

A1.1
9
892

Étude économétrique sur le décrochage scolaire en milieu défavorisé

Simon Landry

Université de Montréal

Directeur :

Claude Montmarquette

Université de Montréal

UNIVERSITÉ DE MONTRÉAL
BIBLIOTHÈQUE
1205, AVENUE J. B. LAPOSTOLLE
MONTRÉAL, QUÉBEC H3T 1J4
TEL: 514 343-7300
FAX: 514 343-7301
WWW.UMONTREAL.CA

Août 2002

Résumé

Ce travail étudie la problématique du décrochage scolaire et de ses déterminants socio-économiques. À l'aide d'une banque de données complexe et unique cette étude cherche à analyser l'apport au modèle traditionnel, qui s'intéresse à la question, de variables provenant de mesures précises telles l'attitude face au risque, le niveau de délinquance, le type d'encadrement scolaire, etc. Ces données proviennent d'une étude longitudinale de grande ampleur qui s'est étalée sur une période de près de vingt ans auprès de garçons montréalais francophones issus d'un milieu défavorisé. Sous les hypothèses d'exogénéité, de données manquantes aléatoires et d'absence de corrélations potentielles nous estimons par forme réduite et par modèle probit la relation entre la probabilité de décrocher et différents facteurs provenant des données disponibles et pertinentes. Nous remarquons ainsi que le niveau de délinquance, la décision de travailler pendant les études, un faible classement en mathématiques et en français et une structure familiale monoparentale influencent positivement la probabilité de décrocher. Nous constatons également que les étudiants ayant des difficultés d'apprentissage ou autres et dont la mère a eu son premier enfant en bas âge sont plus susceptibles d'être sujets au décrochage. Cependant, contrairement aux attentes théoriques, nous observons que les variables représentant la scolarité de la mère et l'attitude face au risque ne jouent pas de rôle significatif. Par ailleurs, il serait pertinent de réaliser une analyse économique expérimentale ultérieure auprès des individus de l'échantillon, afin de raffiner certaines mesures reliées à l'attitude face au risque et au niveau d'impatience.

Table des matières

RESUME	II
LISTE DES TABLEAUX	V
INTRODUCTION	1
CHAPITRE 1	3
1.1 CADRE THEORIQUE GENERAL	3
1.2 L'APPROCHE INDIVIDUELLE	4
1.3 CADRE STATISTIQUE	5
1.4 REVUE DES RESULTATS IMPORTANTS	6
1.4.1 CRITERES DE SELECTION	6
1.4.2 IMPACT DES RESSOURCES FAMILIALES	8
1.4.2.1 La scolarité des parents	8
1.4.2.2 Le revenu familial	8
1.4.3 IMPACT DE LA STRUCTURE FAMILIALE	9
1.4.4 IMPACT DES AUTRES FACTEURS QUALITATIFS	10
1.4.4.1 Facteurs observables	10
1.4.4.2 Facteurs non-observables	10
1.4.5 IMPACT DU TRAVAIL/ETUDE	12
1.5 INTÉGRATION DU RISQUE	14
CHAPITRE 2	16
2.1 LES DONNEES	16
2.2 SPECIFICATION ET CONSTRUCTION DES VARIABLES	17
2.2.1 VARIABLE DEPENDANTE	17

2.2.2 VARIABLES DE TYPE FAMILIAL	17
2.2.3 VARIABLE DE TYPE INDIVIDUEL	18
2.2.4 VARIABLES OMISES	21
2.3 STATISTIQUES DESCRIPTIVES	22
2.4 METHODOLOGIE EMPIRIQUE	24
2.4.1 ENDOGENEITE VERSUS EXOGENEITE	24
2.4.2 CORRELATION	24
2.4.3 OBSERVATIONS MANQUANTES	25
2.5 MODELES	27
CHAPITRE 3	29
<hr/>	
3.1 RÉSULTATS EMPIRIQUES	29
3.2 ANALYSE EMPIRIQUE	30
3.3 CONCLUSION	36
BIBLIOGRAPHIE	39
<hr/>	
ANNEXE	41
<hr/>	

Liste des tableaux

<i>Tableau 1 - Résumé des articles importants</i>	6
<i>Tableau 2 - Statistiques descriptives : caractéristiques familiales</i>	22
<i>Tableau 3 - Statistiques descriptives : caractéristiques individuelles</i>	23
<i>Tableau 4 - Résultats empiriques</i>	29
<i>Tableau 5 - Distribution de la variable GAMBLING</i>	41
<i>Tableau 6 - Distribution de la variable QASTOT</i>	41
<i>Tableau 7 - Distribution de la variable CLASTOT</i>	42
<i>Tableau 8 - Distribution de la variable NBENF84</i>	42
<i>Tableau 9 - Distribution de la variable SCOLMERE</i>	43
<i>Tableau 10 - Distribution de la variable SCOLPERE</i>	43
<i>Tableau 11 - Distribution de la variable AMEREPRE</i>	44
<i>Tableau 12 - Distribution de la variable APEREPRE</i>	44
<i>Tableau 13 - Fréquences de la variable CLASTOT</i>	45

Introduction

Malgré le fait que l'éducation soit un investissement profitable selon plusieurs études qui en ont calculé le rendement¹ et abordable compte tenu de la mise en place de plusieurs politiques pour faciliter l'accès à l'éducation, un bon nombre de jeunes étudiants n'atteignent pas les études post-secondaires. Ce manque de valorisation pour les études post-secondaires se fait également sentir pour les études secondaires elles-mêmes. En effet, selon le ministère de l'Éducation du Québec, le problème du décrochage scolaire touche environ un étudiant sur cinq au Québec. Un étudiant est considéré comme étant un décrocheur s'il a quitté – momentanément ou définitivement – le système scolaire sans avoir obtenu de diplôme du secondaire, et ce, même si l'interruption n'est pas nécessairement définitive.

Le décrochage devient un problème social lorsqu'il provient d'un problème d'information incomplète ou lorsqu'il reflète l'inégalité d'opportunité d'accès. Toujours selon Éducation Québec, de récentes données indiquent que le nombre de jeunes qui quittent l'école sans avoir obtenu leur diplôme d'études secondaires peut atteindre même près de 75% dans certains établissements de la commission scolaire de Montréal². Les écoles concernées sont généralement situées dans des milieux défavorisés où le nombre d'élèves en difficulté est considérable. La Commission scolaire de Montréal compte sur son territoire les écoles présentant les indices de défavorisation les plus élevés de la province. L'impact d'un tel taux de décrochage peut engendrer des coûts sociaux qui peuvent se faire sentir aux niveaux d'une hausse du taux de chômage ou de la criminalité et d'une baisse des revenus d'impôt.

Plusieurs raisons sont généralement évoquées pour tenter d'expliquer le phénomène du décrochage : les décrocheurs accordent une grande valeur au loisir et au présent, ont un faible niveau d'habileté et de motivation, sont confrontés à de bonnes opportunités de marché (avantage comparatif pour le travail non-qualifié), ont des anticipations basses par rapport au rendement de l'investissement ou ont des problèmes familiaux ou personnels, etc. Dans un contexte d'analyse de cette problématique, la principale difficulté demeure celle d'isoler le poids de chacun de ces facteurs.

¹ Chen, 2002.

² C'est le cas notamment à l'école Pierre-Dupuy, où 73,90% des élèves ayant quitté pour de bon l'établissement en 1999-2000 n'avaient pas terminé leur formation. Les écoles Louis-Joseph-Papineau, Chomedey-de-Maisonneuve, Saint-Henri et Père-Marquette suivent avec des taux respectifs de 69,4%, 60,6%, 59,4% et 59,3%.

Ce travail tente d'apporter un éclairage nouveau au problème qui est souvent expliqué, au sein de la littérature sur le sujet, en grande partie par le revenu familial et par le niveau de scolarité parental en testant, grâce à une banque de données complexe et détaillée, l'apport de l'ajout de variables plus précises portant sur les problèmes de comportement, l'attitude face au jeu, le type d'encadrement nécessaire pour l'étudiant, etc. Ainsi, la présente étude s'intéresse aux déterminants du décrochage scolaire au secondaire dans le cadre d'une approche orientée vers les milieux défavorisés pour lesquels les problèmes reliés à l'abandon scolaire sont relativement importants et l'acquisition de capital humain peut représenter une issue importante.

