

# **Origine Sociale et réussite scolaire : un modèle en Chaîne de Markov estimé sur données individuelles belges**

Olivier Donni  
Université de Mons-Hainaut et CREPP

Bernard Lejeune  
Université de Liège, CORE et ERUDITE

---

Cette étude a été réalisée dans le cadre d'un programme de recherche financé par le Fonds National de la Recherche Scientifique Belge. La base de données a été mise à la disposition des auteurs par le Point d'Appui Panel de l'Université de Liège. Les auteurs tiennent à remercier Sergio Perelman, Pierre Pestieau ainsi que deux rapporteurs anonymes pour leurs conseils avisés. Ce travail bénéficia également de nombreux commentaires des participants aux séminaires inter-universitaires en économie de l'éducation.

## Résumé

Cet article étudie les relations entre l'origine socio-économique des élèves belges et leur parcours scolaire primaire et secondaire. Il repose sur l'hypothèse que le parcours scolaire des élèves peut être décrit par un processus de Markov. Sur la base d'un échantillon transversal d'un millier d'enfants, en fonction de variables socio-économiques et pour chacune des années de scolarité obligatoire, les probabilités individuelles de poursuivre un parcours scolaire "normal", c.-à-d. sans redoublement ni relégation dans une filière de second choix, sont estimées. Nos résultats montrent que derrière les substantiels taux agrégés de redoublement qui caractérisent le système scolaire belge se cachent en fait de profondes inégalités sociales.

## Abstract

This paper examines the relationship between socio-economic background and primary and secondary school achievements of Belgian students. It relies on the assumption that the schooling path of students may be described through a Markovian process. Based on a cross-section sample of about 1000 children, the individual probabilities of following a "normal" schooling path, i.e., without repeating or demeaning reorientation, are estimated for each year of compulsory schooling and as a function of socio-economic background variables. Our results show that, behind the very high aggregated rate of repeating which characterizes the Belgian school system are actually hidden considerable social inequalities.

J.E.L. code: I21

# 1. Introduction

Lorsqu'on parle d'inégalités sociales en matière d'éducation, il faut distinguer deux choses : d'une part, l'inégalité des chances de réussite scolaire, et d'autre part, l'inégalité des chances de réussite professionnelle à niveau de formation donné. Ces deux sources d'inégalités se situent à des niveaux très différents. L'une trouve son origine dans le fonctionnement du système éducatif tandis que l'autre résulte du fonctionnement du marché du travail. Cependant, c'est de leurs effets cumulés que dépend l'impact du système éducatif sur la mobilité sociale et la répartition interpersonnelle des revenus.

Notre étude porte sur la première de ces sources d'inégalité. En bref, l'objet de notre recherche est d'évaluer empiriquement la mesure dans laquelle le parcours scolaire primaire et secondaire des enfants belges est lié à leur environnement socio-économique (éducation et statut professionnel des parents, composition et conditions matérielles du ménage, ...), et donc, en fin de compte, la mesure dans laquelle l'école tend à reproduire les inégalités sociales. Plus précisément, nous estimons, en fonction de son origine socio-économique et pour chacune des 12 premières années de sa scolarité, la probabilité pour un élève de conserver un parcours scolaire "normal", c.-à-d. sans redoublement ni, pour l'enseignement secondaire, relégation dans une filière de second choix. Nous utilisons pour cela une enquête belge portant sur environ 4500 ménages.

Cette façon d'approcher les inégalités sociales face à l'éducation est minimale : elle se contente de dégager (en termes de probabilité) le calendrier du "décrochage scolaire initial" selon l'origine socio-économique des élèves. Elle est cependant loin d'être dénuée d'intérêt. Elle constitue en effet la toute première étape du processus de production des inégalités sociales par le système éducatif, ce processus étant lui-même à la base des inégalités résultant du fonctionnement du marché du travail.

Outre le fait qu'il existe peu d'études empiriques de ce type portant sur des données belges<sup>1</sup>, notre contribution se distingue par le modèle économétrique utilisé pour estimer les probabilités de décrochage scolaire. Celui-ci repose sur l'hypothèse que le parcours scolaire primaire et secondaire des élèves peut être décrit par un processus de Markov. Comme nous le verrons plus loin, ce cadre d'analyse permet d'estimer aisément, en fonction de variables socio-économiques et pour chacune des 12 années de scolarité obligatoire, les probabilités individuelles de conserver un parcours normal, tel que nous l'avons défini. Surtout, il permet d'estimer ces probabilités à partir de données transversales, c.-à-d., sans disposer des informations relatives à l'ensemble de la carrière scolaire des élèves, ces informations n'étant pas disponibles lors de notre étude.

L'article est structuré de la manière suivante. La Section 2 trace à grands traits le cadre conceptuel sous-tendant notre étude. La Section 3 décrit brièvement le système éducatif belge. Le modèle économétrique utilisé est exposé dans la Section 4 tandis que les données et la spécification empirique sont discutés dans la Section 5.

---

<sup>1</sup> A notre connaissance, la seule étude belge basée sur des données individuelles est celle de Litt (1981). Toutefois, cet auteur s'est principalement limité à une analyse qualitative bivariée, sans mesurer, en termes de probabilité, l'ampleur des effets de l'origine sociale. Pour une étude récente sur données françaises, voir Duru-Bellat et al. (1993).

La Section 6 présente les résultats et la Section 7 quelques éléments de conclusion.

## 2. Le cadre conceptuel

La littérature reliée à la problématique des inégalités sociales à l'école est extrêmement vaste. Elle couvre des domaines aussi variés que l'économie, la sociologie, la pédagogie ou encore la psychologie. Il est bien entendu hors de question de survoler toute cette littérature. Nous nous contenterons ici de tracer à grand trait le cadre dans lequel s'inscrit la plupart des recherches menées en économie de l'éducation.

Depuis une trentaine d'années, la théorie du capital humain de Becker (1964) et ses divers prolongements dominant les recherches menées en économie de l'éducation. Selon cette théorie, la décision d'aller à l'école s'interprète comme le choix rationnel de dépenser du temps et de l'argent au début de son cycle de vie afin d'accroître sa productivité et donc ses revenus dans le futur. Dans sa forme initiale, cette théorie s'applique essentiellement à l'enseignement supérieur (non-obligatoire). Elle a cependant eu de nombreux prolongements, notamment en économie de la famille.

Ainsi, le ménage est souvent interprété comme un lieu d'échanges et de production de biens et de services. Entre autres, les parents choisissent le nombre d'enfants qu'ils désirent avoir et leur "qualité" (selon la terminologie de Becker et Lewis (1973)), c.-à-d., plus simplement, le capital humain qu'ils désirent leur donner. Pour accroître le capital humain de leur(s) enfant(s), les parents consacrent du temps et d'autres ressources. Ce processus d'accroissement débute dès le plus jeune âge de l'enfant, bien avant que celui-ci n'entre à l'école. Dès le début de l'apprentissage scolaire, les différences de capital humain seront à l'origine des inégalités importantes de résultats qui peuvent se manifester entre les enfants car, comme le souligne Leibowitz (1974), le coût d'acquisition du capital humain est généralement fonction des stocks déjà détenus. Certains enfants peu dotés en capital humain auront donc du mal à assimiler la matière et, le processus étant cumulatif, seront inmanquablement relégués.

Cette approche a mis en évidence un certain nombre de variables explicatives pertinentes. En résumé, on peut dire que les familles où il y a peu d'enfants, où les parents ont un niveau d'éducation élevé, et disposent de temps libre et de ressources financières confortables devraient être un lieu propice à l'accumulation de capital humain pré-scolaire chez les enfants. Ces variables seront utilisées, sous une forme ou sous une autre, dans la plupart des études ultérieures. Les aspects innés ne sont pas négligés; ceux-ci sont simplement considérés comme un intrant parmi d'autres dans le processus de production du capital humain. Celui-ci est seulement plus difficile à mesurer<sup>2</sup>.

On le voit, outre d'éventuelles différences dues aux établissements scolaires fréquentés — autre intrant majeur du processus éducatif —, cette approche at-

---

<sup>2</sup> Certaines théories extrêmes (voire extrémistes), telle celles de Jensen (1969) ou de Herrnstein et al. (1994), attribuent un rôle déterminant à l'hérédité: les bons résultats obtenus à l'école par les enfants des classes sociales favorisées reflèteraient seulement des capacités intellectuelles supérieures, capacités elles-mêmes génétiquement déterminées par l'intelligence de leurs parents. Ces théories se disqualifient toutefois par leurs propres excès (voir par exemple Goldberger *et al.* (1995)).

tribue la disparité des performances scolaires à des différences de capital humain accumulé. Mais quel est le contenu effectif de ce capital humain? La conception qui semble prévaloir dans la littérature économique est celle de considérer le capital humain comme un stock de connaissances, de compétences cognitives, plus ou moins mesurables par des tests standardisés. Cette interprétation est dans la droite ligne du schéma méritocratique dans lequel s'inscrivent les divers aspects de la théorie du capital humain.

