarié < 55 ans)			
<i>Patron < 55 ans</i>	-0.05 (0.02)	-0.05 (0.02)	-0.05 (0.02)
<i>Patron > 55 ans</i>	0.20 (0.08)	0.20 (0.10)	0.20 (0.09)

TABLEAU 2 (suite et fin)

L'effet de l'origine sociale sur la durée de scolarité : un modèle structurel

	Modèle	Estimations VI	
	MCO	DMC	MMG
Équation 2 (Variable dépendante : Revenu à l'entrée) (suite)			
<i>Agriculteur < 55 ans</i>	0.03 (0.03)	0.02 (0.03)	0.02 (0.03)
<i>Agriculteur > 55 ans</i>	-0.08 (0.14)	-0.09 (0.09)	-0.09 (0.09)
<i>Ouvrier < 55 ans</i>	0.03 (0.05)	0.04 (0.05)	0.04 (0.05)
<i>Ouvrier et autre salarié > 55 ans</i>	-0.04 (0.03)	-0.04 (0.04)	-0.03 (0.03)
Prestige père (fin scolarité)	0.010 (0.008)	0.00 (0.01)	0.00 (0.01)
Père décédé (fin de scolarité)	0.00 (0.03)	-0.01 (0.03)	-0.01 (0.03)
Nombre d'observations	1019	1019	1019
R ²	0,183	0,155	0,164
Test d'Hausman			7.9 Df = 10 (P = 0.63)

Note : Les écarts-types sont entre parenthèses.

Source : Enquête Méthodologique sur l'Éducation et les Carrières Professionnelles, 1997, Insee.

TABLEAU 3

L'effet de l'origine sociale sur la durée de scolarité pour les générations nées avant 1968 et pour les générations nées après 1968

	Estimateur	Estimateurs VI	
	MCO (réduit)	DMC	MMG
Équation 1 (Variable dépendante : Durée de la scolarité)			
Constante	11.2 (0.45)	4.2 (7.7)	0.92 (7.0)
<i>E(Ln(revenu))</i>		3.4 (3.7)	5.1 (3.4)
<i>Un des Parents Enseignant</i> génération 1968-1977	0.82 (0.43)	0.68 (0.38)	0.64 (0.37)
génération 1953-1967	0.46 (0.70)	0.39 (0.58)	0.29 (0.57)
<i>Durée Scolarité Mère</i> génération 1968-1977	0.13 (0.04)	0.12 (0.04)	0.11 (0.03)
génération 1953-1967	0.10 (0.05)	0.09 (0.05)	0.08 (0.04)
<i>Durée Scolarité Père</i> génération 1968-1977	0.12 (0.04)	0.11 (0.04)	0.10 (0.04)
génération 1953-1967	0.11 (0.04)	0.09 (0.04)	0.08 (0.04)

TABLEAU 3 (suite et fin)

L'effet de l'origine sociale sur la durée de scolarité pour les générations nées avant 1968 et pour les générations nées après 1968