Ces données proviennent d'une banque construite à partir d'une étude longitudinale qui fut l'une des premières au monde et qui demeure parmi les plus importantes du genre. Cette dernière est dirigée par le professeur de renommée internationale Richard Tremblay, du département de psychologie et de psychiatrie de l'Université de Montréal, qui travail depuis plus de vingt ans sur la délinquance. La richesse d'une telle banque de données rend l'analyse économique très intéressante.

Ce travail est d'abord motivé par la volonté de vérifier dans quelle mesure l'origine socio-économique, et particulièrement les variables à caractère économique telles l'attitude vis-à-vis le risque et l'attitude face au temps, sont bien mesurées à travers l'échantillon. Cette première étape permettra ensuite d'entreprendre une démarche économique expérimentale auprès des individus de l'échantillon pour lesquels les données sont relativement complètes et fiables. Une telle approche rendra possible une analyse plus juste de la distinction entre les impacts de l'origine socio-économique versus les incitatifs courants dans le processus d'acquisition de capital humain.

Ce travail sera présenté comme suit : d'abord le chapitre 1 fait un survol de l'état de la question et des contributions théoriques, ensuite le chapitre 2 présente les données et les variables utilisées, la méthodologie empirique ainsi que les modèles économétriques servant à établir nos estimations et finalement les résultats de régression, l'analyse empirique et la conclusion se retrouvent au chapitre 3.

Chapitre 1

1.1 Cadre théorique général

Un grand nombre de chercheurs en science sociale ont exploré les explications du processus d'épanouissement des jeunes adultes. Par épanouissement on entend « le degré de succès ou d'échec » qui se mesure entre autre par le niveau de scolarité atteint, le niveau de salaire et la situation sociale générale. Pour la plupart des disciplines, l'approche standard met surtout l'emphase sur le contexte familial et individuel pour expliquer le processus d'épanouissement des enfants. Cependant, une approche plus générale consiste à garder à l'esprit l'apport de trois facteurs primaires :

- Les choix faits par la société (ou le gouvernement) qui déterminent les différentes opportunités disponibles pour les parents et pour leurs enfants.
- Les choix faits par les parents reliés à la quantité et à la qualité des ressources accordées aux enfants.
- Les choix faits par les enfants compte tenu des investissements et des opportunités qui leurs sont disponibles.

Pour chaque palier, les choix sont faits en fonction de la maximisation de leurs objectifs propres, compte tenu de certaines contraintes politiques, de ressources familiales ou individuelles. Le gouvernement agit en premier lieu en mettant en place une structure d'investissement pour les enfants et cet environnement économique permet en second lieu aux parents d'intervenir, selon leurs propres intérêts et ceux de leurs enfants. L'enfant lui-même intervient en dernier lieu en choisissant le type et le niveau d'éducation selon ses propres talents et intérêts, les ressources qui ont été mises à sa disposition et les différents incitatifs auxquels il fait face. Autrement dit, on modélise les choix faits par les enfants comme étant le reflet de leurs goûts personnels, de leurs opportunités et de leurs contraintes, ces dernières étant largement déterminées par les choix sociaux et parentaux.

1.2 L'approche individuelle

L'approche générale procure une certaine perspective qui permet de situer la démarche utilisée par les différentes études sur la problématique des choix concernant l'éducation. Au niveau de la littérature, le cadre théorique économique centre son analyse en grande partie sur le contexte de l'individu pour expliquer ses différents choix en matière d'acquisition de capital humain et d'éducation. Ce sont en grande partie les travaux de Becker, Schultz et Mincer dans les années 60 qui servent de référence pour les modèles théoriques actuels. Ils furent les premiers à considérer théoriquement les choix reliés à l'éducation comme étant des décisions d'investissement en capital humain de la part des agents. Leurs travaux apportaient une nouvelle profondeur à la problématique qui était traitée antérieurement de façon similaire à la décision de consommer un bien ordinaire, c'est-à-dire en considérant les prix et le revenu réel pour expliquer la quantité demandée en terme d'éducation.

Parmi les premiers modèle tentant d'expliquer les facteurs qui interviennent dans l'investissement en capital humain, on retient celui de Becker (1967, 1975) pour sa simplicité et son efficacité. En cherchant à maximiser la valeur actualisée nette de ses revenus futurs étalés sur toute sa vie, l'étudiant atteindra un niveau d'investissement pour lequel les rendement et coût marginaux seront égaux. Par le fait même, il arrive à définir le concept de demande pour la scolarité qui devient ainsi fonction des rendements anticipés et de la probabilité de réussir un certain niveau de scolarité. Les rendements sont établis de manière exogène selon les forces du marché du travail, et la probabilité de réussite dépend des capacités individuelles, de l'environnement académique général et des croyances de l'étudiant par rapport à l'impact réel de la scolarité qu'il est en train d'acquérir. Par ailleurs, le coût d'opportunité est défini par rapport au manque à gagner relié à la participation au marché du travail. Ainsi, la quantité demandée d'éducation devient fonction de la valeur actualisée nette des flux associés à l'investissement qu'elle représente.

Toute forme de modélisation économique du comportement de l'individu faisant face à des choix reliés à l'acquisition de capital humain s'inspire du modèle précédent. Cependant le rôle de chacun des déterminants peut être plus nuancé d'un modèle à l'autre. En effet Becker accorde beaucoup d'importance aux incitatifs courants et futurs auxquels l'individu est confronté dans son processus décisionnel, c'est-à-dire le coût d'opportunité, les rendements

anticipés, etc. Alors que d'autres modèles mettent l'emphase davantage sur la perception des différents flux associés à la décision d'investissement. Ainsi cette perception qui se veut plus subjective peut être influencée par les ressources familiales, les valeurs et les goûts de l'étudiant, etc. Par conséquent, aux yeux de certains modèles (voir la revue des résultats importants), le background socio-économique peut avoir un rôle à jouer mais quel peut être son poids par rapport aux incitatifs courants?

1.3 Cadre statistique

Pour analyser empiriquement une question, certaines études vont privilégier une approche théorique afin d'imposer des contraintes aux données et d'avoir ainsi une modélisation formelle. On qualifie cette approche de forme structurelle dans la mesure où on tente d'ajuster les données à un modèle théorique. Il existe une approche alternative qui consiste à ne pas imposer de structure précise aux données. C'est une approche par forme réduite et cette démarche est une démarche beaucoup plus statistique que théorique. Contrairement à l'approche structurelle qui impose différentes formes aux données, l'estimation sous forme réduite ne cherche pas à caractériser les décisions comme des réponses à des incitatifs économiques comme dans le modèle théorique de Becker. Les différents liens de causalité et les interactions entre les déterminants sont plutôt estimés de manière empirique. Cette dernière propriété simplifie grandement l'analyse mais peu toutefois engendrer des résultats ad hoc. Les deux approches de modélisation, soit la forme réduite ou la forme structurelle, peuvent permettre d'obtenir des résultats intéressants dans le cadre de l'analyse des déterminants du processus décisionnel en matière d'éducation (Haveman et Wolfe, 1995). En fait, c'est la question qui détermine en grande partie lequel des deux modèles employer.

L'approche typique des études sur la question de l'impact du background socio-économique sur les décisions relatives à l'éducation s'intéresse à la relation, sous forme réduite, entre les différentes caractéristiques familiales (la structure, le revenu, etc.) et les niveaux atteints par les enfants en contrôlant les autres facteurs pertinents (l'éducation des parents, les caractéristiques du voisinage, etc.) à la lumière des données disponibles. L'utilisation de la forme réduite présuppose implicitement que les variables explicatives sont exogènes en ce qui a trait à leur impact sur la variable dépendante, soit le niveau d'épanouissement atteint (la scolarité atteinte, la décision de décrocher ou non ou le niveau de revenu atteint). L'hypothèse

d'indépendance entre les déterminants potentiels, c'est-à-dire que chaque coefficient reflète son effet total sur la variable expliquée, peut toutefois être restrictive (cette hypothèse peut être violée à différents degrés dans la réalité), chose qui est importante à considérer lorsqu'on analyse les relations estimées de cette façon. L'approche par la forme réduite permet d'obtenir des résultats d'ordre relationnel qui ne sont pas nécessairement dépourvus de sens; cependant l'estimation des liens de causalité demeure fragile, étant donnée la nécessité de montrer l'indépendance entre les déterminants qu'elle soit observée ou non.