D'autres interprétations, issues de la littérature en sociologie de l'éducation, peuvent être avancées. Ainsi, certains auteurs insistent sur le développement d'attitudes et de traits de caractères qui sont valorisés à l'école et ensuite dans l'entreprise, tels que la discipline, la motivation ou l'organisation (Gintis (1971), Bowles (1971), Bowles *et al.* (1976)). D'autres soulignent l'importance de certaines compétences culturelles ou linguistiques, caractéristiques qui sont l'apanage des classes favorisées (Boudon (1973), la notion de "capital culturel" chez Bourdieu (1977)). D'autres encore mettent en lumière le rôle des aspirations qui sont façonnées par l'environnement social (Willis (1977) et Boudon (1973)). Pour un survol (partiel) de cette littérature, voir Mehan (1992).

Compétences cognitives, attitudes et traits de caractères, capital culturel, aspirations, tous ces éléments sont certainement des médiateurs crédibles de la relation entre environnement socio-économique et performances scolaires des élèves. C'est bien entendu de leur identification précise que dépend le succès de toute politique visant à lutter contre les inégalités sociales face à l'école. Tel n'est cependant pas ici notre objectif. Plus modestement, nous nous attachons à quantifier l'intensité de la relation "brute" origine sociale - performance scolaire.

Il reste qu'il existe de nombreuses façons d'appréhender les performances scolaires des élèves, et par delà les performances du système éducatif. En économie, la plupart des recherches ont été menées dans le cadre des fonctions de production d'éducation et utilisent le plus souvent comme "indicateur de performance" les résultats des élèves à des tests standardisés (lecture, mathématique, etc.). Typiquement, ces recherches consistent en l'estimation d'une "fonction de production" reliant l'output du système éducatif — les résultats aux tests standardisés — à ses principaux inputs, d'une part les caractéristiques socio-économiques des élèves (composition du ménage, revenu, profession et éducation des parents, etc.) et, d'autre part, les caractéristiques des écoles qu'ils fréquentent (qualité des professeurs, taille des classes, variabilité des caractéristiques des élèves à l'intérieur des classes, etc.). Selon les études, l'accent est mis sur l'analyse de l'effet des ressources familiales ou de l'effet des ressources des écoles. Pour un résumé de ces recherches, voir Hanusek (1986, 1987) et Cohn et Geske (1990).

Comme le soulignent Morgenstern (1973) et Hanusek (1987), cette façon d'appréhender les performances scolaires est contestable. Pour reprendre les termes d'Hanusek (1987), elle manque de validation externe. Elle repose en fait sur une interprétation "productiviste" de la notion de capital humain : celui-ci serait composé de compétences scolaires objectives, compétences qui détermineraient la productivité des travailleurs et donc leur revenus. A cette vision, on peut opposer celle issue des théories de signalement<sup>3</sup> : le lien entre compétences scolaires et productivité

---

<sup>3</sup> Pour une discussion plus approfondie de ce point, voir Blaug (1985) et Eicher (1990). Il est à noter que

est remis en cause et le diplôme ou le parcours scolaire deviennent des signaux sur lesquels s'appuient les employeurs pour pratiquer une sorte de "discrimination statistique" à l'embauche. Dans cette perspective, évaluer l'output du système éducatif à l'aide de tests standardisés est économiquement non pertinent, tant du point de vue de la répartition des revenus que de l'allocation des ressources.

A l'opposé de la perspective productiviste qui vient d'être décrite, certains auteurs, généralement des sociologues, ont utilisé comme critère de réussite individuel des critères internes au système éducatif (par exemple, le diplôme). Ces auteurs négligent les intrants spécifiques de l'école mais souvent, ils insistent sur les aspects dynamiques du parcours scolaire. Ils abordent alors le parcours scolaire comme une suite de points de bifurcation (Boudon (1973)). Par exemple, Mare (1980) s'appuie sur une représentation du parcours scolaire comme une suite d'étapes au cours desquelles les élèves sont susceptibles de se maintenir ou bien de quitter l'enseignement. Hartog *et al.* (1989) procèdent de manière assez similaire.

Le présent article s'inscrit plutôt dans cette logique mais se consacre exclusivement à l'enseignement obligatoire. Il s'appuie implicitement sur l'idée que la réussite économique et sociale d'un individu sera davantage associée à son parcours scolaire qu'aux compétences très spécifiques qui sont exigées pour la réussite d'un test standardisé.

### 3. Le système éducatif belge

La Belgique est un état fédéral où les compétences en matière d'enseignement sont principalement détenues par deux institutions communautaires : la Communauté Française et la Communauté Flamande<sup>4</sup>.

Quelle que soit la communauté linguistique, le système scolaire<sup>5</sup> est organisé en deux cycles de 6 ans, appelés cycle primaire et cycle secondaire. Le premier cycle commence normalement à l'âge de six ans. De six à dix-huit ans, l'enseignement est obligatoire. De façon plus précise, l'obligation scolaire commence avec l'année scolaire qui prend cours dans l'année civile où l'enfant atteint l'âge de six ans et se termine à la fin de l'année scolaire de l'année civile au cours de laquelle il atteint l'âge de dix-huit ans. On notera que la loi autorise les parents à retarder, ou au contraire à avancer, d'un an l'entrée de leur enfant en 1<sup>ère</sup> primaire. Ils sont cependant en toutes circonstances tenus de respecter l'obligation scolaire. Ainsi, dans le cas où ils décident de la retarder, ils doivent présenter leur enfant en classe préparatoire maternelle.

Le cycle primaire est un cycle commun, homogène. Au début du cycle secondaire, les enfants sont orientés dans différentes filières en fonction de leurs goûts et de leurs résultats passés. Ces filières sont au nombre de trois : la filière générale, la filière technique et la filière professionnelle. La filière générale et la filière tech-

---

ce que nous désignons sous le vocable de "théories de signalement" recouvre un large spectre de théories.

<sup>4</sup>Pour être précis, il faut y ajouter la Communauté Germanophone. Celle-ci étant minuscule (elle représente 0.65% de la population belge), nous traiterons les données qui en émanent comme si elle était partie intégrante de la Communauté Française.

<sup>5</sup>Nous laisserons ici de côté l'enseignement spécial ainsi qu'un certain nombre de caractéristiques du système non indispensables à la compréhension notre étude.

nique ont pour but de préparer les élèves, respectivement, à l'enseignement supérieur général (y compris l'université) et à l'enseignement supérieur technique. Par opposition, la filière professionnelle a pour but de former, directement, les étudiants à une profession. Cette dernière filière ne donne d'ailleurs pas accès à l'enseignement supérieur, à moins de suivre une 7<sup>ème</sup> année de spécialisation.

Le système scolaire belge est caractérisé par des taux agrégés de redoublement extrêmement élevés (voir Eloy (1993)). Dès la fin du premier cycle, en Communauté Française (resp. en Communauté Flamande), le pourcentage d'élèves en retard atteint près de 20% (resp. 15%). Ce chiffre constitue un record parmi les pays de l'Union Européenne. A titre de comparaison, l'Allemagne, l'Espagne et la France affichent des scores de, respectivement, 2, 7, et 9%. A la fin du second cycle, le pourcentage d'élèves en retard est proche de 60% (resp. 40%)<sup>6</sup>. Remarquons également que, en Belgique, contrairement à ce qui se passe dans certains pays, les décisions de redoublement sont très largement du ressort de l'enseignant et du conseil de classe.

A côté du redoublement existe un important phénomène de réorientation en cours de cycle. Pour plus de détails, voir le document de l'IRES (1992). Ces réorientations se font quasi-exclusivement de la filière générale vers les filières technique et professionnelle. Les réorientations en sens inverse sont rarissimes. La réorientation vers la filière professionnelle, en provenance du général ou du technique, est un important substitut du redoublement, en particulier en début de cycle secondaire : il n'est pas rare qu'un élève en difficulté se voit proposer le choix entre un redoublement dans sa filière et le passage en professionnel. Il va sans dire que peu d'élèves choisissent cette filière par vocation.

## 4. Le modèle économétrique

Nous l'avons dit, notre approche repose sur l'hypothèse que le parcours scolaire des élèves peut être représenté par un processus de Markov. La première partie de cette section précise notre modèle. La seconde partie envisage le problème de son estimation.