Équation 1 (Variable dépendante : Durée de la scolarité) (suite)				
<i>Décès Mère</i>	génération 1968-1977	-0.03 (0.98)	-0.02 (1.11)	-0.21 (1.10)
	génération 1953-1967	-0.33 (1.2)	-0.29 (0.73)	-0.42 (0.72)
<i>Décès Père</i>	génération 1968-1977	0.45 (0.51)	0.38 (0.49)	0.38 (0.48)
	génération 1953-1967	1.61 (0.64)	1.40 (0.63)	1.23 (0.63)
<i>Prestige Père</i>	génération 1968-1977	0.73 (0.14)	0.63 (0.17)	0.59 (0.16)
	génération 1953-1967	1.08 (0.18)	0.91 (0.26)	0.95 (0.25)
R²		0,185	0,383	0,467
Test d'Hausman : H (Df)		– 11.1 (14)		
P-value		0.68		
Équation 2 (Variable dépendante : Ln (Revenu entrée marché travail))				
Constante		1.8 (0.3)	1.6 (0.1)	1.6 (0.1)
<i>Durée Scolarité</i>	génération 1968-1977	0.027 (0.002)	0.037 (0.009)	0.038 (0.008)
	génération 1953-1967	0.027 (0.002)	0.037 (0.009)	0.038 (0.009)
<i>Père Patron (fin de scolarité)</i>	génération 1968-1977	-0.06 (0.02)	-0.06 (0.02)	-0.06 (0.02)
	génération 1953-1967	-0.05 (0.03)	-0.04 (0.04)	-0.03 (0.03)
<i>Père Patron >55 ans (fin de scolarité)</i>	génération 1968-1977	0.14 (0.07)	0.14 (0.05)	0.15 (0.05)
	génération 1953-1967	0.13 (0.14)	0.12 (0.28)	0.11 (0.28)
<i>Père Agriculteur (fin de scolarité)</i>	génération 1968-1977	0.03 (0.04)	0.02 (0.03)	0.02 (0.03)
	génération 1953-1967	0.02 (0.05)	0.02 (0.06)	0.01 (0.05)
<i>Père Ouvrier >55 ans (fin de scolarité)</i>	génération 1968-1977	0.01 (0.05)	0.02 (0.05)	0.00 (0.05)
	génération 1953-1967	-0.02 (0.06)	0.00 (0.05)	0.00 (0.05)
<i>Prestige Père</i>	génération 1968-1977	0.02 (0.01)	0.01 (0.01)	0.00 (0.01)
	génération 1953-1967	0.03 (0.01)	0.01 (0.02)	0.01 (0.02)
Nombre d'observations		1019	1019	1019
R²		0,187	0,166	0,162
Test d'Hausman : H (Df)		– 7.7 (13)		
P-value		0.86		

Note : Les écarts-types sont entre parenthèses.

Source : Enquête Méthodologique sur l'Éducation et les Carrières Professionnelles, 1997, Insee.

premier anticipe des revenus d'activité 10 % plus élevés que le second (par exemple s'il a une origine sociale beaucoup plus élevée), un coefficient égal à 8 signifie que le premier restera en moyenne $8 \times 0,1 = 0,8$ ans de plus dans le système éducatif¹².

Au-delà du revenu anticipé, les différentes mesures de l'origine sociale ont toutes un impact significatif sur les durées de scolarité. Notre analyse confirme que des parents diplômés et/ou ayant un statut socio-économique élevé sont des atouts importants pour réussir à l'école. Point important, lorsque nous omettons la variable de revenu anticipé (modèle MCO, tableau 2), l'impact direct de l'origine sociale apparaît beaucoup plus considérable. L'effet de la scolarité des parents est, par exemple, 50 % plus élevé lorsqu'on néglige le revenu anticipé que lorsqu'on en contrôle l'effet.

Conformément aux arguments développés dans les sections précédentes (et en annexe), l'analyse en forme réduite pratiquée habituellement (*ie*, négligeant l'effet du revenu anticipé) conduit donc à surestimer l'effet direct de l'origine sociale sur les destinées scolaires. D'un point de vue économétrique, ce type d'analyse revient à omettre une variable explicative endogène (*ie*, les perspectives de revenus liés à l'investissement éducatif).

5.2 Biais de simultanéité et évolution de la mobilité sociale

Dans le tableau 3, nous avons reporté une analyse similaire à celle du tableau 2, à ceci près que nous introduisons dans l'équation de revenu la possibilité d'une variation du rendement de l'éducation entre les générations nées avant 1968 et celles nées après 1968. De même, nous introduisons dans l'équation de scolarité la possibilité d'une variation de l'effet de l'origine sociale avant et après 1968.

Deux résultats méritent d'être soulignés :

1) le rendement estimé de l'éducation est quasi identique pour les personnes nées avant et après 1968 ;

2) les différents effets de l'origine sociale sur les durées de scolarité ne sont pas significativement différents avant et après 1968. L'effet du diplôme des parents est plutôt plus fort et l'effet du prestige social du père plutôt plus faible après 1968 qu'avant 1968, mais aucune de ces variations n'est significative aux seuils usuels.