1.4 Revue des résultats importants

1.4.1 Critères de sélection

L'article de Haveman et Wolfe (1995) trace un bilan des résultats importants sur la question des déterminants des atteintes des enfants. Ils sélectionnent selon trois critères les meilleurs articles sur la question, c'est-à-dire les articles qui cherchent à établir une relation entre les atteintes des enfants et un certain nombre de variables socio-économiques. Les critères de sélection sont : la qualité de l'étude (qualité des données et des méthodes d'estimation), l'orientation économique et l'utilisation de micro-données longitudinales (panel). Nous nous attarderons aux articles sélectionnés qui s'intéressent aux atteintes académiques (soit le niveau de scolarité et la décision relié au décrochage) des enfants pour fin de comparaison .

Tableau 1- *Résumé des articles importants*

Auteur(s)/Année	Type de données	Caractéristiques de l'échantillon	Méthode d'estimation	Variables dépendantes	Variables ind. pertinentes
Mayer (1991)	HSB	26 000 étudiants de 4 ^e année au secondaire en 1980	OLS et Logit	Décrochage au secondaire entre 1980 et 1982 = 1	Noir = 1 : (-) SES des parents : (-) Résultat math : (-)
Ribar (1991)	NLSY	4741 femmes, 14 à 21 ans en 1979	Probit bivarié	Gradué du secondaire avant 20 ans = 1	Noir = 1 : (+) Mère seul. = 1 : (-) Fam. reconst. = 1 : (-) Nbre d'enfants : (-) Educ. mère : (+) Mère au trav. : (-) 1 ^{er} enfant avant 20 ans = 1 : (-) Prat. rel. = 1 : (+) Accès lect. = 1 : (+)
Astone et McLanahan (1991)	HSB	10 000 étudiants de 2 ^e année au secondaire en 1980	Probit	Avoir complété le secondaire avant 1986 = 1	Fam. mono. = 1 : (-) Fam. reconst. = 1 : (-) Vivre sans parents = 1 : (-)

Haveman, Wolfe, Spaulding (1991)	PSID	1258 enfants, 0-6 ans en 1968, 19-23 ans en 1987	Probit	Gradué du secondaire avant 1987 = 1	Non-blanc = 1 : (0) Femme = 1 : (0) Éduc. père : (+) Éduc. mère : (+) Nbre d'années pauvreté : (0) Nbre d'années de trav. mère : (+) Nbre de déménagement : (-) Nbre de séparations parentales : (0) Nbre d'enfants : (0) Catholique = 1 : (+) Juif = 1 : (0) Protestant = 1 : (0)
Manski et al. (1992)	NLSY	2800 hommes et femmes, 14-17 ans en 1979	Probit, Probit bivarié, Probit trivarié, Modèles non-paramétriques	Avoir complété le secondaire avant 20 ans = 1	Noir = 1 : (+) Femme = 1 : (+) Éduc. père : (+) Éduc. mère : (+) Fam. non intacte à 14 ans = 1 : (-) Éduc. mère > Éduc. père = 1 : (+)
Sandefur, McLanahan et Wojtkiewicz (1992)	NLSY	5246 jeunes (14-17 ans) habitant avec leurs parents	OLS et Probit	Gradué du secondaire avant 1985 = 1	Fam. mono. à 14 ans = 1 : (-) Fam. reconst. à 14 ans = 1 : (-) Pas de parents à 14 ans = 1 : (-) Fam. mono. de 14 à 17 ans = 1 : (-) Fam. reconst. de 14 à 17 ans = 1 : (-) Pas de parents de 14 à 17 ans = 1 : (-) Revenu familial ajusté : (+) Estime de soi : (+)

PSID : Étude de panel sur les dynamiques de revenu au Michigan; HSB : Enquête au secondaire et postsecondaire; NLSY : Enquête longitudinale nationale sur les jeunes; SES : Statut socioéconomique; (+) : impact positif; (-) : impact négatif; (0) : pas d'impact significatif

Parmi les articles sélectionnés, même si certaines études utilisent l'estimateur des moindres carrés ordinaires ou une estimation simultanée, la plupart ont recours au maximum de vraisemblance (*probit*, *tobit*). De plus, le modèle à forme réduite est généralement utilisé par opposition à la forme structurelle qui cherche à caractériser les choix des enfants. Les prochaines sections regroupent par thème les résultats les plus répandus et pour lesquels il y a certains consensus. Toutefois, le problème d'endogénéité potentielle rend l'amplitude des estimateurs difficilement comparable entre les études. C'est pourquoi les comparaisons sont davantage orientées vers les signes et vers la signification statistique des estimateurs.

1.4.2 Impact des ressources familiales

D'un point de vue économique, la famille peut être modélisée comme étant une unité de production qui utilise des inputs réels afin de générer de l'utilité pour ses membres. La quantité, la nature et le moment de la distribution des ressources familiales accordées aux enfants influence le niveau de réalisation de ces derniers. Les ressources familiales peuvent prendre la forme d'aide financière, d'hébergement, de nourriture ou de services domestiques. Ainsi, l'accessibilité accrue à des ressources familiales tend à diminuer le coût d'entreprendre des études. Par ailleurs, il est raisonnable de s'attendre à ce que ces ressources soient proportionnelles au revenu des parents et inversement proportionnelles au nombre d'enfants. Généralement les ressources familiales sont mesurées par des variables reliées aux caractéristiques parentales : la scolarité, le revenu, etc. Celles-ci sont les plus répandues à travers les études.

1.4.2.1 La scolarité des parents

Parmi les études sur le décrochage scolaire, on retrouve souvent le niveau de scolarité des parents. Typiquement, il sert à mesurer le facteur économique le plus fondamental dans le cadre d'analyse des atteintes scolaires des enfants: le capital humain des parents. Les ressources familiales peuvent être vues en quelque sorte comme un transfert de richesse sous forme d'investissement en capital humain, des parents vers les enfants. C'est pourquoi il est raisonnable de s'attendre à ce que ce transfert soit proportionnel au niveau de scolarité des parents qui devient alors un bon indice de l'intérêt parental accordé aux études. Trois faits ressortent : la scolarité des parents est statistiquement significative et quantitativement importante et la scolarité de la mère est plus étroitement liée aux atteintes de l'enfant que celle du père.

1.4.2.2 Le revenu familial

Une autre caractéristique parentale qui revient fréquemment à travers les différentes études sur le décrochage scolaire est le niveau de revenu familial. Ce dernier représente le meilleur indice du niveau de ressources économiques disponible à l'épanouissement de l'enfant. Dans la majorité des cas, la variable du revenu familial est positivement associée avec les atteintes scolaires. Ce résultat est statistiquement significatif dans plus de la moitié des cas.

1.4.3 Impact de la structure familiale

La littérature théorique sur le capital humain ne précise généralement pas le type de structure familiale dans le cadre de ses modèles structurels. Cependant, cela n'a pas empêché certains chercheurs d'obtenir quelques résultats à cet égard avec une approche orientée davantage vers des modèles à forme réduite. En effet, McLanahan et Sandefur (1994) trouvent que le taux de graduation au secondaire et le taux d'inscription et de graduation pour des études post-secondaires sont inférieurs pour les enfants de famille monoparentale par rapport aux enfants de famille avec deux parents. D'autres études se sont intéressées plus particulièrement à l'impact des familles reconstituées. Par exemple, Boggess (1998) trouve que les enfants vivant avec un beau-père (celui avec lequel leur mère s'est remariée) ont des taux de graduation au secondaire plus faible que les enfants vivant avec leurs 2 parents biologiques. Globalement pour ce qui est des articles présentés au tableau 1, les familles monoparentales ou les familles qui ont expérimenté le divorce génèrent une influence négative sur les atteintes académiques des enfants et cette influence est statistiquement significative dans la plupart des études revues. Par ailleurs, l'impact de la présence de la mère sur le marché du travail n'est pas aussi clair. En effet, pour certaines études, l'influence de la mère au travail est négative et significative alors que pour d'autres elle est positive ou pas significative du tout.