### 4.1. Parcours scolaire et chaîne de Markov

Communément, le parcours scolaire d'un élève est décrit par l'évolution de sa "position", repérée par l'année et la filière d'étude, dans le cursus scolaire au cours du temps. Exprimée de la sorte, la notion de parcours scolaire est difficile à modéliser. Elle comprend un trop grand nombre de situations possibles. Il convient donc de la simplifier. Une façon attractive de le faire est de définir un parcours scolaire de référence, puis d'exprimer, en les regroupant de manière plus ou moins fine, les autres situations en tant que déviations par rapport à ce parcours de référence.

Dans le cadre de cette étude, le parcours scolaire de référence est défini par l'absence de redoublements et/ou de réorientation vers la filière professionnelle. Les

---

<sup>6</sup> Au niveau du second cycle, les comparaisons internationales sont difficiles car la durée de l'obligation scolaire diffère d'un pays à l'autre.

autres alternatives sont regroupées en une seule situation, tout simplement celle de s'être écarté du parcours scolaire de référence. Selon cette grille, à tout moment de sa scolarité, un élève peut donc être classé dans l'un des deux états — exhaustifs et mutuellement exclusifs — suivants : “poursuit un parcours scolaire normal” (ci-après dénoté  $R$ , pour réussite) ou “est en situation de retard scolaire” (ci-après dénoté  $E$ , pour échec).

Soit le vecteur  $(2 \times 1)$   $Y_i(t)$ , l'état d'un individu  $i$  à la période  $t$ . Par convention, nous supposons que  $Y_i(t) = [1 \ 0]'$  lorsque l'individu est dans l'état  $R$  et que  $Y_i(t) = [0 \ 1]'$  lorsque l'individu est dans l'état  $E$ . L'index de période  $t$ , appelé période de transition, dénombre les années scolaires passées dans l'enseignement obligatoire. L'indice  $t = 0$  désigne l'année scolaire qui précède l'entrée dans l'enseignement obligatoire, soit l'année scolaire qui prend cours dans l'année civile où l'enfant atteint l'âge de cinq ans. L'indice  $t = 1$  indique l'année scolaire normale de rentrée dans l'enseignement obligatoire, l'indice  $t = 2$ , la seconde année de scolarité obligatoire, et ainsi de suite. Finalement, l'indice  $t = 13$  désigne l'année scolaire au cours de laquelle les élèves ne sont plus légalement tenus de suivre un enseignement, soit, pour les étudiants ayant accomplis un parcours scolaire sans faute et qui poursuivent leurs études, la première année scolaire où ils fréquentent l'enseignement supérieur. On notera que l'indice  $t$  des périodes de transition n'est lié à aucune date calendrier précise; pour deux élèves nés au cours d'années différentes, la période de transition  $t = 0$  correspond à des dates calendrier différentes. Il est en revanche intimement lié à l'âge des élèves, les éventuels décalages n'étant le résultat que de la non-correspondance entre années scolaires et années civiles.

L'état  $Y_i(t)$  dans lequel se trouve un individu  $i$  au moment  $t$  est déterminé par la confrontation entre sa période de transition  $t$  et sa “position”, en termes d'année d'étude et de filière, dans le cursus scolaire. Par définition, en  $t = 0$ , tous les élèves sont dans l'état  $R$  ( $Y_i(0) = [1 \ 0]'$ ). En  $t = 1, 2, \dots, 6$ , l'élève est déclaré poursuivre un parcours “normal” si et seulement si il fréquente respectivement la 1<sup>ère</sup>, 2<sup>ème</sup>, ..., 6<sup>ème</sup> année primaire. De même pour  $t = 7, 8, \dots, 12$ , mais cette fois, outre le fait qu'il doit fréquenter respectivement la 1<sup>ère</sup>, 2<sup>ème</sup>, ..., 6<sup>ème</sup> secondaire, il ne peut être dans la filière professionnelle. Enfin, en  $t = 13$ , est dans l'état  $R$  un élève qui a son diplôme d'enseignement secondaire (général ou technique) en poche.

L'hypothèse fondamentale sur laquelle repose notre analyse est que l'évolution temporelle du vecteur d'état  $Y_i(t)$  peut être représentée par une chaîne de Markov d'ordre 1 dont la matrice des probabilités de transition  $P_i(t)$ , matrice décrivant les probabilités de passage d'un état à un autre entre les périodes de transition  $(t - 1)$  et  $t$ , s'écrit

$$P_i(t) = \begin{bmatrix} p_i^{RR}(t) & p_i^{RE}(t) \\ 0 & 1 \end{bmatrix} \quad (1)$$

où  $p_i^{RR}(t)$  est la probabilité de passer de l'état  $R$  à l'état  $R$  entre  $(t - 1)$  et  $t$ , et  $p_i^{RE}(t) = 1 - p_i^{RR}(t)$  est la probabilité de passer l'état  $R$  à l'état  $E$  entre  $(t - 1)$  et  $t$ . L'écart par rapport au parcours scolaire standard étant irréversible, la probabilité de passer de l'état  $E$  à l'état  $R$  est nulle et celle de passer de l'état  $E$  à l'état  $E$  est unitaire. L'état  $E$  est un état absorbant. Cette chaîne de Markov est dite d'ordre 1 car il est supposé que la probabilité pour un individu de se trouver dans un état donné en  $t$  ne dépend du passé que par l'intermédiaire de son état en  $(t - 1)$ .



Formellement,  $\forall y_i(t) = \{[1 \ 0]', [0 \ 1]'\}$

$$\begin{aligned} & \text{Prob}[Y_i(t) = y_i(t)/Y_i(t-1) = y_i(t-1), \dots, Y_i(1) = y_i(1), Y_i(0) = y_i(0)] \\ & = \text{Prob}[Y_i(t) = y_i(t)/Y_i(t-1) = y_i(t-1)] \end{aligned} \quad (2)$$

Ainsi, le parcours scolaire d'un individu est intégralement décrit par l'ensemble des probabilités conditionnelles  $\text{Prob}[Y_i(t) = y_i(t)/Y_i(t-1) = y_i(t-1)]$  dont les valeurs sont données par ( $t = 1, \dots, 13$ )

$$\text{Prob}[Y_i(t) = y_i(t)/Y_i(t-1) = y_i(t-1)] = y_i(t)' P_i(t)' y_i(t-1) \quad (3)$$

Etant donné la structure très simple des  $P_i(t)$ , toutes ces probabilités conditionnelles sont entièrement gouvernées par les seules probabilités de transition  $p_i^{RR}(t)$ . Celles-ci ont une signification très précise. La probabilité de transition  $p_i^{RR}(1)$  correspond à la probabilité d'entrer dans l'enseignement primaire à l'âge normal. De son côté,  $p_i^{RR}(2)$  correspond, pour un individu entré à l'âge normal dans l'enseignement primaire, à la probabilité de réussir sa première primaire. De même,  $p_i^{RR}(3)$  correspond, pour un individu ayant passé avec succès ces deux premières épreuves c.-à-d. étant en deuxième primaire lors de sa seconde année de scolarité obligatoire, à la probabilité de réussir sa deuxième primaire. Et ainsi de suite. De manière générale,  $p_i^{RR}(t)$  est la probabilité de réussir sa ( $t-1$ ) ième année scolaire, sachant que l'on a déjà réussi toutes les précédentes.

Les probabilités de transition  $p_i^{RR}(t)$  ne sont bien entendu pas constantes. Clairement, elles dépendent des périodes de transition  $t$ . On peut également s'attendre, et c'est là l'objet de notre étude, à ce qu'elles dépendent de l'origine socio-économique des élèves. Ainsi, on supposera finalement que

$$p_i^{RR}(t) = f(X_{it}, \theta), \quad f(.) = \frac{\exp(.)}{1 + \exp(.)} \quad (4)$$

où  $X_{it}$  dénote un vecteur de variables exogènes dépendant des caractéristiques de l'individu  $i$  et/ou de la période de transition  $t$ ,  $\theta$  est un vecteur de paramètres et  $f(.)$  est une fonction de lien logistique garantissant que  $p_i^{RR}(t)$  est compris entre 0 et 1.