Lorsque nous négligeons l'effet du revenu anticipé, nous obtenons de nouveau un impact apparent de l'origine sociale sur la scolarité beaucoup plus élevé que lorsque nous contrôlons l'effet du revenu anticipé (modèle MCO, tableau 3). Nous obtenons, toutefois, la même surestimation pour les effets correspondant à avant et après 1968. Dans le cas de la France contemporaine, l'omission des anticipations ne semble donc pas impliquer d'erreur de diagnostic sur l'évolution des effets de l'origine sociale.

Ce résultat est cohérent avec l'analyse des biais de simultanéité développée dans la deuxième section et en annexe : le rendement de l'éducation n'ayant

12. En terme plus structurel, un coefficient égal à 8 signifie que le coût marginal de la scolarité croît de 12 % chaque année (*ie*, 1/8 %).

pas changé de façon significative, les biais de simultanéité n'évoluent pas au fil des générations et les estimateurs en forme réduite évoluent de la même façon que les estimateurs structurels. Il n'en serait pas de même dans un pays ou à une époque où les rendements de l'éducation évolueraient rapidement au cours du temps.

6 Une nouvelle voie de recherche sur la mobilité sociale

Avant de conclure, il est sans doute utile de rappeler la portée et les limites de cette étude, ainsi que d'indiquer brièvement le programme de recherche qu'elle nous semble indiquer pour l'avenir.

S'agissant de la portée de l'article, elle est essentiellement de mettre en question les mesures statistiques habituelles (a) du rôle propre de l'école dans les mécanismes de reproduction sociale et (b) de l'évolution au cours du temps de l'inégalité des chances, tant à l'école que sur le marché du travail. Dès l'instant où les étudiants prennent en compte l'impact de leurs efforts actuels sur leur situation sociale future, les corrélations statistiques entre origine sociale et destinée scolaire ne sont plus une mesure fiable de la part de la discrimination sociale opérée dans le système éducatif¹³. De même, l'évolution des corrélations entre origine et destinées scolaire ou sociale cesse d'être une indication nécessairement fidèle de la façon dont évoluent les déterminismes sociaux.

Il faut être clair : à chaque date, un coefficient de corrélation entre origine et destinée sociale particulièrement élevé (ou une table de mobilité sociale particulièrement « *diagonale* ») reste bien une indication simple et probante d'un manque de fluidité sociale¹⁴. Mais ces indicateurs statistiques peuvent changer au cours du temps pour des raisons n'ayant rien à voir avec un accroissement ou une diminution des déterminismes sociaux. Il suffit – par exemple – qu'une innovation technologique rende les diplômés plus payants sur le marché du travail et plus attractifs pour tout le monde pour que le lien statistique apparent entre origine et destinée sociale se modifie.

S'agissant des limites de notre exercice, elles sont évidemment nombreuses et représentent autant de pistes pour la recherche sur la mobilité sociale à venir¹⁵. De façon générale, toutes les hypothèses théoriques sur lesquelles se

13. Nos données suggèrent qu'elles en représentent en fait une surestimation.

14. En d'autres termes, ce n'est pas parce qu'une partie des inégalités repérées dans les tables de mobilité ou dans les analyses de corrélations dérivent de décisions individuelles « rationnelles » qu'elles ne correspondent pas à des processus de discrimination sociale, à des défauts de fluidité : les enfants d'origine modeste choisissent peut-être de n'avoir aucune chance d'être très riches plus tard, mais c'est parce que la société ne leur laisse pas de choix plus rationnel.

15. Dans cet article, nous nous sommes concentrés sur le problème des anticipations et sur la façon dont leur prise en compte pouvait affecter l'interprétation des mesures de la mobilité sociale. Pour une approche plus générale, on peut consulter PIKETTY [2000].

fondent notre approche restent à mettre en forme dans un cadre plus général et à valider à l'aide de données nouvelles :

– l'hypothèse selon laquelle les étudiants forment des anticipations rationnelles, c'est-à-dire l'hypothèse selon laquelle ils ont une idée à peu près juste de l'impact moyen futur de leurs efforts scolaires actuels. En réalité, on peut conjecturer qu'ils connaissent imparfaitement le système scolaire d'une part et leurs propres aptitudes d'autre part¹⁶ ;

– l'hypothèse selon laquelle ils ne sont sensibles à rien d'autre (*ie*, rien d'autre déterminé socialement) que le revenu moyen futur. Sans doute sont-ils également sensibles au risque de chômage, à la sécurité de l'emploi et sans doute cette sensibilité varie-t-elle avec l'origine sociale. Peut-être certains aiment-ils aussi les études pour elles-mêmes et – si c'est le cas – il est très probable que ce désintéressement n'est pas réparti au hasard entre les différents milieux sociaux ;

– l'hypothèse d'apparence technique et anodine selon laquelle les coûts varient de façon log-linéaire avec la durée des études. Elle est commode pour le propos de l'article, mais sans réel fondement théorique ou empirique.