Cependant d'autres chercheurs, n'étant pas convaincus des liens de causalité précédents entre la structure familiale et le décrochage scolaire, analysent la question différemment. En postulant plutôt que la structure familiale est endogène, certaines études cherchent à estimer si la corrélation entre la structure familiale et les atteintes scolaires reflète un phénomène de sélection plutôt que de causalité. Par exemple, à l'aide d'un modèle à effet fixe, Ginther et Pollak (2000) arrivent à montrer que la structure familiale (soit le nombre d'années passées avec un beau-parent dans une famille monoparentale ou dans une famille intacte) ne joue pas un rôle significatif sur les atteintes scolaires. Ainsi, ils trouvent plutôt un effet de sélection : l'hétérogénéité non-observée au sein des familles explique les atteintes scolaires mieux que la structure familiale observée. Certaines variables non-observées peuvent donc influencer à la fois la structure familiale et les atteintes scolaires.

1.4.4 Impact des autres facteurs qualitatifs

La plupart des études du tableau 1 incluent également à leur modèle des variables qui représentent des facteurs qualitatifs pour lesquels il est raisonnable de s'attendre à ce qu'il y ait un certain lien entre ces facteurs et les décisions prises en matière d'éducation par les enfants. Par exemple, les changements de structure familiale, les changements de milieu (déménagements), le voisinage et la fratrie peuvent affecter le développement de l'enfant et donc influencer ses choix. Afin de mieux cerner l'impact des différents facteurs d'ordre qualitatif, nous avons regroupé l'analyse des résultats importants en deux groupes : les facteurs observables et les facteurs non-observables.

1.4.4.1 Facteurs observables

Les facteurs qualitatifs observables sont souvent qualifiés « d'investissement parental » dans la mesure où ils correspondent à des décisions parentales concrètes qui ont des impacts directs sur la dynamique du milieu familial. Ces facteurs n'interviennent pas nécessairement de manière uniforme à travers les différentes études. Il en est de même pour l'impact estimé de ces facteurs sur la décision d'acquiescer de la scolarité ou sur la décision de décrocher ou non. Ainsi il est préférable de restreindre notre analyse aux facteurs de type « investissement parental » qui ont généralement un certain impact statistiquement significatif sur les atteintes scolaires des enfants à travers les études. Il s'agit entre autres du nombre de mouvement géographique (déménagement), du nombre d'enfant, du degré de pratique religieuse, de l'implication des parents dans les activités parascolaires et de l'accessibilité à du matériel de lecture à la maison pour les enfants. Par conséquent, l'ajout de telles variables à un modèle sur l'impact du background socio-économique sur le décrochage peut lui permettre des gains en vraisemblance.

1.4.4.2 Facteurs non-observables

Les facteurs non-observables sont par définition plus difficilement modélisables sous forme de variable. Ainsi, les articles du tableau 1 exploitent peu l'apport de telles variables qui peut se faire sentir au niveau de leur influence sur les résidus. En fait, le problème d'endogénéité est principalement caractérisé par l'impact des effets non-observables. Étant donné que la

plupart des études n'utilisent pas de modèle économétrique qui teste l'apport de tels facteurs, il est difficile de formuler des prévisions empiriques. Cependant la théorie économique permet de situer, dans le contexte d'acquisition de capital humain, l'influence des facteurs non-observables. Nous retenons trois types d'effet non-observables : les goûts, l'information et le voisinage.

1.4.4.2.1 Impact des goûts

Il est difficile d'inclure directement les notions de valeurs et de goûts d'un étudiant au sein de sa fonction de demande en capital humain. Cependant il est pertinent de supposer théoriquement que le milieu socio-économique de l'enfant influence ses goûts en matière de scolarisation. Puisque des parents scolarisés valorisent et encouragent généralement l'éducation (on fait ici l'hypothèse que leur niveau de scolarité est un choix personnel), on peut s'attendre, comme dans le cas des ressources familiales, à ce que le niveau de scolarité des parents influence positivement les valeurs et les goûts d'un enfant en matière d'éducation. Ainsi, des valeurs et des goûts prononcés en matière d'éducation peuvent avoir une certaine répercussion positive sur le niveau de motivation de l'étudiant et par conséquent sur la décision de poursuivre ses études.

1.4.4.2.2 Impact de l'information

L'information disponible sur les différentes avenues possibles reliées aux choix en matière d'éducation pour l'étudiant peuvent dépendre en grande partie de son milieu socio-économique. En effet, ce dernier peut affecter les coûts d'acquisition d'information par rapport aux différentes éventualités qui résultent des différentes décisions. L'expérience des parents génère des retombées informatives positives pour les enfants dans la mesure où ces derniers ont, eux aussi, déjà fait face à de telles décisions. Dans cette optique, un milieu scolarisé génère des coûts d'acquisition d'information plus faibles en ce qui a trait à la rentabilité réelle des études post-secondaires. Il n'est pas évident par contre qu'une meilleure information augmente nécessairement la demande pour des études post-secondaires. Cependant l'information joue un rôle sur le niveau de risque d'un investissement en éducation : une meilleure information sur l'impact réel de poursuivre ou non ses études rend la décision moins risquée pour l'étudiant. Nous approfondirons la notion de risque dans le contexte d'investissement en capital humain à la section 1.5.

1.4.4.2.3 *L'effet de voisinage*

Un voisinage tend à isoler, sous forme de réseau, des gens ayant les mêmes antécédents. Ce phénomène engendre des contacts sociaux qui ont de l'influence sur les attitudes culturelles et les opportunités économiques des membres du voisinage (Borjas 1995). Ainsi certaines études tentent de modéliser la liaison entre les externalités du capital humain et la géographie. Cependant une telle approche présente plusieurs complications. D'abord il existe un problème de définition, car la plupart des chercheurs utilisent, faute d'avoir des informations plus précises, le code postal pour isoler chaque voisinage. Or celui-ci correspond rarement au vrai voisinage dans lequel on retrouve les différentes interactions sociales. Ensuite, comme pour plusieurs variables, les problèmes d'endogénéité potentielle compliquent l'analyse. En effet, le choix du voisinage dépend souvent des ressources économiques familiales. Par conséquent, le revenu parental et l'indice de qualité d'un voisinage sont généralement liés dans la mesure où le dernier est choisi en fonction du premier. Toutefois, en dépit de ces complications, certains chercheurs arrivent à isoler différents groupes d'étudiants pour en étudier l'impact sur les membres du groupe. C'est le cas de Summers et Wolfe (1977) et Henderson, Mieszkowski et Sauvageau (1978) qui trouvent, toute chose étant égale par ailleurs, que les étudiants qui évoluent dans un groupe d'étudiants performants atteignent des plus hauts niveaux de réussite. Par contre, les effets de sélection établis précédemment peuvent jouer un rôle important.

1.4.5 Impact du travail/étude

Théoriquement le nombre d'années d'études accomplies est probabiliste et ce, conditionnellement aux performances individuelles antérieures, au niveau de participation sur le marché du travail en terme d'heures travaillées et aux caractéristiques personnelles telles l'habileté et le niveau de motivation. Plusieurs études, autres que celles du tableau 1, se penchent sur l'impact réel d'une participation au marché du travail pendant les études. La grande question demeure au niveau de son impact sur la réussite de l'étudiant et par le fait même de son influence sur le décrochage scolaire.

D'une part un emploi permet à l'étudiant de commencer à acquérir une expérience en parallèle. Cette recherche de préparation pour le marché du travail peut être avantageuse. En effet, différentes études trouvent des taux de chômage plus faibles et des salaires plus élevés

pour les nouveaux arrivants sur le marché du travail qui ont travaillé pendant leurs études, par rapport à ceux qui se sont consacrés uniquement à leurs études (Stephenson 1981, Ruhm 1997).

D'autre part, un emploi permet à l'étudiant d'obtenir un revenu et donc d'accéder à un meilleur niveau de consommation à court terme. Toutefois les attentes quant à la variable explicative du revenu de l'étudiant sont ambiguës. Premièrement, un bon revenu peut permettre à l'étudiant de compenser pour les imperfections de marché du capital financier et donc de mieux assumer les coûts relatifs à des études secondaires et post-secondaires. Deuxièmement, un bon revenu peut aussi représenter le coût d'opportunité d'entreprendre des études (manque à gagner).

Chose certaine, l'impact direct du travail pendant les études se situe au niveau de la diminution du temps accordé au loisir (pour lequel la valeur peut varier d'un individu à l'autre). De plus, certaines études arrivent à montrer qu'il existe même un arbitrage entre le nombre d'heures travaillées pendant les études et le nombre d'heures consacrées à l'étude elle-même (D'Amico et Baker, 1984; Steinberg et Dornbusch, 1991). L'étude de Marsh (1991) va même un peu plus loin en démontrant que le travail pendant les études nuit aux facteurs reliés à la performance scolaire. Dans la mesure où la diminution de la performance nuit à la probabilité de poursuivre ses études, une participation au marché du travail pour un étudiant peut influencer positivement sa décision de décrocher.

