

## **Modélisation séquentielle du cheminement des étudiants en 1<sup>er</sup> cycle universitaire et test d'adéquation des conditions d'accès.**

**HALFAOUI B. & TOUATI-TLIBA M.**

**Abstract :** Cet article traite le problème de l'estimation d'une fonction de production éducative pour un établissement universitaire. En particulier, à travers un modèle logit séquentiel, le cheminement global des étudiants dans un programme de premier cycle universitaire est analysé en dégagant les caractéristiques individuelles des étudiants susceptibles d'influencer les probabilités de performance et de persévérance. Cela nous permet de tester l'adéquation des conditions d'accès. Les estimations menées sur deux cohortes d'étudiants de l'Ecole Supérieure de Commerce (ESC) confirment l'importance des variables pédagogiques (performance antérieure et origine scolaire). En effet, les bacheliers d'origine mathématique et scientifique réussissent mieux. Les littéraires, quant à eux, enregistrent les plus faibles performances. L'influence de ces variables pédagogiques diminue d'intensité lorsqu'il s'agit de la persévérance, mais restent néanmoins significatives. Pour ce qui est de l'âge de l'étudiant, c'est les plus jeunes qui sont performants, mais ce facteur n'est plus déterminant dans l'explication de la persévérance. De même, il semble que l'influence du lieu de résidence sur la persévérance devient plus importante surtout pour les femmes.

**Mots clés :** Capital humain, fonction de production éducative, performance, persévérance, modèle logit séquentiel.

### **Introduction**

Depuis les premiers travaux sur le capital humain de Schultz (1961, 1963) et de Becker (1962, 1964), la valeur économique de l'éducation n'est plus à démontrer. Dans la pensée économique moderne, l'éducation est considérée comme un investissement en capital humain à la fois pour l'individu lui-même et pour la société. Au niveau individuel l'éducation, mesurée par le nombre d'années d'études, joue un rôle central parmi toutes les variables susceptibles d'expliquer le salaire de l'individu (Mincer (1958, 1974)). Quant au niveau social, l'éducation contribue directement à la croissance du revenu national en améliorant les qualifications et les capacités productives de la main d'œuvre (Schultz (1961, 1963)). D'après la théorie de la croissance endogène de Romer (1990), nous savons maintenant que l'accumulation de capital humain contribue au même titre que l'accumulation de capital physique à la croissance économique de long terme ainsi qu'au développement économique (Benhabib & Spiegel (1994)). Ce qui confirme le caractère stratégique de l'éducation dans l'intervention gouvernementale et

explique, en partie<sup>1</sup>, la prédominance du financement public de l'éducation dans tous les pays même les plus développés. D'après les indicateurs de l'O.C.D.E pour l'année 1993, cités par Garboua (2004), les fonds publics représentent, au niveau de l'enseignement supérieure, 82.04% du financement total. La recherche en économie de l'éducation est en expansion rapide. Joanis (2002), cite deux importants constats de la recherche récente dans ce domaine, l'hétérogénéité des rendements individuels de l'éducation et le risque associé à l'investissement en capital humain. Hartog et al. (2004), quant à eux, ont distingué trois sources de risque associées à cet investissement. La première, est liée à la distribution de la performance éducative de l'individu et de sa persévérance. La seconde source d'incertitude est relative à la position individuelle, après l'obtention du diplôme, dans la distribution des gains et la dernière est liée au risque du marché. On s'intéresse dans cet article à l'analyse du risque relatif à la performance éducative ainsi qu'à la persévérance.

En Algérie, devant l'importance des échecs<sup>2</sup> qui se manifestent surtout au niveau du tronc commun à l'université, de nombreuses questions se posent sur les causes de l'inefficacité interne des institutions universitaires<sup>3</sup> (redoublements, exclusions, abandons,.....). On se demande si les échecs universitaires dépendent des caractéristiques individuelles de l'étudiant, si la sélection à l'entrée est le moyen le plus efficace pour réduire les déperditions et si les conditions d'accès en vigueur sont adéquates.

Notre travail se situe, entre autres, dans la lignée des travaux empiriques de Dagenais et Dagenais (1988), Ferroukhi (1989) et Krupa (1993). Dagenais et Dagenais traitent simultanément les phénomènes de performance et persévérance à la fin de la première année et à la fin du cycle. Ferroukhi étudie le comportement des bacheliers en première année universitaire. Alors que, Krupa analyse le problème des réussites et des

<sup>1</sup> Pour une autre explication plausible du financement public des investissements éducatifs, voir Garboua (1997).

<sup>2</sup> A titre d'exemple, à l'Ecole Supérieure de Commerce, les taux de réussite en tronc commun pour les deux années universitaires successives 1993/1994 et 1994/1995 étaient seulement de 55% et 52% respectivement, quant au taux d'exclusion pour la même période, il dépassait les 14%.

<sup>3</sup> Pour le système éducatif ou une institution éducative particulière, on distingue, généralement, deux notions d'efficacité. L'efficacité externe concerne la manière dont ont été préparé les étudiants pour accomplir leur rôle dans la société. Quant à l'efficacité interne, elle concerne la relation entre les inputs et l'output au sein du système où l'output est mesuré par rapport aux objectifs internes de l'institution (taux de réussite, notes,.....).

échecs tout au long d'un cursus. Néanmoins, quelques aspects du problème qu'on se propose d'analyser diffèrent de ceux traités dans les travaux cités ci-dessus. Dans l'étude de Dagenais et Dagenais, le schéma de cheminement et le système d'évaluation sont différents. Tout d'abord, la non persévérance concerne les étudiants absents aux examens finaux, leurs notes ne sont donc pas observables. Alors que dans notre cas, il s'agit plutôt d'exclusion par le jury de délibération après une série d'examens (rattrapage, examens après redoublement). Ferroukhi analyse seulement les résultats obtenus aux épreuves programmées au cours de la première session de l'année universitaire. Il ne tient pas compte des résultats des sessions de rattrapage. La question qui n'a pas été traitée est le cheminement global de l'étudiant dans un programme de 1<sup>ère</sup> année. De plus, son analyse ne permet de dégager que l'effet brut des variables pédagogiques. Quant au modèle séquentiel de Krupa, il prend en compte à chaque étape, les résultats scolaires des étapes antérieures (mentions obtenues) tout au long du cursus. Alors que dans une perspective de sélection, c'est uniquement les caractéristiques individuelles à l'entrée qui sont disponibles. Le problème qu'on se propose d'étudier consiste à estimer une fonction de production éducative pour un établissement universitaire. En particulier, un modèle logit séquentiel est estimé pour expliquer le cheminement global des étudiants dans un programme de premier cycle universitaire et tester l'adéquation des conditions d'accès.

La recherche sur la performance scolaire a connu une croissance rapide depuis le fameux rapport de Colman (1966) aux USA, mais la nature interdisciplinaire de ce domaine l'a rendu difficile à suivre.

Les économistes (Ben Porath (1967), Bowles (1970), Hanushek (1979),...), avec leurs manières d'exprimer, leurs terminologies et leurs outils de travail ont inauguré une nouvelle phase de travaux de recherches dans le domaine de l'éducation où l'institution éducative est assimilée à une unité de production et la relation input output est exprimée à travers une fonction de production éducative. Habituellement, une fonction de production résume le maximum d'output d'un produit homogène pour une combinaison donnée d'inputs, elle suppose donc l'efficacité technique. L'application de la notion de la fonction de production dans le secteur de l'éducation pose des problèmes liés à la nature du secteur lui-même et à l'utilisation potentielle de l'analyse.

Essentiellement, la mesure de l'output présente une difficulté particulière. Une mesure quantitative, comme dans le cas d'un secteur de production d'un produit homogène, n'est pas tout à fait appropriée car en éducation, la différence qualitative est spécialement importante. Au niveau individuel<sup>4</sup>, les

---

<sup>4</sup> Au niveau de l'établissement ou du système éducatif, l'output est généralement mesuré par le taux de réussite ou le taux de déperdition.

notes de tests cognitifs sont souvent utilisées comme une mesure du produit éducatif mais leurs validités et leurs consistances sont remises en question (voir Hanushek (1979)). Le problème ici est de savoir si ces notes reflètent réellement les acquisitions effectives.