Il faut également rappeler que notre modèle repose sur une analyse d'équilibre partiel. En particulier, le rendement des diplômés sur le marché du travail est supposé exogène, indépendant des comportements individuels. En fait, il est probable que ce paramètre est au moins en partie déterminé par l'*agrégation* des décisions des étudiants. Dans le cas de la France, l'impact de l'éducation sur les revenus aurait sans doute augmenté (alors qu'il est resté stable) au cours des décennies récentes si la demande d'éducation ne s'était pas considérablement renforcée, les études allongées et les diplômés diffusés massivement dans l'ensemble de la société. Si l'objectif de la recherche est de mieux prévoir l'effet des politiques éducatives, il est clair que les analyses futures devront prendre en compte non seulement les processus de décisions individuelles, mais également la façon dont l'agrégation des décisions individuelles contribue à modifier les paramètres clefs des processus de décision.

Pour ne pas rester un simple exercice d'école, l'enrichissement du réalisme des modèles théoriques devra aller de pair avec l'amélioration des observations qualitatives et quantitatives. La récente enquête de l'Insee sur l'éducation dirigée par François HÉRAN est un pas dans cette direction (HÉRAN *et al.* [1994]). Selon nous un effort particulier devrait être fait pour recueillir des mesures directes des anticipations, ainsi que de leurs effets. Le choix entre un modèle à anticipation rationnelle et un modèle où familles et étudiants sont « *myopes* » devrait dans l'avenir pouvoir se faire de façon positive, à partir de données statistiques. De même, on sait encore trop peu de choses sur les coefficients d'actualisation, c'est-à-dire sur l'importance accordée à l'avenir (quelle que soit la façon dont il est anticipé) dans les décisions courantes. Il serait évidemment très utile de disposer d'informations sur la façon dont les paramètres d'actualisation se distribuent d'un milieu social à l'autre.

16. Cette méconnaissance se déclinant, en outre, sans doute en fonction de variables sociales, comme le genre (voir BAUDELLOT et ESTABLET [1992]).

7 Conclusion

Dans cet article, nous explicitons une théorie très simple du comportement des étudiants en fonction du rendement de leurs études. Cette approche permet de fonder un modèle empirique de détermination simultanée du statut social et de la scolarité, généralisant le modèle de BLAU et DUNCAN [1967]. Nous estimons les paramètres structurels de ce modèle à l'aide de la méthode des moments généralisée et en utilisant une nouvelle base de données longitudinales française. Nos résultats suggèrent que l'approche statistique en forme réduite, aussi sophistiquée soit-elle, donne une vision déformée des mécanismes de transmission des statuts scolaires et sociaux d'une génération à l'autre. Dans le cas de la France contemporaine, l'approche naïve en forme réduite surestime fortement le rôle propre de l'école dans le déterminisme social. Elle risque, en outre, de conduire à une appréciation biaisée de l'évolution des rigidités sociales.

Notre modèle n'est pas rejeté par les données, cela ne veut pas dire qu'il n'est pas possible de l'enrichir. Dans sa version actuelle, il retient une version très simplifiée des parcours scolaires, négligeant leur nature intrinsèquement séquentielle. Comme le suggèrent plusieurs études récentes, les problèmes d'identification structurelle s'accroissent toutefois très vite au fur et à mesure que l'on approfondit dans cette direction là (voir MAGNAC et THESMAR [1999], CAMERON et HECKMAN [1998]).