Toutefois, l'étude Dagenais, Montmarquette et Viennot-Briot (2001) apporte une certaine nuance à la problématique. En définissant 2 types d'étudiants, soit ceux qui privilégient l'école à la participation au marché du travail à temps plein et ceux qui préfèrent limiter leur éducation au profit d'un accès plus rapide au marché du travail, ils parviennent à mieux isoler les déterminants de la performance scolaire, de la décision de travailler ou non pendant les études et de la décision d'abandonner ou non. Ainsi ils arrivent à montrer qu'un travail parallèle aux études de moins de 15 heures par semaine ne nuit pas nécessairement aux succès académiques. Par ailleurs, ils trouvent que l'âge légal pour accéder au marché du travail, le salaire minimum et la conjoncture économique influencent la décision d'abandonner ou non les études. Par exemple, un faible taux de chômage pousse davantage les étudiants vers l'abandon.

1.5 Intégration du risque

Les différentes études qui ont été abordées jusqu'à maintenant n'incluent pas la notion de risque dans leurs modèles pour analyser les déterminants du décrochage scolaire et ce même si les économistes reconnaissent généralement l'impact considérable du risque dans la décision d'investissement en capital humain. En fait, les contraintes provenant des banques de données font souvent en sorte que la notion de risque est difficilement modélisable dans le cadre d'une approche par forme réduite. C'est pourquoi elle se retrouve davantage au niveau de la littérature récente qui s'intéresse aux formes structurelles. En effet, différentes études ont réussi à inclure, par exemple, de l'aversion pour le risque à travers les hypothèses faites quant aux formes des fonctions d'utilités des agents (Chen, 2002).

La notion de risque dans le contexte d'investissement en éducation s'aborde de la même manière que pour toutes formes d'investissement. Même si a priori, d'un point de vue social, le capital humain s'ajuste assez bien aux diverses circonstances économiques, il est probable que l'acquisition de capital humain s'avère relativement risquée, sur le plan individuel, par rapport à l'acquisition de capital physique. En effet, étant donné qu'il est impossible de le transiger, le capital humain est indissociable de son propriétaire. De plus, l'arbitrage qui existe entre une formation générale et une spécialisation nuit à la possibilité pour un individu de diversifier son avoir en capital humain. Cette diversification qui permet une meilleure répartition du risque est cependant plus accessible dans le processus d'acquisition de capital physique.

Ainsi, l'attitude face au risque d'un individu qui fait face à des choix relativement à son éducation peut avoir une certaine influence. Pour certains économistes, l'aversion pour le risque et le désir d'investir en capital humain sont liés positivement étant donné que le niveau de scolarité réduit la variance du revenu anticipé³, et ce même si l'approche classique permet de présumer le contraire. En fait, tout repose sur le risque réel en soit d'un investissement en capital humain et sur la perception de ce risque par celui qui prend la décision. Particulièrement, la perception du risque réel peut être influencée par différents facteurs d'ordre socio-économique. Par exemple, sous l'hypothèse de l'utilité marginale décroissante

³ Kodde, 1986.

du revenu, il est raisonnable de s'attendre à ce qu'un individu ayant un faible niveau de revenu soit plus averse au risque qu'un individu de haut revenu. Par conséquent, même si un individu est riscophobe à cause de son faible niveau de revenu, si ce dernier lui permet tout juste de subsister, il lui sera très difficile de prendre des décisions d'investissement (Eckel, Johnson et Montmarquette, 2002).

Notre étude s'oriente davantage vers la notion de perception ou d'attitude de l'individu face au risque, plutôt qu'à l'évaluation du niveau de risque d'un investissement en capital humain d'ordre académique. La nature de la banque de données utilisée (qui sera précisée à la prochaine section) permet de mesurer de façon imparfaite l'attitude face au risque et donc d'estimer empiriquement l'impact de cette dernière sur la décision de décrocher ou non.

Chapitre 2

2.1 Les données

L'analyse empirique est faite à partir d'une banque de données unique dans laquelle sont accumulées des informations portant sur un échantillon composé de 1161 garçons, de parents francophones nés au Canada, fréquentant une maternelle de quartier défavorisé, à Montréal en 1984. Des 1161 individus de l'échantillon initial, 124 sujets n'ont pas été retenus étant donné qu'après vérification (les données provenaient initialement des commissions scolaires), on a constaté qu'ils ne répondaient pas à toutes les caractéristiques circonscrites par le champ d'étude. Après la première évaluation à la maternelle, des évaluations annuelles ont été faites à partir de l'âge de 10 ans. Ces évaluations provenaient des parents, des enseignants, des élèves de la classe et des jeunes eux-mêmes.

Cette banque de données a été construite dans le cadre d'une étude unique au monde sur la question de la délinquance dirigée par le professeur Richard Tremblay du département de psychologie et de psychiatrie de l'Université de Montréal. Cette étude à grand déploiement s'intéresse particulièrement à la source de la délinquance et de la violence. En accord avec la nature des données accumulées au fil des ans, la démarche consiste à extraire parmi la multitude de mesures, les données à caractère économique pertinentes, en cohérence avec ce qui a été fait dans ce domaine par la littérature, pour ensuite analyser à l'aide de diverses estimations les déterminants de l'acquisition de capital humain dans le cadre de la problématique du décrochage scolaire.

2.2 Spécification et construction des variables

2.2.1 Variable dépendante

Décrochage (DECROCHE) : variable dichotomique qui indique si l'individu a pris la décision d'abandonner ses études secondaires soit en 1993, 1994 ou en 1995.

DECROCHE = 1 si l'individu a connu un décrochage au cours de ces trois années
= 0 autrement

2.2.2 Variables de type familial

Âge du père à la naissance du premier enfant (APEREPRE) : variable continue qui indique l'âge du père à la naissance du premier enfant de la famille de l'individu.

Âge de la mère à la naissance du premier enfant (AMEREPRE) : variable continue qui indique l'âge de la mère à la naissance du premier enfant de la famille de l'individu.

Scolarité de la mère (SCOLMER) : variable continue qui indique le nombre d'années de scolarité de la mère.

SCOL010 : variable dichotomique qui vaut 1 si SCOLMER est entre 0 et 10; 0 autrement

SCOL1124 : variable dichotomique qui vaut 1 si SCOLMER est entre 11 et 24; 0 autrement

Scolarité du père (SCOLPER) : variable continue qui indique le nombre d'années de scolarité du père.

Statut familial : variable qui s'intéresse à la structure familiale. Elle est divisée en trois catégories principales :

- **Intacte** : variable dichotomique de référence qui indique si la structure familiale est demeuré inchangée (famille biparentale standard), dans quel cas la variable vaut 1 et 0 sinon.
- **Reconstituée (RECONSTI)** : variable dichotomique qui indique si la structure familiale est reconstituée (famille biparentale non-intacte), dans quel cas la variable vaut 1 et 0 sinon.
- **Autres (AUTRE)**: variable dichotomique qui recoupe tous les autres cas de structure familiale, c'est à dire les familles qui ne sont ni intactes ni reconstituées, dans quel cas la variable vaut 1 et 0 sinon.

Nombre d'enfants en 1984 (NBENF84) : variable continue qui indique le nombre de frères et sœurs de l'individu présent au sein de sa famille en 1984.

2.2.3 Variable de type individuel

Année de naissance (ANNAISSA) : variable dichotomique qui indique si l'année de naissance de l'individu est 1977(1) ou 1978 (0).

Classification en maths et en français : variable qui agrège les résultats des tests de classification en français et en mathématiques pour la période de 1988 à 1992. Le test de classification se divise en 5 catégories :

- 1 : excellent
- 2 : très bien
- 3 : bien
- 4 : passable
- 5 : échec

La variable classification totale est la somme des 10 résultats (5 en français et 5 en mathématiques). Puisque cette dernière varie de 10 à 50 de manière discontinue, elle est divisée en 4 variables dichotomiques pour les fins d'estimation :

- **Classification 10-30** : variable dichotomique de référence qui indique si l'individu a obtenu un résultat entre 10 et 30 exclu, dans quel cas la variable vaut 1 et 0 sinon.
- **Classification 30-35 (CLAS3035)** : variable dichotomique qui indique si l'individu a obtenu un résultat entre 30 et 35 exclu, dans quel cas la variable vaut 1 et 0 sinon.
- **Classification 35-40 (CLAS3540)** : variable dichotomique qui indique si l'individu a obtenu un résultat entre 35 et 40 exclu, dans quel cas la variable vaut 1 et 0 sinon.
- **Classification 40-50 (CLAS4050)** : variable dichotomique qui indique si l'individu a obtenu un résultat entre 40 et 50, dans quel cas la variable vaut 1 et 0 sinon.