Motivés par l'explication de la relation qui existe entre gain individuel et niveau d'éducation, plusieurs chercheurs s'interrogent sur le rôle de l'institution scolaire (école, lycée, université) et donc sur la nature même de l'output éducatif. En effet, on se demande si l'école produit des individus qualifiés avec une certaine valeur ajoutée (un ensemble de connaissances additionnelles et de qualifications supplémentaires) ou bien sert-elle simplement à identifier les individus capables. La théorie du filtre (Arrow (1973) et Stiglitz (1975)) considère que l'output de l'éducation se résume essentiellement dans l'information relative aux aptitudes individuelles. En particulier, l'attention est centrée ici sur la variabilité des résultats individuels plutôt que sur les résultats moyens. Elle implique également que la valeur sociale de l'éducation est considérablement inférieure à sa valeur privée. Malheureusement, il n'existe pas encore de tests qui permettent de distinguer parfaitement un modèle du filtre d'un modèle standard de la fonction de production (voir Layard & Psacharopoulos (1974), Blaug (1976) et Hanushek (1979)).

Dans un secteur productif, le cas des outputs multiples ne pose pas habituellement de problème car les prix du marché peuvent être utilisés pour agréger les différents outputs. Néanmoins dans le secteur de l'éducation, on peut envisager plusieurs outputs (connaissances théoriques, qualification professionnelle, valeurs morales, attitudes, ...), mais l'agrégation par les prix n'est ni possible, ni désirable (voir Hanushek, (1979)). Généralement, les analyses supposent un seul output où un modèle de régression suffit pour estimer et expliquer la relation input output, mais lorsqu'on considère plusieurs outputs de sérieuses interactions surgissent et la méthode des moindres carrés n'est plus appropriée. Cependant un système d'équations simultanées est suggéré pour représenter la fonction de production (voir Dagenais & Dagenais (1988)). Hanushek (1979) présente un modèle conceptuel général de la fonction de production éducative où la performance de l'individu au temps  $t$  est exprimée en fonction de trois types de caractéristiques, des variables liées à l'environnement familial de l'individu  $i$  cumulées au temps  $t$  (statut socio-économique de la famille, niveau d'instruction des parents, taille de la famille,.....), des variables liées à l'environnement scolaire de l'individu  $i$  cumulées au temps  $t$  (motivation et expérience du corps enseignant, disponibilité du matériel pédagogique, taille des classes, organisation de l'établissement, caractéristiques des camarades de classe,...) et enfin des caractéristiques individuelles (aptitude personnelle, âge, sexe, état de santé,....). Cette forme générale semble être acceptée par la

majorité des chercheurs, mais des questions se posent sur le problème de mesure des inputs ainsi que sur les besoins énormes du modèle en terme de données. En effet la plupart des travaux sur l'implémentation empirique du modèle semblent être guidés par la disponibilité des données et confronter à de sévères problèmes analytiques et difficultés méthodologiques.

Premièrement, l'omission d'une ou de plusieurs variables essentielles, par exemple l'habilité innée, peut entraîner un biais dans les coefficients estimés et rendre l'interprétation des résultats difficile à cerner notamment dans une perspective de prise de décision. Par exemple sachant que les variables aptitude personnelle et niveau éducatif des parents sont corrélées, si la première est omise le coefficient du niveau éducatif exprime l'impact brut sur les résultats scolaires au lieu de son impact net (voir Hanushek (1979) et Duru & Mingat (1988)). La taille du biais dépend donc du degré de la liaison statistique entre la variable omise et les variables incluses dans le modèle, mais lorsque la variable omise n'est pas corrélée avec les autres variables aucun biais n'est causé, cependant la variance du terme résiduel est plus importante.

Deuxièmement, l'erreur de mesure d'une variable explicative tend à biaiser l'estimation. En effet, fréquemment on remplace des variables non observables par des variables "proxy" c'est à dire des variables approximatives et des caractéristiques individuelles par des caractéristiques moyennes. Par exemple, l'expérience de l'enseignant de chaque élève est remplacée par l'expérience moyenne des enseignants de l'école. La taille du biais dépend du rapport entre la variance de l'erreur et la variance de la variable réelle (voir Hanushek (1979)).

Le troisième problème est lié à l'utilisation des données transversales qui ignore le fait que les phénomènes scolaires se construisent dans le temps selon un processus cumulatif (voir Duru & Mingat (1988)); lorsqu'une variable contemporaine est utilisée en ignorant la source temporelle de son influence sur les résultats scolaires on risque de ne pas identifier son impact réel.

Enfin, parmi d'autres problèmes cités par Hanushek (1979), on se contente d'ajouter le cas particulier des phénomènes de sélection. En effet, parfois, on se trouve dans une situation où on observe qu'un segment de la population due à un mécanisme de sélection. Il s'agit soit d'une sélection administrée (politique d'orientation) soit d'une auto-sélection. Certaines des variables explicatives affectent à la fois la sélection et la performance scolaire étudiée ; ce fait doit, en principe, être pris en compte dans l'analyse.

Une autre version du modèle de la fonction de production éducative présentée par Hanushek (1979), dite version de la valeur ajoutée, est justifiée par le fait que la performance scolaire, dans un cycle donné, est le résultat d'une transformation d'une situation initiale (à l'entrée) à une situation finale. Ce modèle a l'avantage de nécessiter moins de données, de réduire

sensiblement les effets négatifs de biais dus à l'omission des variables, les erreurs de mesure et le manque de variables temporelles. Par exemple si la variable aptitude individuelle est omise, elle est beaucoup moins susceptible de biaiser les estimations puisque son influence est déjà incorporée dans le résultat scolaire initial.

Plusieurs chercheurs ont traité les phénomènes de la performance et de la persévérance dans un programme universitaire et ont essayé, en particulier, d'identifier les facteurs explicatifs susceptibles d'influencer la probabilité individuelle de réussite de l'étudiant.

Krupa (1993) cite deux études qui utilisent l'analyse factorielle de correspondance multiple (A.F.C.M) sur des données de l'université française. La première étude, de Baccini, Mathieu et Mondot (1987), permet de dégager les facteurs qui expliquent la durée d'obtention du diplôme. Pour le DEUG (diplôme de premier cycle), les résultats mettent en évidence l'importance des variables liées à l'origine scolaire (série et mention du baccalauréat) ainsi qu'à l'âge d'obtention du baccalauréat. Concernant la maîtrise (diplôme de deuxième cycle), les variables significatives sont la durée d'obtention du DEUG, l'âge d'obtention du baccalauréat, la profession du père et le type de filière suivi. La seconde étude de Moinard (1989) a pour but la détermination des facteurs de réussite en première année. L'analyse fait ressortir quatre groupes d'individus. Le premier est composé d'étudiants inscrits mais qui ne se sont présentés à aucun examen. Ces individus sont généralement âgés, mariés, ont un ou plusieurs enfants et exercent une activité salariale. Le deuxième groupe concerne les étudiants qui réussissent bien, leur origine scolaire est souvent scientifique. Le troisième groupe rassemble les étudiants moyens qui sont fréquemment d'origine littéraire ou économique. Le dernier groupe concerne les étudiants qui échouent. Ils sont habituellement de nationalité étrangère, des doublants ou dispensés des contrôles continus.

Millot et Orivel (1980), quant à eux, ont estimé une fonction de production du diplôme pour douze universités françaises et ont trouvé que les variables qui favorisent la réussite sont le temps de travail personnel, le type de filière suivi et l'origine scolaire de l'étudiant.