Connaissant les relations (1), (2) et (3), et en s'appuyant sur le fait que, par définition,  $Y_i(0) = [1 \ 0]'$ , on peut aisément dégager les probabilités non conditionnelles pour un individu  $i$  de se trouver dans un état donné en  $t = 1, 2, \dots, 13$ . Ainsi,  $\text{Prob}[Y_i(1) = y_i(1)] = \text{Prob}[Y_i(1) = y_i(1)/Y_i(0) = y_i(0)] = y_i(1)' P_i(1)' y_i(0)$ , avec  $y_i(0) = [1 \ 0]'$ . Pour  $t = 2$ , on a

$$\begin{aligned} & \text{Prob}[Y_i(2) = y_i(2)] \\ & = \sum_{\forall y_i(1) = \{[1 \ 0]', [0 \ 1]'\}} \text{Prob}[Y_i(2) = y_i(2)/Y_i(1) = y_i(1)] \text{Prob}[Y_i(1) = y_i(1)] \\ & = \sum_{\forall y_i(1) = \{[1 \ 0]', [0 \ 1]'\}} y_i(2)' P_i(2)' y_i(1) y_i(1)' P_i(1)' y_i(0) \\ & = y_i(2)' P_i(2)' P_i(1)' y_i(0) \end{aligned} \quad (5)$$

Par récurrence, on obtient la forme générale

$$\text{Prob}[Y_i(t) = y_i(t)] = y_i(t)' \left( \prod_{s=0}^{t-1} P_i(t-s)' \right) y_i(0) \quad (6)$$

Etant donné la structure des  $P_i(t)$ , l'équation (6) est nettement plus simple qu'il n'y paraît à première vue. En effet, sous forme détaillée, la probabilité non conditionnelle pour un individu  $i$  de se trouver dans l'état  $R$  (parcours scolaire normal) en  $t$  s'écrit simplement ( $t = 1, 2, \dots, 13$ )<sup>7</sup>

$$\pi_i^R(t) = \text{Prob}[Y_i(t) = [1 \ 0]'] = \prod_{s=1}^t p_i^{RR}(s) \quad (7)$$

tandis que la probabilité non conditionnelle de se trouver dans l'état  $E$  en  $t$  est égale à

$$\pi_i^E(t) = \text{Prob}[Y_i(t) = [0 \ 1]'] = 1 - \pi_i^R(t) \quad (8)$$

Vue comme une fonction de  $t$ , la probabilité  $\pi_i^R(t)$  s'interprète comme une fonction de survie dans l'état "parcours scolaire normal".

## 4.2. Estimation

Nous l'avons dit, les probabilités de transition  $p_i^{RR}(t)$  gouvernent l'ensemble de la chaîne de Markov décrite ci-dessus. Celles-ci étant, selon la relation (4), fonction de variables exogènes, l'estimation du modèle revient donc à estimer le vecteur de paramètre  $\theta$ .

Le jeu de données idéal pour estimer notre modèle consisterait assez naturellement en un panel décrivant l'ensemble de la carrière scolaire de  $n$  individus, c.-à-d., pour chaque individu, l'ensemble des  $Y_i(t)$  ( $t = 1, 2, \dots, 13$ ). Ce type de données, qui n'était pas disponible au moment de l'étude, n'est cependant nullement nécessaire. L'estimation de ce modèle est parfaitement possible lorsque l'on n'observe, pour chaque individu  $i$ , qu'une partie seulement de l'ensemble des  $Y_i(t)$ , voire même, et c'est notre cas, lorsque l'on n'en observe qu'un seul élément.<sup>8</sup>

Cette dernière situation correspond typiquement au cas des données en coupe (données transversales): l'échantillon est composé d'un ensemble de  $n$  individus répartis dans les différentes périodes de transition  $t$ ; Pour chaque individu, on n'observe qu'un seul  $Y_i(t)$ . La procédure d'estimation est alors basée sur l'expression des probabilités non conditionnelles (6).

Ainsi, sous l'hypothèse d'un tirage indépendant dans une surpopulation de taille infinie, l'estimateur maximum de vraisemblance de  $\theta$  est donné par la solution du

<sup>7</sup> Exprimé sous la forme (7), on peut voir la parenté de notre modèle avec un modèle logit séquentiel. Tant que l'élève réussit, il est confronté au cours de chaque année au même "choix": réussir ou échouer. Cependant, dès qu'il a échoué, tout choix disparaît. Il reste par définition dans une situation de retard scolaire. Pour une application du modèle logit séquentiel à la carrière scolaire (aux Pays-Bas), voir Hartog, Pfann et Ridder (1989).

<sup>8</sup> Pour un aperçu général des méthodes d'estimation des chaînes de Markov, voir Gouriéroux (1989).

problème d'optimisation

$$Max_{\theta} \sum_{i=1}^n \ln \left( y_i(t_i)' \left( \prod_{s=0}^{t_i-1} P_i(t_i - s)' \right) y_i(0) \right), \quad \text{avec } y_i(0) = [1 \ 0]' \quad (9)$$

et où  $t_i$  désigne la période de transition observée de l'individu  $i$ . Sous des hypothèses de régularité adaptées, cet estimateur possède les propriétés habituelles du maximum de vraisemblance : convergence, efficacité et normalité asymptotique.

Une dernière remarque. Si, comme nous venons de le voir, l'estimation de notre modèle peut être réalisée sans observer l'ensemble des trajectoires individuelles  $Y_i(t)$ , il n'en va évidemment pas de même pour les variables exogènes  $X_{it}$ . La valeur des  $X_{it}$  doit être connue pour toutes les périodes précédant celle de l'observation  $Y_i(t_i)$ , soit pour tout  $t = 1, 2, \dots, t_i$ . Dans le cas qui nous occupe, ce problème est limité car la majorité des variables considérées sont soit par nature invariantes, soit peuvent être supposées stables dans le temps. Il n'est toutefois pas totalement absent. Nous y reviendrons.

## 5. Les données et la spécification empirique

Notre étude s'appuie sur une enquête menée sur l'ensemble du territoire belge en 1992 auprès de 4445 ménages<sup>9</sup>. Les informations contenues dans cette banque de données touchent à de nombreux aspects de la vie familiale, sociale et culturelle du ménage, ainsi qu'à la formation, aux activités professionnelles et aux revenus des parents. Elle inclut également des renseignements concernant les enfants, leur santé, leurs activités et leur scolarité.

De cette base de données, nous utilisons un sous-échantillon se rapportant aux enfants en âge de scolarité obligatoire. La section 5.1 décrit sa composition. Les variables explicatives considérées et la forme précise de la relation (4) — relation déterminant les  $p_i^{RR}(t)$  — sont discutées dans la Section 5.2.

### 5.1. Echantillon

Notre échantillon est composé de tous les enfants de la base de données satisfaisant simultanément les critères suivants : (1) être en âge de normalement fréquenter l'enseignement obligatoire ou la première année d'enseignement supérieure, c.-à-d. être dans l'une des périodes de transition 1 à 13, (2) ne pas être dans l'enseignement spécial, (3) appartenir à un ménage biparental composé de ses parents biologiques, et (4) ne pas présenter de valeurs manquantes pour la variable endogène<sup>10</sup> et/ou les variables exogènes.

<sup>9</sup> Cette enquête constitue la première vague du *Panel Study on Belgian Household*, réalisé conjointement par les universités de Liège et d'Anvers.

<sup>10</sup> Un nombre important d'observations (de l'ordre de 300) ont été perdues suite à l'impossibilité de reconstituer la variable "état de retard" sur base des renseignements disponibles. Plus précisément, toutes les observations correspondant à des individus interrogés durant les vacances d'été ont dû être écartées car nous ne savons pas l'année scolaire qu'ils désignaient — l'année venant de se terminer ou la suivante — lorsqu'ils nous rapportaient leur année et filière d'étude.

L'exigence d'appartenance à un ménage biparental composé de ses parents biologiques est liée à la nécessité, évoquée ci-dessus, de disposer des valeurs des variables explicatives  $X_{it}$  — en l'occurrence ici l'environnement socio-économique des enfants — pour toutes les périodes précédant celle de l'observation d'un individu. Nous ne disposons pas de l'information nécessaire pour reconstituer l'histoire des cellules familiales non biparentales. Nous nous sommes donc limités aux seuls ménages biparentaux, ménages pour lesquels nous pouvons raisonnablement supposer la stabilité dans le temps du noyau familial (absence de ruptures et/ou de recompositions). Bien entendu, en contrepartie, nos résultats ne valent que pour ce type de ménage.

Sur base de l'application de ces critères, nous avons obtenu un sous-échantillon utile de 1083 enfants. Le Tableau 1 décrit sa structure.

Tableau 1 : Description de l'échantillon

Période de transition	Age normal	Proportions d'enfants			Effectifs de la cohorte
		Normal	En retard		
			En prof. (sans redoubl.)	Autres	
1	6	93.5	0.0	6.5	92
2	7	85.4	0.0	14.6	103
3	8	86.3	0.0	13.6	102
4	9	93.1	0.0	6.9	73
5	10	78.6	0.0	21.3	89
6	11	84.2	0.0	15.8	76
7	12	79.1	7.7	13.2	91
8	13	74.0	7.0	19.0	100
9	14	56.1	12.2	31.7	82
10	15	60.3	6.4	33.1	78
11	16	50.7	5.7	43.6	69
12	17	42.7	5.3	52.0	75
13	18	45.3	0.0	54.7	53

Pour rappel, les périodes de transition dénombrent les années passées dans l'enseignement obligatoire. La colonne intitulée "Age normal" désigne, pour chaque période de transition, l'âge atteint par les individus au cours de l'année civile correspondant au début de l'année scolaire.