Type de classe : variable nominale qui représente une évaluation sur 5 années (1988 à 1992) du statut de l'étudiant par rapport à ses besoins en terme d'encadrement académique. Cette variable prend les valeurs suivantes :

1 : régulier	J : handicapé moteur
2 : difficulté d'apprentissage	K : infirme moteur
3 : enrichi	L : handicapé moteur grave
5 : immersion	M : handicapé physique
A : francisation	Q : handicapé auditif
B : immersion	R : handicapé intellect profond
C : accueil	S : handicapé visuel
D : intellect moyen	T : difficulté grave de comportement
E : bilingue	V : troubles graves d'apprentissage
G : intellect léger	W : troubles sévères d'apprentissage
H : intellect moyen	X : déviations multiples

Pour des fins d'estimation, la variable est regroupée en trois variables dichotomiques :

- **Régulier-enrichi** : variable dichotomique de référence qui indique si l'individu a obtenu majoritairement (au moins 3 sur 5) sur les 5 années le type de classe : 1 ou 3; dans quel cas la variable vaut 1 et 0 sinon.
- **Difficultés d'apprentissage (DIFFAPPR)** : variable dichotomique qui indique si l'individu a obtenu majoritairement (au moins 3 sur 5) sur les 5 années le type de classe : 2, V ou W; dans quel cas la variable vaut 1 et 0 sinon.

- **Autres types de difficultés (AUTRDIFF)** : variable dichotomique qui indique si l'individu a obtenu majoritairement sur les 5 années n'importe quel autre type de classe, dans quel cas la variable vaut 1 et 0 sinon.

Travail du sujet (TRAVSUJ) : variable dichotomique qui indique s'il y a eu ou non, de la part de l'individu, participation au marché du travail sur la période de 1990 à 1995. 1 si l'individu a travaillé pendant cette période; 0 sinon.

Échelle de délinquance (QASTOT) : variable continue qui est la somme des variables « total délinquance » de 1988 à 1995. Les variables « total délinquance » sont elles-mêmes une agrégation pour chaque année des indices sur la consommation de drogues et l'implication au sein d'actes de bataille, de vol et de destruction/vandalisme tel que fournis par l'individu lui-même dans le cadre d'un questionnaire. La variable QASTOT varie de façon continue de 202 à 506. Le niveau de délinquance est directement proportionnel au total obtenu.

Échelle d'attitude face au risque (GAMBLING) : variable continue qui est la somme des variables « déjà joué pour argent » de 1988 à 1992. Ces dernières varient de 1 à 4 :

- 1 : jamais
- 2 : 1 ou 2 fois
- 3 : plusieurs fois
- 4 : très souvent

Ainsi la variable GAMBLING varie potentiellement de 5 à 20. On fait l'hypothèse qu'un résultat faible peut être interprété comme étant un comportement riscophobe et vice versa.

GAMB2 : variable continue qui prend la valeur de la variable GAMBLING au carré.

2.2.4 Variables omises

Les variables suivantes n'ont pas été retenues en raison soit d'un nombre relativement élevé de données manquantes ou soit d'un problème de définition :

Travail de la mère : participation ou non au marché du travail pour la mère sur la période de 1991 à 1993.

Prestige occupationnel de la mère et du père : classification des types de professions.

Mesure d'impulsivité : délai versus gratification.

2.3 Statistiques descriptives

Les statistiques descriptives permettent de présenter les caractéristiques générales de l'échantillon analysé. La question centrale de cette étude s'intéresse au problème du décrochage au secondaire qui est conceptualisé par la variable dépendante. Ainsi, le taux de décrochage pour les garçons formant l'échantillon est de 24 % par rapport aux 1002 observations valides. Pour ce qui est des variables indépendantes, notre analyse se divisera en deux parties : les caractéristiques familiales et les caractéristiques individuelles. (Voir tableaux 1 et 2).

Tableau 2 - *Statistiques descriptives : caractéristiques familiales*

Symbole	Définition	Moyenne	Nombre d'observations
APEREPRE	Âge du père à la naissance du 1 ^{er} enfant	26,3832	902
AMEREPRE	Âge de la mère à la naissance du 1 ^{er} enfant	23,2970	997
SCOLMER	Scolarité de la mère	10,4946	1011
SCOLPER	Scolarité du père	10,5628	947
INTACTE	Statut familial : intacte	0,68	1008
AUTRE	Statut familial : autre	0,25	1008
RECONSTI	Statut familial : reconstituée	0,07	1008
NBENF84	Nombre d'enfants à la maison en 84	1,10	991

Plusieurs faits intéressants ressortent de l'analyse des statistiques descriptives des caractéristiques familiales. D'abord, en moyenne la scolarité de la mère est très proche de celle du père. Ensuite, on constate que les familles reconstituées sont minoritaires avec une proportion d'environ 6%. Les familles intactes étant majoritaire avec 68%. Par ailleurs, pour plus de la moitié des familles considérées dans l'échantillon l'individu dont le comportement est étudié a un frère ou une sœur. Finalement, l'histogramme en annexe nous permet également de conclure que les familles avec enfant unique sont les deuxièmes plus répandues.

Tableau 3 - *Statistiques descriptives : caractéristiques individuelles*

Symbole	Définition	Moyenne	Nombre d'observations
ANNAISSA	Année de naissance; 1 :1977, 0 : 1978	0,25	1161
CLAS1030	Classification 10-30	0,32	1051
CLAS3035	Classification 30-35	0,23	1051
CLAS3540	Classification 35-40	0,27	1051
CLAS4050	Classification 40-50	0,18	1051
REG_ENRI	Type de classe : régulier-enrichi	0,80	783
AUTRDIFF	Type de classe : autres types de difficultés	0,11	783
DIFFAPPR	Type de classe : difficultés d'apprentissage	0,09	783
TRAVSUJ	1 : si le sujet n'a jamais travaillé, 0 : autrement	0,16	830
QASTOT	Échelle de délinquance	266,4668	966
GAMBLING	Attitude face au risque	6,6990	960

L'échantillon est majoritairement composé de garçons nés en 1978 (75%). Les regroupements par classification en math-français sont relativement bien distribués mais on remarque que la classification 10-30 qui correspond à l'excellence est majoritaire, suivie de la classification 35-40. La plupart des étudiants sont considérés comme étant capables d'évoluer dans le cadre d'une classe régulière ou enrichie (80%). Les étudiants ayant des difficultés d'apprentissage sont minoritaires avec une proportion de 9%. Une grande partie des étudiants ont eu des expériences de travail pendant leurs études (84%). Finalement, pour ce qui est des variables continues reliées à la délinquance et à l'attitude face au risque, l'analyse de leur distribution (qui sont semblables) nous permet de conclure que les comportements ne sont pas distribués de façon normale et que la grande majorité des individus ont des comportements modérés.

2.4 Méthodologie empirique

La banque de données ainsi que les contraintes économétriques influencent en grande partie le design de cette étude. L'analyse empirique repose sur des données très riches en information et en variables qui ont pour effet d'orienter notre analyse vers une approche axée sur les impacts des caractéristiques et les choix individuels dans le cadre du processus d'acquisition de capital humain. Plus précisément l'estimation par probit permet d'évaluer l'impact des différents déterminants socio-économiques pertinents sur la probabilité de décrocher au secondaire.