Dans une série d'études longitudinales réalisées à l'I.R.E.D.U, Mingat (1976) et Mingat et Rasera (1981) ont utilisés des modèles logit pour examiner les déterminants de la réussite en première et en deuxième année respectivement. Les résultats de la première étude confirment l'importance de la série et de la mention du baccalauréat, de l'âge de l'étudiant et de l'établissement fréquenté avant l'entrée à l'université. Pour la réussite en deuxième année, les auteurs affirment que «globalement les facteurs liés à l'origine scolaire ont un effet qui diminue d'intensité mais qui restent néanmoins importants». Ils ajoutent «de façon beaucoup plus qu'en cours de la première année d'étude, les caractéristiques économiques des étudiants

(ressources financières des parents) ont un effet sur la réussite en deuxième année ». D'autres chercheurs se sont intéressés au problème de sélection à l'entrée du système et au choix de scolarité. Dagenais (1984) propose un modèle Probit pour la validation des critères de sélection. De même, Nieradzik (1995) évalue, à travers un modèle logit multinomial non ordonné, l'impact du risque d'échec sur les choix d'orientation réalisés par les étudiants à l'issue du baccalauréat et conclue que les probabilités individuelles de réussite dans chacune des filières étudiées semblent jouer un rôle significatif dans la prise de décision. Cannings, Montmarquette & Mahseredjian (1993), quant à eux, traitent la même question en exploitant la base de données de N.L.S.Y (National Longitudinal Survey of Labor Force behavior Youth Survey) et confirment que le choix d'une discipline à l'université dépend de la probabilité de succès prévue dans cette discipline. Néanmoins, il existe des différences significatives entre les groupes sociaux dans leur attitude envers le risque. En particulier, les étudiants de sexe masculin, issus d'une famille aisée semblent avoir plus d'aversion au risque.

Duru & Mingat (1979, 1988) présentent un modèle où l'étudiant, au moment de l'admission dans un programme universitaire, fait un arbitrage entre le rendement et le risque lors de son choix. Les résultats montrent que les étudiants titulaires d'un baccalauréat scientifique accordent plus d'importance, dans leurs choix d'orientation, au rendement alors que les autres bacheliers semblent être préoccupés par le risque.

Le travail de Dagenais et Dagenais (1988) consiste à étudier le cheminement des étudiants dans un programme universitaire du premier cycle. Le modèle proposé est constitué de deux équations et permet de traiter simultanément les phénomènes de performance et de persévérance ainsi que leur interdépendance. Il tient compte également du fait que pour les étudiants qui n'ont pas persévéré, leurs résultats scolaires ne sont pas observables. Les variables explicatives sont de deux types, celles communes aux deux équations et celles spécifiques à l'une ou à l'autre. L'estimation a été effectuée à deux moments du cheminement des étudiants ; à la fin de la première année puis à la fin du cycle<sup>5</sup>.

Pour l'équation de la performance, les variables regroupant les capacités de l'étudiant (résultats scolaires antérieurs et résultats aux tests d'admission), le profil de formation (origine scolaire scientifique ou non) et le

<sup>5</sup> Les auteurs notent l'intérêt de situer l'étude de performance en début du programme et celle de persévérance à la fin du programme car l'influence de certaines variables semblent devenir moins importantes en fin de cursus alors que pour d'autres l'influence s'accroît. De plus, parmi les étudiants les moins performants, un nombre important quitte le programme après la première année.

niveau du collège de provenance ont des coefficients positifs et statistiquement significatifs. Par contre, pour les variables "sexe" et "années d'admission", les résultats semblent ne pas être stables dans le temps. Quant à l'équation de persévérance, il est à remarquer que l'importance de l'influence de l'origine scolaire de l'étudiant et de ses capacités a sensiblement diminué mais ces variables demeurent tout de même significatives. Concernant les variables spécifiques à cette équation, on note que les étudiants moins âgés persévèrent mieux. La corrélation qui lie les termes d'erreur des deux équations est non significative et de valeur presque nulle.

Les résultats basés sur les données de fin d'études, comparés avec ceux de première année, font ressortir une diminution importante des coefficients en valeur absolue, au niveau de l'équation de la performance, cependant les variables qui étaient significatives dans l'équation de la première année le demeurent en fin d'études. Par contre, au niveau de l'équation de la persévérance, on note une augmentation de la plupart des coefficients en valeur absolue avec des *t* de Student plus élevés, ce qui confirme qu'il vaut mieux situer l'étude du phénomène de la persévérance à la fin du programme. Enfin, le coefficient de corrélation des termes d'erreur estimé est élevé.

La thèse de Krupa (1993) traite le problème des réussites et échecs tout au long d'un cursus universitaire. L'analyse porte sur plusieurs années où la réussite de l'étudiant à une étape dépend des résultats antérieurs, une méthode d'estimation séquentielle est donc retenue. Les variables explicatives proposées pour la modélisation de la réussite individuelle, sont le sexe de l'étudiant, l'âge à l'entrée, l'origine géographique, la catégorie socioprofessionnelle du père (C.S.P), la série, la mention et l'année d'obtention du baccalauréat.

Les résultats de première année, dans les trois filières étudiées, montrent que l'influence des variables C.S.P du père et sexe est faible ou absente. Quant à l'âge, il est très significatif dans les trois filières. Toutes choses étant égales par ailleurs, les jeunes réussissent mieux. Concernant l'origine scolaire de l'étudiant, les résultats montrent que les bacheliers scientifiques (C et D) réussissent mieux dans les trois filières et que la mention du baccalauréat est très significatives.

Pour ce qui est de la réussite en deuxième année, le modèle a été estimé sur la sous population des étudiants admis en deuxième année qui se sont inscrits et présentés aux examens. Il est à noter qu'une variable explicative additionnelle, la mention obtenue aux examens de première année, est introduite pour expliquer la réussite en deuxième année. Les résultats font ressortir une importante baisse (en valeur absolue) des coefficients indiquant les caractéristiques de l'étudiant à l'entrée. L'origine scolaire a encore une

influence sur la réussite individuelle comme en première année mais le phénomène est d'importance moindre.

L'hypothèse d'un lien entre la qualité des résultats obtenus en première année et la probabilité de réussite en deuxième année est vérifiée. La variable "mention obtenue aux examens de première année" est de loin la plus déterminante dans le modèle. Cela explique la diminution de l'influence des autres variables individuelles.

L'étude de Ferroukhi (1989) consiste à modéliser les comportements des bacheliers en première année universitaire et à tester l'adéquation des conditions d'accès avec les programmes du supérieur afin de proposer une politique d'orientation des bacheliers. A l'exception de la variable sexe de l'étudiant, toutes les variables retenues sont de nature pédagogique<sup>6</sup>. L'auteur propose trois modèles différents, le premier concerne le comportement global de l'étudiant (succès ou échec), le second porte sur l'analyse des comportements des nouveaux bacheliers vis à vis de chacune des matières enseignées en première année de technologie et enfin le dernier modèle traite le phénomène du comportement des redoublants. Pour le premier modèle, uniquement la série du baccalauréat est retenue comme variable explicative à un seuil inférieur à 5 %. Quant aux étudiants non admis en deuxième année, les résultats montrent que les variables mention du baccalauréat, note de mathématiques obtenue à cet examen et surtout la moyenne générale réalisée durant l'année précédente à l'université sont toutes significatives. Il est à noter que la série du baccalauréat n'a pas d'influence sur le comportement des étudiants doublants. Halfaoui (1998) a présenté une analyse du cheminement global des étudiants dans un programme de 1<sup>ère</sup> année pour dégager les caractéristiques individuelles des étudiants susceptibles d'influencer leurs performances et persévérances. Les estimations, sur une cohorte d'étudiants de l'E.S.C, confirment globalement les résultats des études déjà citées. La performance antérieure (moyenne générale obtenue au baccalauréat) et l'origine scolaire (série du baccalauréat) expliquent une part importante des probabilités de performances et de persévérances de l'étudiant. Les bacheliers d'origine mathématique et scientifique réussissent mieux, par contre les littéraires enregistrent les plus faibles performances. L'influence de ces variables pédagogiques diminue d'intensité lorsqu'il s'agit de la persévérance, mais restent néanmoins significatives. Pour ce qui est de l'âge de l'étudiant, c'est les plus jeunes qui réussissent mieux. Ce facteur devient plus déterminant pour l'explication de la persévérance.

<sup>6</sup> Les notes obtenues à l'examen du baccalauréat en mathématiques et en sciences physiques, la série du baccalauréat (mathématiques ou sciences) et la mention (baccalauréat avec mention ou sans mention).