Les colonnes 3 à 5 reportent la répartition des effectifs de chaque cohorte selon leur état. La tendance à la baisse des effectifs par cohorte dans les périodes de transition élevées traduit essentiellement la raréfaction, avec les années qui passent, des ménages biparentaux. De manière à illustrer l'importance du phénomène de substitution entre redoublement et passage en professionnel, l'état "retard scolaire" est décomposé dans les colonnes 4 et 5 entre, respectivement, les élèves se trouvant en professionnel mais sans redoublement, et les autres. On notera que les enfants en avance d'un an sont traités comme s'ils poursuivaient un parcours scolaire normal.

A la lecture du Tableau 1, on constate que, dans notre échantillon, 6.5% des enfants qui sont en période de transition 1 accusent un retard : ils sont toujours

à l'école maternelle alors qu'ils devraient normalement être en 1ière primaire. A l'autre extrémité du cursus, soit après 13 ans, plus de 50% des élèves de la cohorte sont en retard : soit ils n'ont toujours pas terminé leurs études secondaires, soit ils les ont terminées, mais en professionnel. Comme le suggère l'évolution des chiffres repris à la colonne 4 — une augmentation suivie d'une diminution —, les élèves passés en professionnel sans redoublement ont une tendance nette à redoubler dans les années futures.

## 5.2. Spécification empirique

Le choix de la spécification empirique du modèle recouvre deux problèmes distincts : le choix d'une forme analytique générique et le choix des variables exogènes représentant l'origine socio-économique des élèves. Nous traiterons ces deux problèmes successivement.

Les paramètres d'un modèle à variables dépendantes qualitatives sont toujours, par la nature même du modèle, difficiles à estimer avec précision. Dans ce type de modèle, il est donc particulièrement important de sélectionner des formes analytiques génériques qui, outre leur adéquation au problème étudié, sont les plus parcimonieuses possibles en paramètres. C'est à cette logique que répond la spécification empirique décrite ci-dessous.

La spécification générique retenue pour l'expression (4) des probabilités de transition  $p_i^{RR}(t)$  comporte trois ensembles de paramètres. Le premier se rapporte à la modélisation de l'évolution temporelle des probabilités de transition. Pour des raisons institutionnelles évidentes, cette évolution temporelle est supposée différente en Communauté Française et en Communauté Flamande. Le second ensemble de paramètres est lié aux effets relatifs des différentes composantes (variables) de l'origine sociale des individus sur leur succès scolaire. Nous nommerons cet ensemble de paramètres, joint aux variables qui y correspondent, le “noyau”. Enfin, le dernier groupe de paramètres introduit des “paramètres d'échelle” relativisant l'effet des variables du noyau selon les périodes de transition. Formellement,

$$p_i^{RR}(t) = f(Z_{it}\alpha + I_{2 \leq t \leq 6}(W_{it}\beta) + I_{t=7}\lambda_1(W_{it}\beta) + I_{t \geq 8}\lambda_2(W_{it}\beta)) \quad (10)$$

où  $I_{(\cdot)}$  est une variable indicatrice ( $I_{(\cdot)} = 1$  si  $(\cdot)$  est vérifié,  $I_{(\cdot)} = 0$  sinon),  $Z_{it}$  désigne un vecteur de variables modélisant l'évolution temporelle “autonome” (différenciée selon la communauté où l'individu poursuit sa scolarité) des probabilités de transition,  $W_{it}$  dénote un vecteur de variables exogènes (noyau) représentant l'origine socio-économique des élèves, et  $\alpha$ ,  $\beta$ ,  $\lambda_1$  et  $\lambda_2$  sont des (vecteurs de) paramètres.

Cette forme analytique distingue donc quatre “régimes d'influence” des variables du noyau sur les probabilités de succès scolaire. En  $t = 1$ , soit lors de la décision d'entrée en 1ière primaire, ces variables sont censées ne pas jouer. Un second régime couvre les périodes de transition 2 à 6, soit la réussite des 5 premières années primaires<sup>11</sup>. Un régime spécial est réservé au passage de la 6ième primaire à la 1ière secondaire ( $t = 7$ ). Enfin, un dernier régime est affecté à l'ensemble du

---

<sup>11</sup> Rappelons que  $p_i^{RR}(t)$  désigne la probabilité de réussir sa  $(t - 1)$  ième année scolaire, sachant que l'on a déjà réussi toutes les précédentes. Au cours des période 2 à 6, le paramètre d'échelle est implicitement fixé à 1 (normalisation).

cycle secondaire.

Toujours en ce qui concerne la forme analytique générique (10), on notera finalement sa forte non-linéarité, introduite par la fonction (monotone et strictement croissante) de lien logistique  $f(\cdot)$ . Il est important de conserver cette caractéristique à l'esprit lors de l'interprétation des résultats car elle est à l'origine d'effets d'interaction entre les variables du noyau et les variables représentant l'évolution temporelle "autonome" des probabilités de transition. En clair, l'effet marginal des premières sur les  $p_i^{RR}(t)$  dépend du niveau des secondes.

L'origine sociale ou environnement socio-économique des élèves est un concept par nature multidimensionnel. Dans cette étude, nous l'avons appréhendé au travers des variables suivantes :

- 1- Niveau d'éducation du père (3 niveaux).
- 2- Niveau d'éducation de la mère (6 niveaux).
- 3- Statut professionnel du père (ouvrier - autre).
- 4- Statut professionnel de la mère (sans profession : jamais travaillé - autre).
- 5- Nationalité à la naissance du père (autochtone - étranger).
- 6- Revenu du ménage.
- 7- Aisance financière du ménage (6 niveaux : très difficile à très facile).
- 8- Langue parlée à la maison (différente - identique à celle parlée à l'école).
- 9- Activités culturelles du ménage (existantes - inexistantes).
- 10- Abandon prématuré des études par la mère (car mauvais résultats : oui-non).
- 11- Taille du ménage (nombre d'enfants).

Cet ensemble de variables reprend, en y ajoutant quelques-unes d'ordre culturelles, la plupart des caractéristiques socio-économiques des ménages généralement prises en compte dans la littérature empirique<sup>12</sup>.

Avant de poursuivre, une précision s'impose. Notre objectif ici n'est nullement d'expliquer, au sens causal strict du terme, les performances scolaires des élèves au moyen des variables décrites ci-dessus. Comme on l'a déjà souligné dans la Section 2, les variables caractérisant l'environnement socio-économique des élèves ne peuvent au mieux être considérées que comme des variables "proxy" des déterminants fondamentaux de la carrière scolaire. La présente étude doit plutôt être appréhendée en termes d'analyse discriminante. Il s'agit d'évaluer la mesure dans laquelle il est possible de discriminer, de prédire, le parcours scolaire des individus sur base de leur origine sociale. C'est dans cette optique qu'il convient également d'apprécier l'absence de toute variable relative aux établissements scolaires.

Il nous reste à examiner le problème de la stabilité temporelle des variables exogènes prises en compte. Clairement, les variables (1), (2), (5), (8) et (10) sont par nature invariantes dans le temps. Leur valeur observée durant la période d'observation peut donc être reportée sur les périodes de transition précédentes. Le même raisonnement vaut pour les variables (3), (4) et (9). Ces variables ne sont pas invariantes par nature mais peuvent être supposées stables de par leur construction dichotomique tranchée. Le passé de la variable "Taille du ménage" a pu être entièrement reconstitué. Reste donc le problème des variables "Revenu du ménage" et "Aisance financière du ménage". Ne pouvant faire autrement, nous les avons

---

<sup>12</sup> Voir à ce propos par exemple Alwin et al. (1984).

utilisées en l'état.

## 6. Les résultats empiriques

Le modèle a été estimé par la méthode du maximum de vraisemblance. Tous les calculs, en particulier la résolution du programme d'optimisation (9), ont été réalisés en Gauss.

Nous présenterons nos résultats en deux étapes. Dans un premier temps, nous discuterons le signe et la significativité des coefficients estimés de l'expression (10) — les probabilités de transition  $p_i^{RR}(t)$  —, expression qui, rappelons-le, gouverne l'ensemble de la chaîne de Markov. Dans un second temps, de manière à faire apparaître l'ampleur de l'effet des variables exogènes en termes de probabilité, nous les illustrerons graphiquement sur base de simulations.

### 6.1. Paramètres estimés

Le Tableau 2 présente les coefficients estimés du modèle simplifié — c.-à-d. débarrassé des variables non significatives<sup>13</sup> — finalement choisi. La définition précise des variables retenues est détaillée en annexe.