En se focalisant sur les revenus en début de carrière, notre modèle retient également une vision simplifiée de tout ce qui peut déterminer les étudiants à poursuivre leurs études. Une extension tout à fait envisageable du modèle serait de prendre en compte non seulement les perspectives de revenus professionnels, mais également les risques de chômage. ■

• Références bibliographiques

- BAUDELLOT C., ESTABLET R. (1992). – *Allez les Filles !*, Seuil, Paris.
- BLAU P., DUNCAN O.D. (1967). – *The American Occupational Structure*, Wiley & Sons: New York.
- BOUDON R. (1981). – *The Logic of social Action: An Introduction to Sociological Analysis*, Translated by D. Silverman with the Assistance of G. Silverman, Boston: Routledge & Kegan Paul.
- BOUDON R. (1974). – *Education, Opportunity and Social Inequality*, New York: Wiley.
- CAMERON S.V, HECKMAN J.J. (1998). – « Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males », *Journal of Political Economy*, 106, pp. 262-334.
- CARD D. (1998). – « The Causal Effect of Education on Earnings », *Document de travail n° 2*, University of Berkeley.
- CHAMBAZ C., MAURIN É., TORELLI C. (1998). – « L'évaluation sociale des professions en France », *Revue Française de Sociologie*, 39(1).
- ERIKSON R. GOLDTHORPE J. (1992). – *The Constant Flux. A Study of Class Mobility in Industrial Societies*, Oxford Clarendon Press.

- FREEMAN R.B. (1971). – *The Market for College-trained Manpower*, Cambridge Mass: Harvard University Press.
- GANZEBOOM H., LUIJKX R., TREIMAN D. (1989). – « Intergenerational Class Mobility in Comparative Perspective », *Research in Social Stratification and Mobility*, Vol. 8.
- GAREN J. (1984). – « The Returns to Schooling: A Selectivity Biased Approach to Continuous Choice Variables », *Econometrica* 52(5), pp. 1199-1218.
- GOUX D., MAURIN É. (1994). – « Éducation, expérience et salaire : tendances récentes et évolution de long terme », *Économie et Prévision*, 116(5).
- GOUX D., MAURIN É. (1997). – « Meritocracy and Social Heredity in France: Some Aspects and Trends », *European Sociological Review*, 13(2), pp. 159-177.
- GRILICHES Z. (1977). – « Estimating the Returns to Schooling: Some Econometric Problems », *Econometrica*, 45, pp. 1-22.
- GRILICHES Z. (1979). – « Sibling Models and Data in Economics: Beginnings of a Survey », *Journal of Political Economy*, 87, pp. S37-S65.
- HARTOG J., PFANN G., RIDDER G. (1989). – « (Non-)Graduation and the Earnings Function », *European Economic Review*, 33, pp. 1373-1395.
- HAVEMAN R., WOLFE B. (1995). – « The Determinants of Childrens Attainments: A Review of Methods and Findings », *Journal of Economic Literature*, 33, pp. 1829-1878.
- HAVEMAN R., WOLFE B., WILSON K. (1995). – « Childhood Poverty and Adolescent Schooling and Fertility Outcomes: Reduced-Form and Structural Estimates », in *Consequences of Growing up Poor*, Greg J. Duncan and Jeanne Brooks-Gunn Editors. New York: Russell Sage Foundation, pp. 419-460.
- HÉRAN F., GISSOT C., MANON N. (1994). – « Les efforts éducatifs des familles », *Collection INSEE-Résultats, Série Consommation Mode de Vie*, INSEE, Paris.
- LEE M.J. (1996). – *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*, Springer, New York.
- MAGNAC T., THESMAR D. (1999). – « Identifying Dynamic Discrete Choice Models. An Application to School-leaving in France », *Working Paper*, CREST.
- MANSKI C.F. (1993). – « Adolescent Econometricians: How Do Youth Infer the Returns to Schooling? », in C. Clotfelter and M. Rothschild eds., *Studies of Supply and Demand in Higher Education*, Chicago, University of Chicago Press.
- MANSKI C.F., WISE D.A. (1983). – *College Choice in America*, Cambridge Mass, Harvard University Press.
- PIKETTY T. (2000). – « Theories of Persistent Inequality and Intergenerational Mobility », in *Handbook of Income Distribution*, Chapter 6, A. Atkinson and F. Bourguignon Editors.
- SEWELL W.H., HAUSER R.M. (1975). – *Education, Occupation and Earnings: Achievement in the Early Career*, New York, Academic Press.
- SEWELL W.H., HAUSER R.M., WOLF W.C. (1980). – « Sex, Schooling and Occupational Status », *American Journal of Sociology*, 86, pp. 551-583.
- TACHIBANAKI T. (1977). – « Mobilité sociale et formation en France », *Annales de l'INSEE*, 28, pp. 129-142.
- THÉLOT C. (1977). – « Mobilité sociale et formation en France : commentaire sur l'article de Tachibanaki », *Annales de l'INSEE*, 28, pp. 145-159.
- WHITE H. (1982). – « Instrumental Variables Regression with Independent Observations », *Econometrica*, 50, pp. 483-499.
- WILLIS R.J., ROSEN S. (1979). – « Education and Self-Selection », *Journal of Political Economy*, 87, pp. S7-S36.
- WONG R.S.-K. (1992). – « Vertical and Nonvertical Effects in Class Mobility: Cross-National Variations », *American Sociological Review*, 57(3), pp. 396-410.
- XIE Y. (1992). – « The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables », *American Sociological Review*, 57(3), pp. 380-395.