2.4.1 Endogénéité versus exogénéité

Les variables utilisées pour expliquer la probabilité de décrocher peuvent se regrouper en 2 catégories : les caractéristiques parentales et familiales et les caractéristiques individuelles. Parmi celles-ci plusieurs peuvent avoir un caractère endogène, c'est-à-dire qu'ils peuvent avoir une certaine corrélation avec les résidus qui incorporent les facteurs non-observables. Par exemple, la variable qui classe les individus selon leur niveau en mathématique et en français peut être corrélée avec le niveau de motivation de l'individu qui représente un facteur non-observable. Cependant, l'existence de décalage entre le moment de la mesure des caractéristiques et de la prise de décision de décrocher ou non permet de soutenir l'hypothèse d'exogénéité stricte pour les variables utilisées. En effet, les caractéristiques parentales et familiales proviennent de mesures prises en 1984 et les caractéristiques individuelles sont l'agrégation de mesures prises sur plusieurs années débutant en 1988 pour la plupart alors que la décision de décrocher ou non n'est observée qu'à partir de 1993, d'où le décalage de quelques années.

2.4.2 Corrélations

Par ailleurs, certaines corrélations sont soupçonnées à l'intérieur des 2 catégories. Par exemple, les caractéristiques qui se rattachent au père de l'individu peuvent être corrélées avec celles de la mère du même individu, ou bien les caractéristiques qui influencent le caractère délinquant d'un individu peuvent également avoir un impact sur sa propension à jouer pour de l'argent, dans la mesure où les deux comportements de « sensation seeking»

peuvent être issus de distorsions d'évaluation du risque (Zuckerman, 1994). Cependant, puisque le test de corrélation de Pearson ne donne aucun résultat plus haut que 0,95, nous soutenons comme hypothèse que les corrélations possibles ne sont pas en mesure d'influencer les estimations.

2.4.3 Observations manquantes

Le problème de non-réponse dans le cadre d'évaluations diverses pour une étude donnée peut s'avérer très important lorsque ces données attendues ont un rôle considérable au niveau des analyses par régression. Restreindre l'analyse seulement aux observations qui sont complètes peut introduire des pertes d'efficacité et engendrer des résultats biaisés si celles-ci ne forment pas un sous-ensemble aléatoire de l'ensemble original. Par exemple, lors d'une enquête de type « panel », les pertes de données sont rarement aléatoires car elles sont plutôt dues aux effets de sélection. En fait, lorsque les trous de données sont systématiquement reliés au phénomène modélisé, on parle d'enquêtes avec auto-sélection.

On distingue trois types généraux de données manquantes selon la nature de l'absence de l'observation : observations manquantes aléatoires, catégories manquantes, observations manquantes non-aléatoires⁴. Le premier type fait référence à un oubli de répondre par manque d'attention ou de distraction de la part du répondant. Le type « catégories manquantes » est utilisé lorsque certaines questions peuvent générer différents types de réponses qui ne sont pas nécessairement tous intégrés dans les catégories de réponses disponibles au répondant. Finalement, une observation manquante non-aléatoire fait allusion à une réponse qui est volontairement non-divulguée par le répondant pour une quelconque raison. Par exemple, les questions d'enquêtes touchant le salaire du répondant peuvent être intentionnellement omises par souci de confidentialité. Évidemment, l'approche à utiliser pour résoudre un tel problème diffère selon le type d'observations manquantes auquel le chercheur est confronté.

Lorsque les données sont non-disponibles sans raison apparente et sans lien avec les autres observations complètes de l'échantillon alors le sous-échantillon complet est utilisable (« ignorable case » Griliches, 1986). Autrement dit, si la considération d'efficacité n'est pas la priorité principale, mieux vaut ignorer le problème car le remplacement des valeurs

⁴ Haitovsky, 1967.

manquantes par les moyennes d'échantillon peut sous-estimer les variances et covariances tout en biaisant l'estimation des pentes (Santos, 1981). Dans notre cas, nous considérons que les données manquantes sont aléatoires pour simplifier l'analyse.

2.5 Modèles

La décision de décrocher ou non est examinée à l'aide d'un modèle binaire *probit* avec variable latente D^* . Celle-ci représente les préférences de l'individu par rapport à la décision de décrocher du système scolaire ou non. La variable D^* n'est pas directement observable : elle décrit le niveau d'utilité de l'individu dans le contexte de sa prise de décision en matière d'éducation. Ainsi :

$$D^* = X\beta + \varepsilon$$
$$\left| \begin{array}{l} D = 1 \text{ si } D^* < 0 \\ D = 0 \text{ sinon} \end{array} \right.$$

Pour les fins d'estimation, nous employons un processus itératif qui tient compte entre autres des observations manquantes, de la présence de colinéarités potentiels entre les variables dépendantes et de la possibilité d'avoir une forme fonctionnelle. Ainsi plusieurs modèles sont estimés sur des sous-échantillons créés par les contraintes amenées par le processus.

Modèle 1

Le modèle 1 sert de régression de référence. Il comprend toutes les variables pertinentes et disponibles sauf la variable TRAVMERE étant donné un manque relativement important d'observations (599 données manquantes sur 1161) et compte tenu que pour 62% des données disponibles, la mère n'a jamais travaillé.

$$D^* = f(\text{ANNAISSA, APEREPRE, AMEREPRE, SCOLMER, SCOLPER, AUTRE, RECONSTI, NBENF84, CLAS3035, CLAS3540, CLAS4050, AUTRDIFF, DIFFAPPR, TRAVSUJ, QASTOT, GAMBLING})$$

Modèle 2

Ce modèle est une variante du précédent auquel est ajouté, pour tenter de capter les comportements extrêmes, une forme fonctionnelle quadratique pour la variable « GAMBLING ».

$D^* = f$ (ANNAISSA, APEREPRE, AMEREPRE, SCOLMER, SCOLPER, AUTRE, RECONSTI, NBENF84, CLAS3035, CLAS3540, CLAS4050, AUTRDIFF, DIFFAPPR, TRAVSUJ, QASTOT, GAMBLING, GAMBLING²)

Modèle 3

Le modèle 3 diffère du modèle 2 par le retrait des variables « APEREPRE » et « SCOLPERE » qui sont des reliées aux caractéristiques du père du sujet. De plus, la variable « SCOLMERE » est raffinée pour laisser place à la variable « SCOL010 »

$D^* = f$ (ANNAISSA, AMEREPRE, SCOL010, AUTRE, RECONSTI, NBENF84, CLAS3035, CLAS3540, CLAS4050, AUTRDIFF, DIFFAPPR, TRAVSUJ, QASTOT, GAMBLING, GAMBLING²)

Modèle 4

Ce modèle est presque identique au modèle précédent. L'unique différence provient du retrait de la variable « QASTOT » pour tester son influence sur la variable « GAMBLING ».

$D^* = f$ (ANNAISSA, AMEREPRE, SCOL010, AUTRE, RECONSTI, NBENF84, CLAS3035, CLAS3540, CLAS4050, AUTRDIFF, DIFFAPPR, TRAVSUJ, GAMBLING, GAMBLING²)

Modèle 5

Puisque les variables « AUTRDIFF » et « DIFFAPPR » possèdent relativement plus de données manquantes que les autres et que le type d'encadrement scolaire est généralement lié aux performances académiques (classification math-français), ce modèle omet ces deux variables.

$D^* = f$ (ANNAISSA, AMEREPRE, SCOL010, AUTRE, RECONSTI, NBENF84, CLAS3035, CLAS3540, CLAS4050, TRAVSUJ, QASTOT, GAMBLING, GAMBLING²)