Tableau 2: Paramètres estimés du modèle réduit

Paramètres	Variables	Estimation	Ecart-type
$\alpha_1$	$I_{(\text{Communauté Française})}$	3.210**	0.357
$\alpha_2$	$I_{(\text{Communauté Française})} \times \ln(t)$	1.312**	0.562
$\alpha_3$	$I_{(\text{Communauté Française})} \times I_{(t \geq 8)}$	-4.563**	1.318
$\alpha_4$	$I_{(\text{Communauté Flamande})}$	2.708**	0.313
$\alpha_5$	$I_{(\text{Communauté Flamande})} \times \ln(t)$	2.012**	0.674
$\alpha_6$	$I_{(\text{Communauté Flamande})} \times I_{(t \geq 8)}$	-4.042**	1.054
$\beta_1$	Niveau d'éducation du père <sup>a</sup>	0.676**	0.231
$\beta_2$	Niveau d'éducation de la mère <sup>a</sup>	0.392**	0.112
$\beta_3$	Aisance financière <sup>a</sup>	0.888**	0.257
$\beta_4$	Père ouvrier	-1.016**	0.318
$\beta_5$	Autre langue parlée dans le ménage	-1.049*	0.478
$\beta_6$	Absence de vie culturelle	-0.497*	0.235
$\lambda_1$	—	1.572**	0.569
$\lambda_2$	—	0.477**	0.133

\* significatif à 5%, \*\* significatif à 1%.

<sup>a</sup> variable centrée à son mode.

Six variables du noyau sur 11 se sont donc avérées — plutôt fortement — non significatives. Il s'agit du statut professionnel de la mère (sans profession : jamais travaillé - autre), de la nationalité du père (autochtone - étranger), du revenu du ménage, de l'abandon prématuré des études par la mère (suite à de mauvais résultats : oui-non) et enfin de la taille du ménage (nombre d'enfants).

<sup>13</sup> Le modèle a été réduit par éliminations successives des variables non-significatives au seuil de 5%.

Interprétée dans l’optique de la théorie du capital humain, la non-significativité des variables “taille de la famille” et “statut professionnel de la mère” semble indiquer que la quantité de temps<sup>14</sup> que les parents peuvent consacrer à leurs enfants n’est jamais contrainte au point de nuire à leur scolarité. Bien plus déterminante semble être la “qualité” du temps, dont il est généralement supposé que le niveau d’éducation des parents est un bon indicateur. On constate en effet que tant le niveau d’éducation du père que de la mère ont un impact positif très significatif sur la réussite.

Quelle que soit la spécification utilisée, la variable “Revenu du ménage” n’a jamais été trouvée significative. Est-ce à dire que les conditions matérielles du ménage n’influencent guère la scolarité des enfants? Pas nécessairement, et ce pour au moins deux raisons. D’abord, on l’a dit, notre mesure du revenu familial est très médiocre (les valeurs historiques du revenu n’ont pu être reconstituées). Mais surtout, la variable “Aisance financière” s’est, elle, avérée très significative. Or on peut penser que cette variable rend plus finement compte des ressources financières du ménage que le revenu courant et qu’elle est relativement plus stable dans le temps<sup>15</sup>.

La profession du père semble également jouer un rôle : les enfants d’ouvriers réussiraient nettement moins bien que ceux des autres classes sociales. La profession traduit, sans doute, des effets de culture familiale qui ne sont pas repris par les autres variables. On pense par exemple au rôle des aspirations (ambitions). En revanche, ainsi que l’indique la non-significativité de la variable “Abandon prématuré des études par la mère”, une mauvaise expérience scolaire des parents ne semble pas être aussi “transmissible”.

Comme on pouvait s’y attendre, pratiquer à la maison une langue étrangère à celle de l’école constitue un sérieux handicap scolaire. De même pour l’absence d’activités culturelles à l’extérieur du ménage. On retrouve là des éléments chers aux tenants de la notion de “capital culturel”.

Pour en terminer avec les variables du noyau, on notera finalement la non-significativité de la nationalité du père. Pas de différences notables donc entre parents d’origine autochtone et étrangère<sup>16</sup>. En clair, la nationalité en tant que telle ne permet de capturer aucune caractéristique pertinente supplémentaire à celles déjà prises en compte par les autres variables du modèle.

Il nous reste à examiner les variables modélisant l’évolution temporelle “autonome” des probabilités de transition ainsi que les paramètres d’échelle. Pour ce qui est des variables temporelles, on notera tout d’abord la forme fonctionnelle finalement retenue pour chaque communauté. Il s’agit d’un simple trend logarithmique “fracturé” par une variable binaire à la période  $t = 8$  (dans le cadre d’un parcours normal, passage de la 1<sup>ière</sup> à la 2<sup>ième</sup> année du secondaire)<sup>17</sup>. On remarquera ensuite que, bien qu’ils ne soient pas significativement différents l’un de l’autre, nous avons

<sup>14</sup> Dans une famille nombreuse, la quantité de temps que les parents peuvent consacrer à chaque enfant est forcément plus limitée que dans les ménages à enfant unique. De même, une mère qui travaille a moins de temps à consacrer à ses enfants qu’une mère femme au foyer.

<sup>15</sup> On peut voir cette variable comme un indicateur plus ou moins fiable du revenu permanent.

<sup>16</sup> La distinction entre étranger CEE et étranger non-CEE n’est pas plus significative.

<sup>17</sup> Cette forme fonctionnelle pourrait paraître exagérément rigide. Nos investigations nous ont cependant montré qu’une forme plus souple n’était guère pertinente.



conservé des profils temporels distincts pour chaque communauté. Deux raisons à cela. La première est d’ordre institutionnelle. Elle a déjà été évoquée. La seconde est de nature statistique. Etant donné la structure de l’information fournie par l’échantillon, on ne pouvait *a priori* pas s’attendre à obtenir des estimations très précises de l’évolution temporelle des probabilités de transition<sup>18</sup>. La non-significativité de la différence des profils temporels communautaires doit donc davantage s’interpréter comme la conséquence d’imprécisions d’estimations que comme le reflet d’une absence de différences structurelles, différences structurelles qui, comme en attestent les chiffres issus de données agrégées reportés à la section 3, existent bel et bien.

Cela étant, au vu du signe des paramètres  $\alpha_1 - \alpha_6$ , on peut assez facilement tracer un premier portrait-robot de l’évolution au cours du temps, *ceteris paribus*, des probabilités de dévier du parcours scolaire “normal”. Il se résume à deux grandes caractéristiques : (1) Au sein de chaque cycle (primaire et secondaire), une diminution progressive des risques d’encourir un échec (reflet du trend), et (2), un fort accroissement de ce risque en début de cycle secondaire (reflet de la binaire  $I_{(t \geq 8)}$ ). Ce schéma est valable tant en Communauté Française qu’en Communauté Flamande.

Finalement, en ce qui concerne les paramètres d’échelle, étant donné la forme non-linéaire de la relation (10), on se gardera de toute interprétation hâtive. Ainsi, le fait que  $\lambda_2$  soit significativement inférieur à 1 (le paramètre d’échelle implicite du cycle primaire) ne signifie nullement que l’origine socio-économique a un impact plus faible dans le cycle secondaire que dans le cycle primaire. Il en est évidemment de même pour le paramètre  $\lambda_1$ <sup>19</sup>. En fait, l’effet de ces paramètres ne peut pas être appréhendé isolément, mais seulement au travers de l’ensemble de la relation (10)<sup>20</sup>. Ceci nous amène tout naturellement aux simulations.

## 6.2. Simulations

Les résultats de l’estimation reportés au Tableau 2 ne permettent guère de se faire une idée précise de l’ampleur de l’effet des variables exogènes — leur pouvoir discriminant — sur le parcours scolaire des élèves.

La façon la plus simple d’illustrer ce point est de représenter graphiquement le profil des probabilités de survie  $\pi_i^R(t)$  ( $t = 1, 2, \dots, 13$ ) des élèves dans l’état “parcours scolaire normal” pour différents ensembles particuliers de valeurs des variables exogènes, chaque ensemble particulier de valeurs pouvant être assimilé à un scénario-type.

Nous avons retenu quatre scénarios-types. Ces scénarios, ainsi que leur représentativité dans chaque communauté linguistique, sont décrits par le Tableau 3.

---

<sup>18</sup> Les effectifs disponibles par période de transition sont bien trop faibles. En fait, l’information contenue dans l’échantillon concerne essentiellement les différences inter-individuelles. A ce niveau, la non-significativité d’une variable peut légitimement être interprétée, sous les réserves d’usage, comme une absence structurelle d’effet.

<sup>19</sup> On notera que ce paramètre, contrairement à  $\lambda_2$ , n’est guère significativement différent de 1. Pour des raisons semblables à celles évoquées ci-dessus à propos des profils temporels communautaires, nous l’avons néanmoins conservé. En fait, le fixer à 1 ne modifie que très marginalement nos résultats.