### Les données

Cette étude vise l'analyse du cheminement des étudiants de l'École Supérieure de Commerce (ESC) durant le tronc commun (de la première année jusqu'au passage en troisième année ou l'exclusion) et porte sur le suivi de deux cohortes (coh96 et coh97) de taille totale de 731 étudiants inscrits durant les années universitaires 1995/1996 et 1996/1997 et dont les noms figurent sur la liste officielle d'orientation des bacheliers de l'année. Il convient de signaler que les étudiants inscrits en dehors de la liste officielle<sup>7</sup>, ceux qui ont procédé à un transfert ainsi que les absents aux examens finaux de la première année sont exclus de l'échantillon.

Pour chaque étudiant, on observe, à l'entrée, des caractéristiques pédagogiques telles que la moyenne et la série du baccalauréat, ainsi que des caractéristiques non pédagogiques telles que l'âge, le sexe et le lieu de résidence. On observe également, durant le séjour en premier cycle, les résultats des sessions de juin et de septembre.

Pour ce qui est des variables pédagogiques, la série du baccalauréat est susceptible d'influencer la performance à l'université pour au moins deux raisons. Premièrement, les programmes du secondaire diffèrent d'une série à l'autre. Deuxièmement, la série du baccalauréat peut être considérée comme une variable "proxy" mesurant la performance antérieure<sup>8</sup>. Les étudiants des deux cohortes étudiées sont d'origine scolaire variée. Les séries Gestion-Economie (G) et Sciences (S) sont les plus fréquentes avec des parts respectives de 49% et de 25% (53% et 33% pour la coh96). Quant aux séries Lettres (L), Comptabilité (C) et Mathématiques (M), leurs parts sont respectivement de 12%, 10,5% et 3% (5%, 5% et 3,5% pour la coh96). Le nombre de bacheliers d'origine technologique (T) est négligeable (3 étudiants).

La politique d'orientation en vigueur attribue la priorité une aux bacheliers de série G, la seconde regroupe les étudiants d'origine scolaire M, S et T, alors que les priorités trois et quatre sont réservées aux détenteurs d'un baccalauréat L et C respectivement.

La performance à la sortie du secondaire est mesurée par la moyenne générale du baccalauréat<sup>9</sup> (Moyb). Cette dernière peut fournir de précieuses informations pour l'explication de la performance universitaire. Néanmoins, elle ne représente pas le même instrument de mesure pour tous les étudiants

<sup>7</sup> Ce choix est motivé par le fait qu'on s'intéresse particulièrement au test d'adéquation de la politique d'orientation des bacheliers.

<sup>8</sup> Il est connu que les étudiants les plus performants optent déjà à l'entrée du lycée pour les séries scientifiques.

<sup>9</sup> Cette moyenne représente le critère d'orientation vers l'E.S.C pour toutes les séries à l'exception de la série comptabilité.

considérés<sup>10</sup>. Le tableau suivant indique quelques résumés statistiques de la moyenne du bac par priorité :

|            | COH96 |       |       |       | COH97 |       |       |       |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|
|            | moyb1 | moyb2 | moyb3 | moyb4 | moyb1 | moyb2 | moyb3 | moyb4 |
| N          | 220   | 153   | 20    | 21    | 138   | 56    | 67    | 56    |
| Moyenne    | 9,94  | 11,73 | 11,22 | 10,80 | 13,27 | 12,22 | 10,84 | 11,34 |
| Ecart type | 0,58  | 0,43  | 0,39  | 0,86  | 0,36  | 0,32  | 0,47  | 0,67  |
| Minimum    | 9,23  | 11,26 | 10,75 | 9,47  | 12,84 | 11,90 | 10,36 | 10,19 |
| Maximum    | 12,59 | 13,22 | 12,00 | 12,27 | 14,36 | 13,26 | 12,95 | 13,43 |

Les nouveaux bacheliers désirant s'inscrire à l'université doivent déposer une fiche de vœux exprimant l'ordre de préférence (de 1 à 15) pour les établissements d'enseignement supérieur. L'ordre de préférence de l'étudiant pour l'E.S.C (ch) est une variable qui pourrait mesurer sa motivation pour les études dans cette école.

Pour les cohortes étudiées, presque 3 étudiants sur 4 ont exprimé leur premier choix pour l'E.S.C. On note que la quasi-totalité des étudiants d'origine C, G et L ont choisi l'E.S.C en première position.

Concernant les variables extra pédagogiques, l'âge de l'étudiant pourrait avoir une influence sur la probabilité de réussite car il représente un indicateur de son cheminement scolaire antérieur. En général, les étudiants plus âgés sont souvent ceux qui ont subi du retard dans leur scolarité passée. Toutefois, l'âge peut être aussi un indicateur de la maturité de l'individu. L'âge moyen des étudiants est de 18,7 ans avec un écart type de 1,04. Il est à signaler que presque 10% des étudiants sont âgés de moins de 18 ans (dage1), alors que 20% dépassent 19 ans (dage4). A priori le sexe de l'étudiant et son succès scolaire ne sont pas liés. Néanmoins, cette variable pourrait avoir un impact significatif sur les choix de disciplines des étudiants et donc une filière peut avoir des étudiants d'un sexe donné plus au moins brillants. En ce qui concerne les étudiantes, il peut exister un barrage de type social, celles parvenant à poursuivre des études supérieures seraient soit les plus brillantes, soit celles appartenant aux milieux sociaux favorisés. La part des filles dans les cohortes considérées est de 51%. Enfin, la proportion des étudiants qui résident à Alger et ses environs est de 63%. Quant à ceux qui résident hors Alger, mis à part ceux qui habitent dans les cités universitaires caractérisées par des conditions de vie détériorées, le reste est confronté à des problèmes de transport au quotidien. Pour ces raisons, on s'attend, toutes choses étant égales par ailleurs, à une meilleure performance des étudiants habitant chez leurs parents. L'introduction de cette variable dans l'étude nous permet aussi

<sup>10</sup> Différents examens pour les différentes séries et années du baccalauréat.

d'avoir une information sur le niveau des lycées d'origine (wilaya d'Alger par rapport aux autres régions).

La politique actuelle d'évaluation permet à l'étudiant de terminer une année donnée après les premières épreuves annuelles de juin ou celles de la session de rattrapage en septembre. Les étudiants non admis après cette série d'examens peuvent redoubler. Toutefois, un étudiant n'a le droit de doubler qu'une seule fois en tronc commun. De ce fait, l'observation du cheminement d'une cohorte en premier cycle nécessite le suivi de l'étudiant pendant trois années successives.

Pour étudier la performance et la persévérance en premier cycle, on peut définir une variable ordinale, groupe de performance de l'étudiant  $i$  ( $GP_i$ ), qui prend neuf modalités selon le nombre de sessions de rattrapage et d'année de doublement subit par l'étudiant durant son séjour en tronc commun. Ainsi, les modalités 1, 2 et 3 concernent les étudiants non doublants avec 0, 1 et 2 sessions de rattrapage respectivement. De même, les modalités 4, 5 et 6 classent les doublants selon le nombre de sessions de rattrapage. Enfin, les groupes de performance 7, 8 et 9 regroupent les étudiants non persévérants et signifient respectivement les exclus en deuxième année, les exclus en première année et les abandons après un premier échec. Le tableau suivant présente la distribution des étudiants des deux cohortes par groupes de performance et par séries de baccalauréat.