ANNEXE

Le problème des biais de simultanéité

Le modèle développé dans cet article permet de se donner une idée des problèmes d'interprétation posés par les modèles en forme réduite utilisés habituellement dans la littérature. Notons Z_{12} les variables expliquant à la fois le revenu et le niveau de formation du répondant, Z_{11} les variables expliquant le niveau de formation du répondant seulement et enfin Z_{22} les variables n'ayant d'effet propre que sur le revenu. Le modèle structurel considéré se réécrit :

$$(A1) \quad \ln R = \alpha S + \gamma_{12} Z_{12} + \gamma_{22} Z_{22} + u$$

$$(A2) \quad S = \theta \ln R + \delta_{12} Z_{12} + \delta_{11} Z_{11} + v,$$

En remplaçant $\ln R$ par sa valeur, l'équation de choix éducatif se réécrit en forme réduite selon :

$$(A3) \quad S = \left(\frac{\theta \gamma_{12} + \delta_{12}}{1 - \theta \alpha} \right) Z_{12} + \frac{\theta \gamma_{22}}{1 - \theta \alpha} Z_{22} + \frac{\delta_{11}}{1 - \theta \alpha} Z_{11} + v_1.$$

$$\text{où } v_1 = \frac{v + \theta u}{1 - \theta \alpha}.$$

L'équation (A3) correspond à l'équation de détermination de l'éducation dans un modèle en forme réduite traditionnel, où les seuls facteurs explicatifs considérés sont les variables d'origine sociale. Dès l'instant où le revenu espéré joue sur les choix éducatifs ($\theta > 0$), on constate que les coefficients obtenus à partir de cette forme réduite fournissent une estimation biaisée des coefficients structurels, les seuls ayant un intérêt pour la politique éducative et sociale. Considérons, par exemple, une variable susceptible d'affecter à la fois la scolarité des enfants et leur statut social ultérieur. En suivant nos notations, cette variable appartient à Z_{12} . Dans le modèle en forme réduite, l'impact de cette variable sur la scolarité est estimé par $\frac{\theta \gamma_{12} + \delta_{12}}{1 - \theta \alpha}$ alors que son impact structurel est δ_{12} . En supposant ($\theta \alpha < 1$, $\delta_{12} > 0$ et $\gamma_{12} > 0$), cette estimation en forme réduite de δ_{12} a le bon signe, mais souffre d'un biais $B_{\delta_{12}}$ positif potentiellement très important, soit :

$$(A4) \quad B_{\delta_{12}} = \frac{\theta \gamma_{12} + \delta_{12}}{1 - \theta \alpha} - \delta_{12} = \frac{\theta(\gamma_{12} + \alpha \delta_{12})}{1 - \theta \alpha}$$

Ce biais est d'autant plus grand que θ , α et γ_{12} sont forts. Il peut également être très important alors même que γ_{12} est faible ou nul.