<sup>20</sup> Cela découle bien entendu des effets d’interaction évoqués ci-dessus. Ainsi, l’ampleur de l’effet “multiplicateur” des paramètres d’échelle sur les variables du noyau dépend de la valeur prise par les variables temporelles.

Tableau 3: Scénarios de simulation

Scénario	(1)	(2)	(3)	(4)
Educ. du père	Supérieur	Secondaire	Secondaire	Primaire
Educ. de la mère	Supérieur long	Secondaire sup.	Secondaire inf.	Primaire
Statut prof. du père	Non-ouvrier	Non-ouvrier	Ouvrier	Ouvrier
Aisance financière	Facile	Plutôt facile	Plutôt facile	Plutôt difficile
Langue parlée	Commun.	Commun.	Commun.	Commun.
Activités culturelles	Oui	Oui	Non	Non
Proportion de l'échantillon dans une situation <i>au moins</i> équivalente*				
Com. Française	5.6%	33.7%	51.0%	81.1%
Com. Flamande	4.1%	31.0%	58.2%	83.8%

\* Est déclaré dans une situation au moins équivalente un individu se trouvant dans un environnement socio-économique identique ou plus favorable à la réussite que celui décrit dans le scénario

Le scénario (1) correspond clairement à une population d'élèves fortement favorisés. Seuls plus ou moins 5% des élèves peuvent s'en prévaloir. A l'autre extrême, le scénario (4) illustre une situation fortement défavorable. Seuls un peu moins de 20% des élèves sont plus mal lotis. Les scénarios (2) et (3) représentent des cas intermédiaires.

Les Figures 1 et 2 reportent, pour respectivement la Communauté Française et la Communauté Flamande, les profils de probabilités de survie des élèves dans l'état "parcours scolaire normal" calculés sur base de chacun des 4 scénarios repris au Tableau 3. Il convient d'abord de remarquer que, à cause du type de données utilisées, les profils décrits sur ces figures peuvent être trompeurs si les taux de redoublement ne sont pas stables dans le temps. On peut néanmoins raisonnablement supposer que ce problème n'a pas une grande importance pour la période considérée et que la comparaison des différents scénarios n'est pas affectée.

Figure 1: Profils de probabilités de survie dans l'état "parcours scolaire normal" en Communauté Française

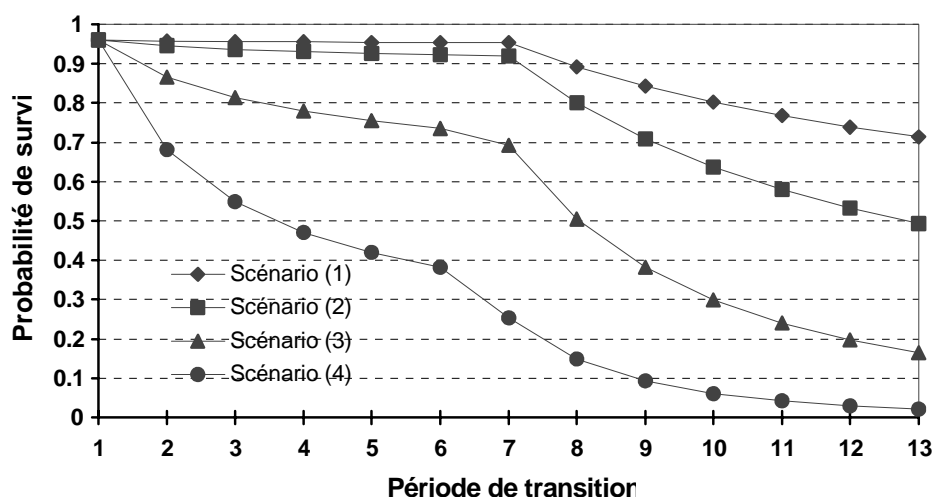
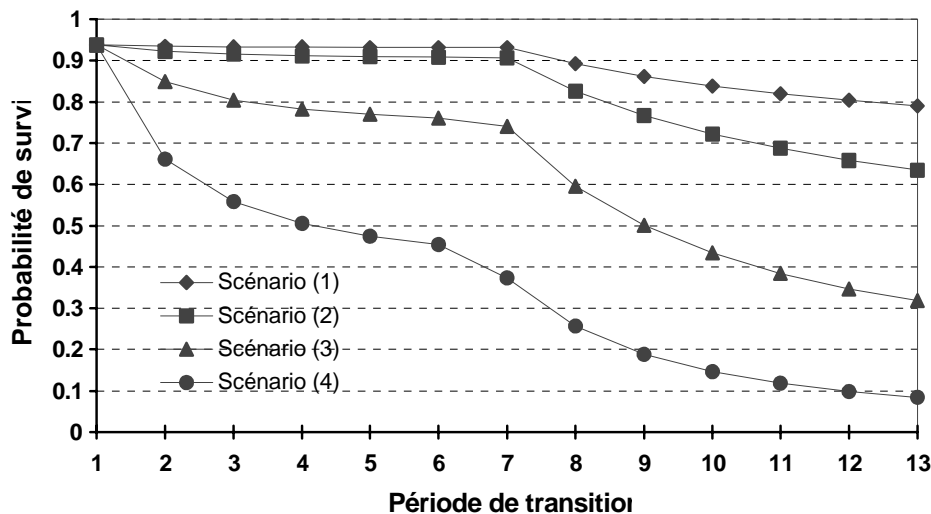


Figure 2: Profils de probabilités de survie dans l'état "parcours scolaire normal" en Communauté Flamande



Les différences observées entre les communautés linguistiques appellent quelques commentaires. On l'a vu, les statistiques agrégées indiquent des taux d'échec sensiblement plus importants en Communauté Française qu'en Communauté Flamande, en particulier dans le secondaire. Bien qu'ils ne permettent guère de faire valoir à ce niveau des différences statistiquement significatives (au seuil de 5%), nos résultats s'inscrivent résolument dans cette perspective<sup>21</sup> : presque toutes les probabilités de survie reportées à la Figure 1 sont inférieures à leur homologue de la figure 2<sup>22</sup>. De manière générale, si l'on raisonne en termes relatifs (c.-à-d. si on fait le rapport des probabilités de survie), on notera que le désavantage (resp. avantage) des élèves francophones semble d'autant plus (resp. moins) marqué que, d'une part, on avance dans la scolarité et que, d'autre part, on a affaire à des individus issus de milieux défavorisés.

Pour le reste, les Figures 1 et 2 parlent d'elles-mêmes : quelle que soit la communauté linguistique considérée, on ne peut être que frappé par les différences considérables entre les profils de probabilités de survie correspondant aux divers scénarios. Ainsi, si plus de 90% des étudiants appartenant aux 5% les plus favorisés de la population (Scénario (1)) ont toutes les chances d'effectuer un cursus primaire (jusque  $t = 7$ ) sans accroc, moins de 40% (45% en Communauté flamande) des élèves appartenant aux 20 % les plus défavorisés de la population (Scénario (4)) peuvent en dire autant. Plus impressionnant encore, à la fin de la période de scolarité obligatoire ( $t = 13$ ), si plus de 70% (resp. 80% en Communauté Flamande) des individus de la première catégorie seront nantis d'un diplôme de l'enseignement secondaire leur donnant accès à l'enseignement supérieur, seul moins de 5% (resp. 10%) des individus de la seconde catégorie pourront s'en prévaloir. Au delà de ces différences absolues, on notera également le contraste des calendriers de décrochage scolaire. Si les individus plutôt favorisés semblent y être essentiellement exposés au cours du cycle secondaire, les individus moins favorisés le sont pour leur part dès le début du

<sup>21</sup> On gardera à l'esprit que nous ne travaillons pas avec un échantillon représentatif de l'ensemble de la population, mais seulement des ménages biparentaux.

<sup>22</sup> Seules les probabilités de survie correspondant aux premières années de scolarité et aux individus les plus favorisés ont été estimées plus importantes en Communauté Française qu'en Communauté Flamande.

cycle primaire.

## 7. Conclusion

L'objet de cet article était d'évaluer empiriquement la mesure dans laquelle le parcours scolaire primaire et secondaire des enfants belges est lié à leur environnement socio-économique (éducation et statut professionnel des parents, composition et conditions matérielles du ménage, ...), et donc, en fin de compte, la mesure dans laquelle l'école tend à reproduire les inégalités sociales. Pour ce faire, un modèle économétrique basé sur un processus de Markov a été construit et estimé avec des données transversales issues d'une enquête.