| G<br>P | SERIES DU BACCALAUREAT |             |             |             |              |            | TOTAL        |
|--------|------------------------|-------------|-------------|-------------|--------------|------------|--------------|
|        | C                      | G           | L           | M           | S            | T          |              |
| 1      | 11<br>14,3%            | 72<br>20,1% | 5<br>5,7%   | 17<br>77,3% | 106<br>57,6% | 1<br>33,3% | 212<br>29,0% |
| 2      | 7<br>9,1%              | 73<br>20,4% | 3<br>3,4%   | 4<br>18,2%  | 36<br>19,6%  |            | 123<br>16,8% |
| 3      | 12<br>15,6%            | 44<br>12,3% | 6<br>6,9%   |             | 8<br>4,3%    |            | 70<br>9,6%   |
| 4      | 6<br>7,8%              | 17<br>4,7%  | 5<br>5,7%   |             | 2<br>1,1%    |            | 30<br>4,1%   |
| 5      | 7<br>9,1%              | 47<br>13,1% | 4<br>4,6%   |             | 9<br>4,9%    | 1<br>33,3% | 68<br>9,3%   |
| 6      | 4<br>5,2%              | 19<br>5,3%  | 3<br>3,4%   |             | 6<br>3,3%    |            | 32<br>4,4%   |
| 7      | 13<br>16,9%            | 25<br>7,0%  | 16<br>18,4% | 1<br>4,5%   | 6<br>3,3%    |            | 61<br>8,3%   |
| 8      | 16<br>20,8%            | 48<br>13,4% | 18<br>20,7% |             | 8<br>4,3%    | 1<br>33,3% | 91<br>12,4%  |
| 9      | 1<br>1,3%              | 13<br>3,6%  | 27<br>31,0% |             | 3<br>1,6%    |            | 44<br>6,0%   |

|       |            |             |            |            |             |           |             |
|-------|------------|-------------|------------|------------|-------------|-----------|-------------|
| TOTAL | 77<br>100% | 358<br>100% | 87<br>100% | 22<br>100% | 184<br>100% | 3<br>100% | 731<br>100% |
|-------|------------|-------------|------------|------------|-------------|-----------|-------------|

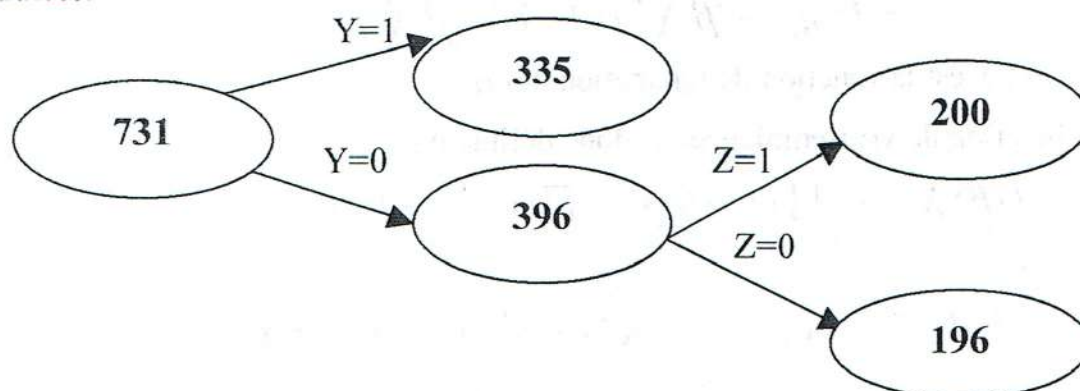
En ignorant les rattrapages, l'indice d'inefficacité interne<sup>11</sup> pour l'établissement est évalué à 1.50, ce qui indique que pour produire 100 étudiants qui passe du premier cycle vers le deuxième cycle, l'établissement doit fournir, en moyenne, 150 places pédagogiques pendant les deux années de tronc commun dont 50 places sont réservées aux doublants et exclus. Cet indice est évalué à un minimum de 1,03 pour l'audience d'origine M et à un maximum de 3,36 pour l'audience d'origine L.

### Modélisation du cheminement de l'étudiant: Modèle séquentiel

Pour la modélisation économétrique du cheminement des étudiants en premier cycle à travers un modèle logit séquentiel<sup>12</sup>, on peut introduire deux variables dichotomiques qui définissent respectivement la performance et la persévérance de l'étudiant  $i$  comme suit :

$$Y_i = \begin{cases} 1 & \text{si } GP_i \leq 2 \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad Z_i = \begin{cases} 1 & \text{si } GP_i \leq 6 \\ 0 & \text{sinon} \end{cases}$$

Le schéma ci-dessous représente le cheminement simplifié des deux cohortes étudiées:



L'estimation du modèle séquentiel se fait en deux étapes. La première concerne l'estimation de la probabilité de performance, elle se fait sur la base de l'échantillon complet (731 étudiants). La seconde concerne l'estimation de

<sup>11</sup> L'indice d'inefficacité interne est défini comme le rapport entre le nombre des places pédagogiques réellement offertes sur le nombre minimal de places pédagogiques.

<sup>12</sup> Pour plus de détails sur le modèle logit ordinaire et séquentiel, voir Maddala (1983).

la probabilité de non persévérance, ce phénomène touche uniquement les étudiants non performants durant le tronc commun. A cette étape, l'estimation se fait, donc, sur le sous échantillon des 396 étudiants restants. On associe aux deux variables (observables)  $Y_i$  et  $Z_i$  deux variables latentes (non observables)  $Y_i^*$  et  $Z_i^*$  telles que :

$$Y_i = \begin{cases} 1 & \text{si } Y_i^* > 0 \\ 0 & \text{si } Y_i^* \leq 0 \end{cases} \quad \text{et} \quad Z_i = \begin{cases} 1 & \text{si } Z_i^* > 0 \\ 0 & \text{si } Z_i^* \leq 0 \end{cases}$$

Où

$Y_i^*$  et  $Z_i^*$  peuvent s'exprimer respectivement par les deux modèles linéaires suivants :

$$Y_i^* = \beta' X^i + \mu_i \quad \text{et} \quad Z_i^* = \alpha' X^i + \varepsilon_i \quad i = 1, 2, \dots, n$$

$\beta$  et  $\alpha$  sont des vecteurs de paramètres à estimer

$X^i$  est un vecteur de caractéristiques individuelles

$\mu_i$  et  $\varepsilon_i$  sont des termes d'erreurs aléatoires

Dans un modèle séquentiel, chaque étape nécessite l'estimation par la maximisation de la fonction de vraisemblance. Pour la première étape<sup>13</sup>, on peut écrire :

$$\begin{aligned} P(Y_i = 1) &= P(Y_i^* > 0) = P(\beta' X^i + \mu_i > 0) \\ &= P(\mu_i > -\beta' X^i) = 1 - F(-\beta' X^i) \end{aligned}$$

où  $F(\cdot)$  est la fonction de répartition de  $\mu_i$

La fonction de vraisemblance est donc définie par :

$$L(\beta; X^i) = \prod_{y_i = 0} F(-\beta' X^i) \prod_{y_i = 1} (1 - F(-\beta' X^i))$$

$$L(\beta; X^i) = \prod_{i=1}^n (F(-\beta' X^i))^{(1-y_i)} (1 - F(-\beta' X^i))^{y_i}$$

On peut alors exprimer le logarithme de la fonction de vraisemblance

$$l(\beta; X^i) = \ln(L) = \sum_{i=1}^n (1 - Y_i) \ln(F(-\beta' X^i)) + \sum_{i=1}^n Y_i \ln(1 - F(-\beta' X^i))$$

La forme fonctionnelle de  $F(\cdot)$  dépend des conditions supposées sur le terme d'erreur  $\mu_i$ . On utilise soit la fonction de distribution associée à une variable

<sup>13</sup> L'estimation de la deuxième étape ne sera pas présentée car elle est identique à la première.

normale, soit une fonction logistique. Bien que les deux approches donnent des estimateurs de  $\beta$  proportionnels, c'est le modèle logit qui est le plus utilisé car il permet, entre autre, d'exprimer explicitement la fonction de répartition en une forme fermée.

On suppose ici que le terme d'erreur suit une loi logistique dont la fonction de répartition est donnée par :  $F(u) = \frac{\exp(u)}{1 + \exp(u)}$ .

On peut vérifier les relations suivantes où  $f(\cdot)$  est la densité de la loi logistique :

$$F(-u) = 1 - F(u)$$

$$f(u) = F(u)(1 - F(u))$$

$$F(u) = \frac{f(-u)}{F(-u)}$$

$$\therefore P(Y_i = 1) = 1 - F(-\beta' X^i) = F(\beta' X^i)$$

$$\text{D'où } l(\beta; X^i) = \sum_{i=1}^n (1 - Y_i) \ln(F(-\beta' X^i)) + \sum_{i=1}^n Y_i \ln(F(\beta' X^i))$$

Le système d'équation suivant représente les conditions du premier ordre

$$\frac{\partial l(\beta; X^i)}{\partial \beta} = - \sum_{i=1}^n (1 - Y_i) X^i F(\beta' X^i) + \sum_{i=1}^n Y_i X^i F(-\beta' X^i) = 0$$

$$\text{On a alors } \sum_{i=1}^n X^i F(\beta' X^i) = \sum_{i=1}^n Y_i X^i$$

Ce système d'équations est non linéaire en  $\beta$ , sa résolution peut être accomplie par une méthode itérative telle que l'algorithme de Newton-Raphson. Il possède les particularités suivantes :

$$\text{Si un terme constant est inclus dans le modèle on a : } \sum_{i=1}^n F(\beta' X^i) = \sum_{i=1}^n Y_i$$

Ce qui implique que la moyenne des probabilités estimées est égale au pourcentage observé. Si le modèle inclut une variable dichotomique, on déduit un résultat similaire pour les deux modalités.