On peut tenir le même type de raisonnement pour les variables affectant uniquement la scolarité. L'impact structurel de ce type de variable est mesuré

par δ_{11} . En raisonnant sur la forme réduite habituelle (A3), on obtient non pas δ_{11} , mais $\frac{\delta_{11}}{1 - \theta\alpha}$. En supposant $\theta\alpha < 1$, cet effet en forme réduite est positif, mais surestime l'effet. Raisonner sur un modèle en forme réduite conduit ainsi typiquement à surestimer l'effet des contextes culturels. De façon générale, même si l'école fonctionnait de façon parfaitement neutre, sans générer la moindre discrimination sociale (*ie*, même si δ_{11} et δ_{12} étaient nuls), il resterait possible de continuer à observer (en forme réduite) une corrélation positive – éventuellement très élevée – entre origine sociale et niveau de formation.

Les paramètres estimés en forme réduite ont le même signe que ceux estimés en forme structurelle. L'identification d'effets significatifs en forme réduite est bien équivalente à l'existence d'effets structurels significatifs, même si on ne sait pas lesquels. En revanche, ce n'est pas parce que les coefficients de corrélation obtenus en forme réduite changent au cours du temps, que les coefficients structurels changent eux aussi : tout chose égale par ailleurs, une augmentation des rendements de l'éducation (*ie*, hausse de α) ou une baisse de ses coûts (*ie*, hausse de θ) entraînent mécaniquement une augmentation des corrélations observées entre origine et destinée sociale, sans que rien n'ait changé du point de vue des processus de discrimination sociale. En d'autres termes, les tables de mobilité et les mesures statistiques traditionnelles sont bien, à chaque date, une mesure agrégée, simple et probante des défauts de fluidité d'une société donnée. En revanche, la façon dont elles évoluent au cours du temps ne s'interprète pas simplement en terme de fluidité sociale.

Erreurs de mesure et aptitudes non mesurées

S'agissant des effets de l'éducation sur le revenu, une équation aussi simple que l'équation (A1) pose des problèmes économétriques qui ne sont pas moins redoutables – même s'ils sont mieux connus et étudiés – que ceux exposés ci-dessus à propos des effets de l'origine sociale sur l'éducation.

Comme le prouve la relation (A3), la variable éducation dépend des variables d'origine sociale, mais dépend également du résidu $v_1 = \frac{v + \theta u}{1 - \theta\alpha}$. Dès l'instant où $\theta \neq 0$, ce résidu est corrélé avec u . En d'autres termes, dès l'instant où les anticipations pèsent significativement sur les décisions individuelles, la variable éducation n'est plus orthogonale à u et ne peut plus être considérée comme exogène. Dans ce cas, la régression de la variable $\ln R$ sur S , Z_{12} et Z_{22} expose à des biais d'endogénéité. L'application de la méthode des moindres carrés ordinaires a toute chance de fournir une estimation biaisée du paramètre α , et, par conséquent, de l'ensemble des paramètres structurels. Dans la mesure où θ est positif, l'endogénéité de S entraîne *a priori* une surestimation des rendements de l'éducation.

Indépendamment des problèmes d'endogénéité, il faut toutefois également envisager que la variable éducation soit mal mesurée. Les erreurs de mesure créent un problème symétrique à celui occasionné par l'endogénéité. De fait, les erreurs sur la mesure de l'éducation reviennent à oublier une variable explicative corrélée négativement avec la variable à expliquer. Une telle omis-

sion implique un biais vers zéro des estimateurs standards du rendement salarial de l'école (voir GRILICHES [1977, 1979]).

Par ailleurs, même dans le cas où il n'y aurait pas d'erreurs de mesure et où $\theta = 0$, rien ne garantirait l'orthogonalité des deux résidus u et v . En présence de facteurs non mesurés déterminant simultanément les performances scolaires et salariales (*ie*, u et v ne sont pas orthogonaux), l'analyse statistique naïve des corrélations entre éducation et revenus d'activité n'a de nouveau rien d'une relation causale. Comme le rappelle CARD [1998], en présence d'erreur de mesure, on tend à sous-estimer le rôle de l'école, alors que les aptitudes non mesurées déterminant simultanément les salaires et la scolarité exposent à surestimer l'impact de l'école.