En s'appuyant sur une définition de l'échec scolaire prenant en compte à la fois le redoublement et la relégation vers une filière de second rang, nos résultats montrent qu'il existe d'énormes disparités de probabilités de décrochage scolaire entre les individus issus d'horizons socio-économiques différents. En clair, derrière les substantiels taux agrégés de redoublement qui caractérisent le système scolaire belge se cachent en fait de profondes inégalités sociales.

L'ampleur des inégalités sociales constatées est plutôt surprenante, en particulier au regard de la forte décentralisation des décisions de réussite, qui sont très largement du ressort de l'enseignant et du conseil de classe. En effet, si l'on veut bien admettre, d'une part et comme le prétendent les psychopédagogues, que les enseignants ont tendance à évaluer leurs élèves par rapport au niveau moyen du groupe qu'ils ont en charge, ce qui produit des taux d'échec relativement indépendants du niveau absolu du groupe (loi dite de Posthumus), et d'autre part, que les élèves de même origine sociale ont tendance à se regrouper dans les mêmes établissements scolaires, ce système décentralisé est plutôt de nature à atténuer les inégalités sociales observées en matière de réussite (mais pas nécessairement en matière de niveau de compétence atteint).

Nos résultats n'illustrent qu'une petite partie de la problématique des inégalités sociales face au système scolaire : celle du décrochage scolaire (initial) dans l'enseignement obligatoire. Elle montre déjà un tableau très inquiétant. On peut cependant craindre qu'elle ne constitue que la partie émergée de l'iceberg. En effet, comme l'indiquent plusieurs études, les difficultés scolaires semblent présenter une tendance très nette à s'auto-alimenter<sup>23</sup>, créant des cercles vicieux. En outre, même si ils finissent par surmonter leur handicap, les enfants des milieux plutôt défavorisés sortiront généralement de filières et d'écoles moins réputées, et éprouveront probablement davantage de difficultés (financières ou autres) à accéder à l'enseignement supérieur.

Les remèdes qu'il conviendrait d'apporter à cette situation sont complexes. On l'a dit, on ne les trouvera pas dans nos résultats. Par contre, ceux-ci indiquent clairement les caractéristiques minimales que devraient satisfaire toute politique efficace de lutte contre l'échec scolaire : sélectivité (la population "à risques" est aisément identifiable) et mise en oeuvre dès le début du cycle primaire.

---

<sup>23</sup> En ce sens, voir Litt (1980) et l'IRES (1992).

## Annexe : Définition des variables exogènes

Variable	Descriptif	Fréquences observées
Niveau d'éducation du père	1 = pas de diplôme ou diplôme primaire 2 = diplôme secondaire 3 = diplôme de l'ens. supérieur	17.1% 51.4% 31.5%
Niveau d'éducation de la mère	1 = pas de diplôme 2 = diplôme primaire 3 = diplôme du secondaire inférieur 4 = diplôme du secondaire supérieur 5 = diplôme de l'ens. supérieur court 6 = diplôme de l'ens. supérieur long	6.9% 11.7% 21.1% 31.2% 16.3% 12.6%
Aisance financière	ln 1 = très difficile ln 2 = difficile ln 3 = plutôt difficile ln 4 = plutôt facile ln 5 = facile ln 6 = très facile	3.5% 5.4% 22.8% 34.9% 26.3% 7.2%
Père ouvrier	1 si le père est ouvrier 0 sinon	38.7% 61.3%
Autre langue parlée dans le ménage	1 si la langue parlée dans le ménage est différente de celle parlée à l'école 0 sinon	4.3% 95.7%
Absence d'activités culturelles	1 si aucunes activités culturelles (cinéma, exposition, théâtre, concert) au cours de la dernière année écoulée 0 sinon	24.6% 75.4%

## Références

- Alaluf M. *et al.* (1994), "Prolongation de la scolarité obligatoire: 10 ans après", Recherche effectuée à la demande du Ministre P. Mahoux, Université Libre de Bruxelles, Bruxelles.
- Alwin D.F. et A. Thornton (1984), "Family origins and the schooling process: early versus late influence of parental characteristics", *American Sociological Review*, Vol. 49, pp. 784-802.
- Blaug M. (1985), "Where are we now in the economics of education?", *Economics of Education Review*, Vol. 4(1), pp. 17-28.
- Becker G. (1964), *Human Capital*, The University of Chicago Press, Chicago and London.
- Becker G. et H.G. Lewis (1973), "On the interaction between the quantity and quality of children", *Journal of Political Economy*, Vol. 81(2), pp. S279-S288.
- Bourdieu P. (1977), "Cultural reproduction and social reproduction", dans *Power and Ideology in Education*, édité par Karabel J. et Halsey A.H., Oxford Univer-

- sity Press, Oxford.
- Boudon R. (1973), *L'inégalité des chances*, Armand Colin, Paris.
- Bowles S. (1971), "Schooling and inequality from generation to generation", *Journal of Political Economy*, Vol. 80(3), pp. S219-S251.
- Bowles S. et H. Gintis (1976), *Schooling in Capitalist America*, Basic books, New-York.
- Cohn E. et T.G. Geske (1990), *Economics of Education*, Pergamon Press, Oxford.
- Duru-Bellat M., J.-P. Jarousse et A. Mingat (1993), "Les scolarités de la maternelle au lycée: étapes et processus dans la production des inégalités sociales". *Revue Française de Sociologie*, Vol. 34, pp. 43-60.
- Eloy M. (1993), "Comparaisons internationales en matière d'enseignement", *Courrier Hebdomadaire du CRISP*, n° 1412-1413, Bruxelles.
- Eicher J-C. (1990), "Education", Ch. 36 dans *Encyclopédie Economique*, édité par Greffe X., Mairesse J. et Reiffers J-L., Economica, Paris.
- Eicher J-C. et L. Levy-Garboua (1979), *Economique de l'éducation*, Economica, Paris.
- Gintis H. (1971), "Education technology and the characteristics of worker productivity", *American Economic Review*, Vol. 61, pp. 266-279.
- Goldberger A.S. et C.F. Manski (1995), "Review article: The *Bell Curve* by Herrnstein and Murray", *Journal of Economic Literature*, Vol. 63, pp. 762-776.
- Gouriéroux C. (1989), *Econométrie des variables qualitatives*, Economica, Paris.
- Hanusek E.A. (1986), "The economics of schooling: production and efficiency in public schools", *Journal of Economic Literature*, Vol. 24, pp. 1141-1177.
- Hanusek E.A. (1987), "Educational production functions", dans *Economics of Education: Research and Studies*, édité par Psacharopoulos G., Pergamon Press, Oxford.
- Hartog J., G. Pfann et G. Ridder (1989), "(Non-)Graduation and the earning function", *European Economic Review*, Vol. 33, pp. 1373-1395.
- Herrnstein R.J. et C. Murray (1994), *The Bell Curve: Intelligence and Class Structure in American Life*, Free Press, New-York.
- IRES (1992), "Les dépenses dans l'enseignement fondamental et secondaire de la Communauté Française", Radioscopie de l'enseignement en Communauté Française de Belgique, Rapport final, Université Catholique de Louvain, Louvain-la-Neuve.
- Jensen A.R. (1969), "How much can we boost IQ and scholastic achievement?", *Harvard Educational Review*, Vol. 39, pp. 1-123.
- Lafontaine D. et A. Grisay (1988), "Réduire l'échec en début de secondaire: que suggère la recherche?", Recherche subventionnée par la Communauté Française de Belgique, Université de Liège, Liège.
- Leibowitz A. (1974), "Home investments in children", *Journal of Political Economy*, Vol. 82, pp. S111-S131.
- Leibowitz A. (1974), "Education and home production", *American Economic Review*, Vol. 64 (2), pp. 243-250.
- Litt J.L. (1981), *Origine sociale et scolarité*, Dissertation doctorale, Université Ca-

tholique de Louvain, Louvain-la-Neuve.

- Mare R.D. (1980), "Social Background and School Continuation Decisions", *Journal of the American Statistical Association*, Vol. 75, pp. 295–305
- Mehan H. (1992), "Understanding inequality in schools: the contribution of interpretive studies", *Sociology of Education*, Vol. 65, pp. 1–20.
- Morgenstern R.D. (1973), "Direct and indirect schooling effects on earning of schooling and socio-economic background", *The Review of Economics and Statistics*, Vol. 55, pp. 225–133.
- Mulkey L.M., R.L. Crain et A.J.C. Harrington (1992), "One-parent households and achievement: economic and behavioral explanations of a small effect", *Sociology of Education*, Vol. 65, pp. 48–65.
- Murnane R.J., R.A. Maynard et J.C. Ohls (1981), "Home resources and children's achievement", *The Review of Economics and Statistics*, Vol. 63, pp. 369–377.
- Willis P.E. (1977), *Learning to Labor*, Columbia University Press, New-York.