Les conditions du deuxième ordre sont :

$$\frac{\partial^2 l(\beta; X^i)}{\partial \beta \partial \beta'} = \sum_{i=1}^n -(X^i)(X^i)' F(\beta' X^i) (1 - F(\beta' X^i))$$

La matrice  $(X^i)(X^i)'$  est définie positive, par conséquent la matrice  $\frac{\partial^2 l(\beta; X^i)}{\partial \beta \partial \beta'}$  est définie négative. On conclut que la fonction de vraisemblance est concave. L'estimateur de maximum de vraisemblance de  $\beta$  suit asymptotiquement une loi normale multi variée de moyenne  $\beta$  et de matrice de variance covariance  $(I(\beta))^{-1}$  où  $I(\beta) = E\left(\frac{\partial^2 l(\beta; X^i)}{\partial \beta \partial \beta'}\right)$ .

Pour estimer les probabilités de performance et de non persévérance à l'aide d'un modèle logit séquentiel, on écrit :

$$P(Y_i = 1) = F(\beta'X^i)$$

$$P(Y_i = 0, Z_i = 1) = F(\alpha'X^i)F(-\beta'X^i)$$

$$P(Z_i = 0) = F(-\alpha'X^i)F(-\beta'X^i)$$

Comme dans le modèle logit ordinaire, pour le modèle séquentiel l'impact d'une variable explicative sur les probabilités individuelles d'appartenir à un groupe donnée n'est pas linéaire en  $\beta$  et  $\alpha$  mais dépend des caractéristiques individuelles. Pour déterminer cet impact on écrit :

$$\frac{\partial P(Y_i = 1)}{\partial X_k^i} = \beta_k F(\beta'X^i)F(-\beta'X^i)$$

$$\frac{\partial P(Z_i = 0)}{\partial X_k^i} = -F(-\beta'X^i)F(-\alpha'X^i) [\beta_k F(\beta'X^i) + \alpha_k F(\alpha'X^i)]$$

La qualité d'ajustement d'un modèle peut être testée par le rapport du maximum de vraisemblance :

$$W = -2\ln(L_0/L) \sim \chi_d^2 \quad \text{où :}$$

$L_0$  = La valeur maximale de la vraisemblance sous l'hypothèse que les coefficients des variables sont nuls.

$L$  = La valeur maximale de la vraisemblance du modèle

$d$  = Nombre de variables du modèle

Lorsque  $W > \chi_d^2(1-\alpha)$ , le modèle est significatif à un seuil  $\alpha$ .

La qualité d'ajustement d'un modèle à variables qualitatives est souvent mesurée par le coefficient de MacFadden (MFC) :

$$MFC = \left[ \sum_{m=1}^M \left( (N_{mm}/N) - (N_m/N)^2 \right) \right] / \left[ 1 - \sum_{m=1}^M (N_m/N)^2 \right]$$

où :

$N$  = Nombre total d'observations

$N_{mm}$  = Nombre d'individus correctement prédits par le modèle pour le groupe  $m$ .

$N_m$  = Nombre d'individus pour lesquels le groupe est effectivement  $m$ .

### Estimation et interprétation des résultats

Pour le modèle logit séquentiel, la première étape concerne l'estimation, sur la base de l'échantillon complet (731 étudiants), de la probabilité de performance (probabilité de terminer le tronc commun sans doublement avec au plus une session de rattrapage). A la seconde étape, l'estimation de la probabilité de non persévérance (probabilité d'exclusion du programme ou d'abandon après échec) touche uniquement les étudiants non performants durant le tronc commun (le sous échantillon des 396 étudiants restants). Les principaux résultats du modèle pour les deux étapes sont regroupés dans les tableaux n°1 et n°2.

Tableau n°1 : Estimation de la probabilité de performance

| VAR   | COEF   | Ecart Type | Z     | $\alpha$ |
|-------|--------|------------|-------|----------|
| Coh96 | 1,323  | 0,284      | 4,66  | 0,000    |
| Dage1 | 1,212  | 0,341      | 3,55  | 0,000    |
| Dage4 | -0,561 | 0,228      | -2,46 | 0,014    |
| Dfha  | -0,888 | 0,213      | -4,16 | 0,000    |
| Mbac  | 0,488  | 0,101      | 4,83  | 0,000    |
| Dpl   | -1,339 | 0,402      | -3,33 | 0,001    |
| Dpm   | 3,065  | 1,040      | 2,95  | 0,003    |
| Dps   | 1,245  | 0,227      | 5,47  | 0,000    |
| Const | -6,527 | 1,272      | -5,13 | 0,000    |

Le modèle converge après 5 itérations, Log-vraisemblance = -387.886  
 $\chi^2(8) = 232.51$ ,  $\alpha_m \leq 10^{-4}$ .

Pour ce qui est de la performance, comme le confirment les travaux déjà réalisés dans ce domaine, on note l'importance des variables liées directement à la performance antérieure de l'étudiant (Mbac) et à l'origine scolaire (séries du baccalauréat). En effet, par rapport aux variables

dichotomiques témoins représentant les séries G et C les coefficients qui affectent les séries M et S sont positifs, par contre celui de la série L est négatif. Ces coefficients sont tous significatifs.

On peut donc affirmer, toutes choses étant égales par ailleurs, que la probabilité de performance d'un étudiant d'origine M est plus grande que celle d'un étudiant d'origine S qui est elle même supérieure à celle d'un étudiant d'origine scolaire G ou C alors que les littéraires enregistrent les plus faibles probabilités de performance.

Les résultats montrent que le lieu d'habitation est non significatif, cependant les femmes habitant loin de la capitale ou de ses environs immédiats semblent avoir des probabilités inférieures aux autres. On constate également une meilleure performance des étudiants âgés de moins de 18 ans, par contre la performance des plus âgés (plus de 19 ans) est plus faible.

Quant à l'estimation des probabilités individuelles de non persévérance des étudiants, les résultats du modèle logit estimé dans la deuxième étape sont indiqués dans les tableaux n°2.

Tableau n°2 : Estimation de la probabilité de non persévérance.

| VAR     | COEF   | Ecart Type | Z     | $\alpha$ |
|---------|--------|------------|-------|----------|
| Lieu    | 0,611  | 0,261      | 2,34  | 0,019    |
| Dfha    | -0,834 | 0,281      | -2,97 | 0,003    |
| Mbac    | 0,341  | 0,130      | 2,62  | 0,009    |
| Mbacg96 | 0,074  | 0,039      | 1,90  | 0,057    |
| Dpl     | -1,175 | 0,348      | -3,37 | 0,001    |
| Const   | -3,935 | 1,584      | -2,48 | 0,013    |

Le modèle converge après 3 itérations, Log-vraisemblance = -250.489  
 $\chi^2(5) = 47.95$ ,  $\alpha_m \leq 10^{-4}$ .

A ce niveau, l'importance des variables pédagogiques diminue légèrement par rapport à l'estimation de l'étape précédente. En effet, le coefficient qui affecte la moyenne générale du baccalauréat (Mbac) enregistre une faible baisse mais reste tout de même significatif, alors que la variable dichotomique Dpl (série lettres) demeure très significative. Néanmoins, il semble que pour la persévérance, l'influence du lieu de résidence devient plus importante surtout pour les femmes, la variable Dfha (variable représentant les femmes habitant hors Alger) reste significative. Quant au facteur âge, il semble qu'il n'est plus déterminant dans l'explication de la persévérance. Il est utile de souligner que les variables<sup>14</sup> "Coh96" et "Mbacg96" sont introduites dans le

<sup>14</sup> Coh96 est une variable dichotomique, par contre Mbacg96 est la moyenne d'un bachelier de série G de la cohorte 96 et 0 sinon.

modèle pour corriger les effets d'instabilité des résultats des examens du baccalauréat<sup>15</sup> et ceux de l'ESC.

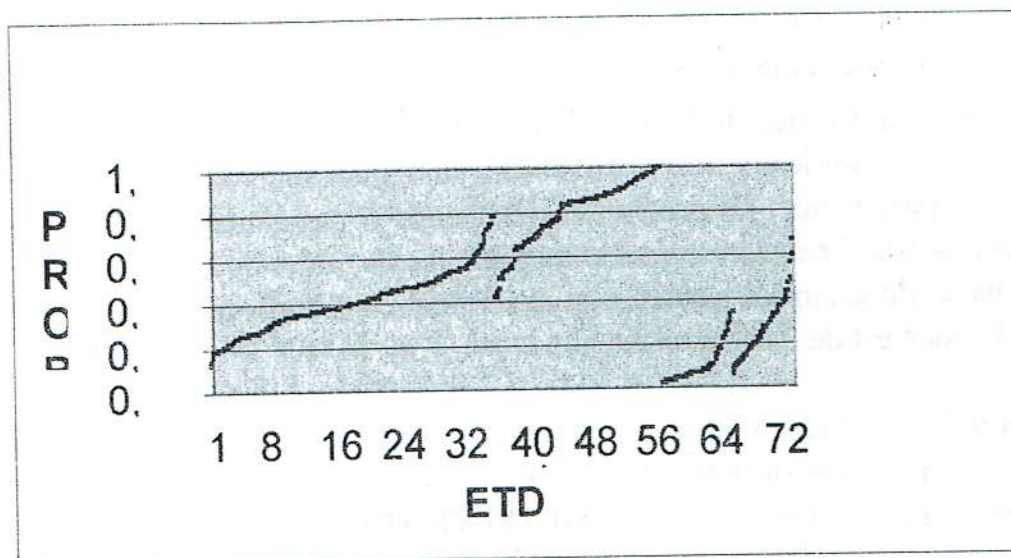
En ce qui concerne le taux de bonne classification globale du modèle séquentiel (voir tableaux n°3), 58% des étudiants sont classés dans leur groupe d'appartenance. En particulier, 63% et 55% des étudiants appartenant aux groupes 1 et 2 respectivement sont bien classés, alors que 52% seulement des étudiants du groupe 3 sont reconnus comme tels. En résumé, le coefficient de MacFadden est de 34,2%.

Tableau n°3

| GP<br>REEL | GP PREVISION |             |            |            |
|------------|--------------|-------------|------------|------------|
|            | 1            | 2           | 3          | TOTAL      |
| 1          | 210 (63%)    | 95 (28%)    | 30 (9%)    | 335 (100%) |
| 2          | 37 (18,5%)   | 110 (55%)   | 53 (26,5%) | 200 (100%) |
| 3          | 26 (13%)     | 68 (35%)    | 102 (52%)  | 196 (100%) |
| TOTAL      | 273 (37,5%)  | 273 (37,5%) | 185 (25%)  | 731 (100%) |

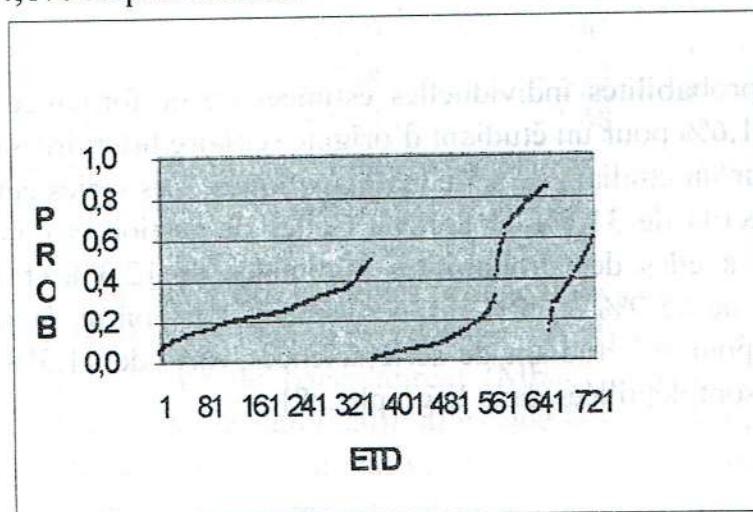
Les probabilités individuelles estimées de performance varient d'un minimum de 1,6% pour un étudiant d'origine scolaire littéraire à un maximum de 99,3% pour un étudiant de série mathématiques. Ces séries enregistrent des étendus respectifs de 31,8% et de 8%. Celles de gestion et comptabilité ont connu, quant à elles des probabilités minimales de 12,3% et 5% avec des grands écarts de 65,9% et 60% respectivement. Par contre, la probabilité de performance pour un étudiant de série sciences varie de 41,3% à 97% ; ces informations sont détaillées dans le graphe n°1.

<sup>15</sup> Par exemple, la différence entre les minimums des moyennes des bacheliers de série G des cohortes 97 et 96 est supérieure à 3 points.



Graphe n°1 : Probabilités de performance individuelles estimées pour les séries Gestion, Mathématiques & Sciences, Lettres et Comptabilité respectivement.

Les probabilités individuelles estimées de non persévérance qui figurent dans le graphe n° 2 montrent que les plus grandes probabilités (83,5% et 62,1%) sont enregistrées pour les étudiants de série littéraires et comptabilité suivie des gestionnaires avec une valeur de 47,9%. De plus, on note que ces séries ont des écarts respectifs de 65,9%, 48,9% et 44,2%, quant aux étudiants de séries sciences et mathématiques leurs probabilités varient de 0,8% à 30,7% et de 0,2% à 4,1% respectivement.



Graphe n°2: Probabilités de non persévérance individuelles estimées pour les séries Gestion, Mathématiques & Sciences, Lettres et Comptabilité respectivement.

Comme nous l'avons déjà mentionné, pour les modèles à variables qualitatives, l'impact d'une variable explicative sur les probabilités individuelles d'appartenir à un groupe donné n'est pas linéaire mais dépend

des caractéristiques individuelles. Pour déterminer l'impact des caractéristiques pédagogiques, on considère un étudiant âgé de 19 ans, de sexe masculin et résidant à Alger. Dans ce cas, à titre d'exemple, la variation de la variable M<sub>bac</sub> de 12 à 13 augmente la probabilité de performance de 64% à 74.3% pour un bachelier d'origine S et de 91.6% à 94.7% pour un bachelier d'origine M. De même, la diminution de la probabilité de non persévérance varie de 52.9% à 42.4% pour un étudiant littéraire et de 21% à 13.5% pour un comptable.

Dans une perspective d'orientation<sup>16</sup>, nos résultats peuvent être utiles. Si le décideur tolère, par exemple, une probabilité de déperdition maximale de 10%, la moyenne du baccalauréat minimale exigée à l'entrée doit être de 13.6 pour les séries C et G et de 12.24 pour les bacheliers d'origine S. Par contre, cette politique permet l'inscription de tous les étudiants d'origine M et refuse l'accès à la quasi-totalité des littéraires (La probabilité de déperdition d'un bachelier d'origine M ayant une moyenne de 10 est estimée à 9.3%. Quant à la note minimale exigée pour les étudiants d'origine L, elle doit être de 16.69).

## Conclusion

Cette conclusion présente un résumé global des résultats essentiels des estimations effectuées, des recommandations pour une meilleure sélection des bacheliers ainsi que les limites de ce travail.

Le but de cette étude est d'analyser le cheminement global des étudiants dans un programme de premier cycle universitaire et de déterminer les caractéristiques individuelles des étudiants susceptibles d'influencer leurs performances et leurs persévérances.

Les résultats des estimations menées sur deux cohortes d'étudiants de l'ESC confirment globalement les résultats des études déjà citées dans la revue de la littérature.

La performance antérieure (moyenne générale obtenue au baccalauréat) et l'origine scolaire (série du baccalauréat) expliquent une part importante des résultats scolaires de l'étudiant ainsi que sa probabilité de performance. Les bacheliers d'origine mathématique et scientifique réussissent mieux. Par contre les littéraires enregistrent les plus faibles performances. L'influence de ces variables pédagogiques diminue d'intensité lorsqu'il s'agit de la persévérance, mais restent néanmoins significatives.

Pour ce qui est de l'âge de l'étudiant, c'est les plus jeunes qui sont performants, mais ce facteur n'est plus déterminant dans l'explication de la

---

<sup>16</sup> Pour les calculs de risque de déperdition dans ce paragraphe, on suppose un vecteur de caractéristiques identique à celui du paragraphe précédent.

persévérance. Il semble, également, que l'influence du lieu de résidence sur la persévérance, devient plus importante surtout pour les femmes.

Si on suppose qu'une politique d'orientation des bacheliers vise l'amélioration du rendement interne des établissements universitaires, nos résultats confirment l'importance des variables pédagogiques actuellement utilisées dans le système d'orientation. Néanmoins, les critères peuvent être améliorés, du moins, pour l'ESC. Les résultats montrent que la moyenne générale peut servir comme un critère de classement pour les étudiants titulaires d'un baccalauréat de même série. Pour ce qui est de la priorité par série, la politique actuelle semble ne pas être adéquate. Tout d'abord, les détenteurs d'un baccalauréat de séries M et S ont la même priorité alors que les premiers sont plus performants. La première priorité accordée aux étudiants de série Gestion-Economie n'est pas justifiée par leurs résultats. Quant aux littéraires qui enregistrent les plus faibles performances, ils sont prioritaires par rapport aux comptables.

Pour pouvoir recommander une politique d'orientation adéquate, il nous paraît nécessaire de disposer de plus d'informations. Bien que notre modèle prend la forme de la version de valeur ajoutée d'une fonction de production, ce qui réduit l'ampleur des biais causés par les variables omises (on dispose d'aucune variable du milieu socio-économique de l'étudiant), l'introduction d'autres variables explicatives facilitent d'avantage la détermination de l'impact net des variables pédagogiques nécessaire pour une politique d'orientation efficace. Toutefois, quelques indications peuvent être tirées de ce travail. On propose, donc, l'allègement des conditions d'accès pour les bacheliers d'origines scolaires scientifiques surtout la série mathématiques, la diminution du quota accordé aux bacheliers de série G et l'exigence des conditions plus sévères pour les comptables et surtout pour les littéraires.

Les limites de ce travail sont nombreuses, on se contente d'en énumérer les plus importantes. Tout d'abord, on cite le problème de variables importantes manquantes liées à l'environnement familiale (niveau d'instruction des parents, leur C.S.P.,....) et l'environnement scolaire individuel. Le second problème concerne le phénomène de sélection administrée, certaines des variables explicatives affectent à la fois la sélection et /où le choix de scolarité et la performance scolaire étudiée.

## Bibliographie

- Arrow K. (1973), "Higher Education as a Filter", *Journal of Public Economics*.
- Becker G. (1962), "Investment in Human Capital", *Journal of Political Economy*.
- Benhabib J. & Spiegel M. (1994) "The Role of Human Capital in economic development Evidence from aggregate cross-country data", *Journal of Monetary Economics*.
- Ben Porath Y. (1967) "The Production of Human Capital and the Life Cycle of Earning", *Journal of Political Economy*.
- Blaug M. (1970), "Economics of Education", *Penguin Books*.
- Blaug M. (1976), "The Empirical Statues of Human Capital Theory: A Slightly Jaundiced Survey", *Journal of Economic Literature*.
- Bowles B. (1970), "Toward an Education Production Function", *National Bureau of Economic Research*.
- Cannings K., Montmarquette C. & Mahseredjian S. (1993), "Major Choices, undergraduate Concentrations and the Probability of Graduation", *CNRS/IREDU*.
- Dagenais L. (1984), "The Use of Probit Model for the Validation of Selection Procedures", *Educational and Psychological Measurement*.
- Dagenais L. & Dagenais G. (1988), "Modèle d'Analyse du Cheminement des Etudiants dans un Programme du Premier Cycle", *Annales d'économie et de statistique*.
- Duru-Bellat M. & Mingat A. (1979), "Comportement des bacheliers: Modèle de choix de disciplines", *Consommation*.
- Duru-Bellat M. & Mingat A. (1988), "Les disparités de carrières individuelles à l'université: une dialectique de la sélection et de l'autosélection", *L'année Sociologique*.
- Duru-Bellat M. & Mingat A. (1988), "Analyse de la Genèse Temporelle des Trajectoires Scolaires", *CNRS/IREDU*.
- Ferroukhi D. (1989), "Modélisation des Comportements des Bacheliers en Première Année de l'Université et Tests d'Adéquation des Conditions d'Accès avec les Programmes du Supérieur", *Collections statistiques, O.N.S.*
- Levy-Garboua L. (1997), "Le Financement Public Des Investissements Educatifs", *Options politiques*.
- Halfaoui B. (1998), "Etude de la performance et de la persévérance dans un programme universitaire de 1<sup>ère</sup> année", *Mémoire de Magister, INPS, ALGER*.
- Hanushek E. (1979), "Conceptual and Empirical Issues in the Estimation of Educational Production Functions", *Journal of Human Resources*.
- Hanushek E. (1986), "The Economics of Schooling: Production and Efficiency in Public Schools", *Journal of Economic Literature*.

- Hartog J. (2000)**, "Human Capital as an Instrument of Analysis for the Economics of Education", *European Journal of Education*.
- Hartog J., Van Ophem H. & Bajdechi S. M. (2004)**, "How Risky is Investment in Human Capital", *Tinbergen Institute Discussion Paper, University of Amsterdam*.
- Joanis M. (2002)**, "L'Economie de l'Education: Méthodologies, Constats et Leçons", *Scientific Series, CIRANO, Montréal*.
- Krupa S. (1993)**, "Réussites et Echecs à l'Université : Analyse Théorique et Empirique", *thèse de doctorat, Université des Sciences Sociales de Toulouse*.
- Layard R. & Psacharopoulos G. (1974)**, "The screening Hypotheses and the Returns to Education", *the Journal of Political Economy*.
- Maddala G. S. (1983)**, "Limited Dependent and Qualitative Variables in Econometrics", *Cambridge University Press*.
- Millot B. & Orivel F. (1980)**, « L'économie de l'enseignement supérieur », *Cuja, Paris*.
- Mincer J. (1958)**, "Investment in Human Capital and Personnel Income Distribution", *Journal of Political Economy*.
- Mincer J. (1974)**, 'Schooling, Experience and Earnings', *NBER, Columbia University Press*.
- Mingat A. (1976)**, « Enquête longitudinale 2. La première année d'études, la réussite, l'abandon, l'échec » I.R.E.D.U, Dijon.
- Mingat A. & Rasera (1981)**, « Enquête longitudinale 3. Après la première année d'études, Réorientation, réussite, Scolarité » I.R.E.D.U, Dijon.
- Montmarquette C. & Mahseredjian S. (1989)**, "Could Teacher Grading Practices Account for Unexplained Variation in School Achievements?", *Economics of Education Review*.
- Nieradzik C. (1995)**, "Probabilité Individuelle de Réussite et Choix de Scolarité : une Etude Empirique", *C.R.E.S.E.P.-Université d'Orléans*.
- Psacharopoulos G. (1985)**, "Returns to Education : A Further International Update and Implications", *Journal of Human Resources*.
- Psacharopoulos G. (1994)**, "Returns to Investment in Education : A Global Update", *World Development*.
- Psacharopoulos G. & Woodhall M. (1985)**, "Education for Development: an Analysis of Investment Choices", *Oxford University Press*.
- Romer P. (1990)**, "Human Capital and Growth: theory and evidence", *WCarnegie Rochester Conference Series on Public Policy*.
- Schultz T.W. (1961)**, 'Investment in Human Capital', *American Economic Review*
- Schultz T.W. (1963)**, "The Economic Value of Education", *New York Columbia University Press*.
- Stiglitz J. (1975)**, "The Theory of Screening, Education and the Distribution of Income", *American Economic Review*.