Chapitre 3

3.1 Résultats empiriques

Tableau 4 - Résultats empiriques

	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5
Nbre d'obs.	473	473	525	525	744
Log fct. de vraisemblance	-169,93	-169,22	-199,05	-205,76	-300,24
CONSTANTE	-2,1593* (0,8876)	-1,0865 (1,2669)	-0,6572 (1,0273)	0,2386 (0,9959)	-0,9366 (0,8693)
ANNAISSA	-0,01149 (0,1868)	-0,01436 (0,1875)	-0,0451 (0,17)	-0,0820 (0,1670)	0,0548 (0,1353)
APEREPRE	-0,00769 (0,0175)	-0,008756 (0,1764)			
AMEREPRE	-0,0456** (0,0247)	-0,04606** (0,0248)	-0,0605* (0,0203)	-0,0628* (0,0201)	-0,0627* (0,0170)
SCOLMER	-0,03582 (0,0333)	-0,03595 (0,0334)			
SCOL010			0,07626 (0,151)	0,0438 (0,1489)	0,2437* (0,1206)
SCOLPER	0,02889 (0,0309)	0,0253 (0,312)			
AUTRE	0,3619** (0,19)	0,3860* (0,1911)	0,4432* (0,1655)	0,4395* (0,1633)	0,4546* (0,1321)
RECONSTI	0,512 (0,404)	0,5182 (0,4074)	0,2329 (0,348)	0,4360 (0,3242)	-0,2009 (0,2774)
NBENF84	0,1156 (0,0758)	0,1151 (0,07616)	0,126** (0,07)	0,1002 (0,0686)	0,0599 (0,0609)
CLAS3035	0,6914* (0,2382)	0,6765* (0,2394)	0,572* (0,2137)	0,5759* (0,2098)	0,6474* (0,1859)
CLAS3540	0,9268* (0,2398)	0,9153* (0,2402)	0,7522* (0,2113)	0,8706* (0,2053)	0,9527* (0,1774)
CLAS4050	1,0503* (0,2884)	1,055* (0,289)	1,002* (0,2618)	1,1110* (0,2571)	1,2029* (0,1894)
AUTRDIFF	0,95722* (0,2737)	0,9131* (0,277)	0,7661* (0,2307)	0,9206* (0,2238)	
DIFFAPPR	0,6261* (0,2615)	0,611* (0,262)	0,5075* (0,2438)	0,4681** (0,2427)	
TRAVSUJ	-0,4009** (0,2279)	-0,4111** (0,2292)	-0,4824* (0,2214)	-0,5124* (0,2175)	-0,4824* (0,1916)
QASTOT	0,00764* (0,00234)	0,00804* (0,00239)	0,00756* (0,00211)		0,0076* (0,0017)
GAMBLING	-0,06184 (0,0498)	-0,3502 (0,2486)	-0,4046** (0,23)	-0,1676 (0,2152)	-0,3370** (0,1866)
GAMB2		0,017 (0,014)	0,01998 (0,0132)	0,0118 (0,0128)	0,0161 (0,0109)

L'écart-type est entre parenthèses. * Significativement différent de 0 au niveau 5%. ** Significativement différent de 0 au niveau 10%.

3.2 Analyse empirique

Le modèle 1 sert de modèle de référence dans la mesure où toutes les variables qui semblent pertinentes a priori sont estimées, et ce sans tester de forme fonctionnelle quelconque. Par ailleurs, étant donné que chaque variable ou presque possède des données manquantes qui ne sont pas forcément reliées aux mêmes individus, plus le nombre de variables estimées est grand, moins le nombre d'observations retenues pour la régression est élevé (effet d'omission). De plus, cette variation au niveau des observations retenues rend la comparaison entre les modèles difficile. Ainsi, pour les fins de l'analyse, nous regrouperons dans notre discussion les modèles dont la régression est effectuée sur le même nombre d'observations.

Modèle 1 et 2

Les 2 premiers modèles utilisent le même ensemble de variables, soit la totalité de ce qui est disponible et pertinent. L'unique différence provient du test de la forme fonctionnelle quadratique pour la variable « GAMBLING » dans le modèle 2. Cette dernière permet d'amplifier les comportements extrêmes en ce qui a trait à l'attitude face au jeu et au risque. Les 2 modèles sont comparables étant donné qu'ils couvrent le même nombre d'observations, soit 473. La forme quadratique permet un gain en ajustement de modélisation qui est ressenti au niveau d'une légère augmentation de la fonction de vraisemblance. Cependant, la variable « GAMBLING » n'est pas significativement différente de 0 par rapport à son impact sur la variable dépendante, et ce qu'elle soit quadratique ou non. Outre le gain en vraisemblance, le modèle 2 se différencie également du premier par une amélioration de la précision de la variable « AUTRE ». En effet, elle devient significative à un niveau de confiance inférieur, soit 5%.

Pour ce qui est des conclusions communes et pertinentes aux deux premiers modèles, on remarque que les variables qui ont un rôle significatif au niveau 5% sont d'ordre individuel. Il s'agit du niveau de classement en math-français, du type d'encadrement scolaire et de l'indice de délinquance. Finalement, même si elle ne sont pas significatives à 5%, on rejette l'hypothèse du coefficient nul à 10% pour les variables « AMEREPRE » et « TRAVSUJ ». Nous reviendrons un peu plus loin sur le sens et sur l'ordre de grandeur des coefficients lorsque le nombre de variables significatives sera plus important.

Modèle 3 et 4

Les modèles 3 et 4 se distinguent du modèle de base par l'omission de certaines variables qui, par leur certain niveau de corrélation et leur manque de pertinence, peuvent nuire à la spécification de certaines variables plus fondamentales. D'abord pour chacun d'entre eux, les variables « APEREPRE » et « SCOLPER » sont retirées. La première variable correspond à l'âge du père à la naissance du premier enfant qui est corrélé significativement au niveau 1% à l'âge de la mère pour la naissance du même enfant, et ce dans un ordre de grandeur de 0,551. Il en est de même pour les variables « SCOLPER » et « SCOLMER » qui sont également corrélées à 0,499 au niveau 1%. Dans les deux cas, il ne s'agit pas de corrélations très fortes (0,9 et plus) mais la littérature nous informe que les caractéristiques de la mère sont généralement plus étroitement liées aux atteintes des enfants que celles du père. Il est donc intéressant d'observer l'apport de l'omission de ces deux variables. Par ailleurs, la scolarité de la mère est redéfinie de façon plus précise afin de pouvoir tester si un niveau de scolarité plus élevé que 11 années (qui correspond à l'obtention d'un diplôme d'études secondaires) a plus d'impact sur les atteintes de l'enfant que lorsque le niveau de scolarité est inférieur à ce chiffre. Cet effet est caractérisé par la variable dichotomique « SCOL010 » qui a pour référence dans le modèle la variable « SCOL1124 ».

Modèle 3

En terme de vraisemblance, le modèle 3 ne se compare pas avec le modèle de base dans la mesure où le nombre d'observations utilisées pour la régression diffère. En effet on passe de 473 à 525 observations. Cette augmentation est due à l'effet d'omission. Cependant, le principal apport de ce modèle est au niveau du gain en précision du degré de signification de certaines variables. Dans un premier temps, les variables « AMEREPRE » et « TRAVSUJ » deviennent significatives au niveau 5% et dans un deuxième temps, les coefficients des variables « GAMBLING » et « NBENF84 » diffèrent désormais de 0 et ce de manière significative au niveau 10%. Par ailleurs, il est à noter que l'impact d'une scolarité pour la mère inférieure à 11 années n'est pas statistiquement différent de celui d'une scolarité supérieure. Finalement, compte tenu du grand nombre de variables significatives, le modèle 3 nous permet d'approfondir notre analyse.

Caractéristiques familiales

D'abord, pour ce qui est des variables dites de type familiales, seules les variables « SCOL010 » et « RECONSTI » ne sont pas significatives. Pour le premier cas, cela implique

qu'une scolarité élevée pour la mère n'a pas plus d'impact qu'une scolarité faible sur la probabilité de décrocher pour l'enfant. Bien que ce résultat soit surprenant puisqu'à travers la littérature la scolarité de la mère est généralement importante dans le processus d'acquisition de capital humain pour l'enfant, il peut s'expliquer par la faible variance pour la scolarité de la mère au sein de l'échantillon estimé. Pour le deuxième cas, toujours en terme d'impact sur le décrochage, le type de structure familiale reconstituée ne se différencie pas de la structure intacte. En terme économique, ce dernier résultat implique que dans le cadre de notre échantillon, les familles reconstituées génèrent des ressources familiales de même nature pour l'enfant (en ce qui a trait à leur impact sur le décrochage) que les familles dites intactes. Cependant, la faible proportion relative de famille reconstituée pourrait également expliquer le manque de précision de l'estimation du coefficient de la variable.

Ensuite, toujours pour les caractéristiques familiales, les variables « AMEREPRE » et « AUTRE » sont significatives à 5% alors que la variable « NBENF84 » est statistiquement différente de 0 à 10%. Premièrement, le coefficient négatif de la variable « AMEREPRE » nous indique que plus l'âge de la mère à son premier enfant est élevé, plus la probabilité de décrocher pour l'individu issu de cette famille est faible. Ainsi un milieu familial dont la mère a commencé à avoir ses enfants en bas âge engendre moins de ressources d'encadrement contre le décrochage qu'un milieu où la mère a eu ses enfants plus tard. Ce dernier résultat correspond à celui de Ribar (1991) sur la même question. Deuxièmement, pour ce qui est de l'impact de la structure familiale, un individu qui provient d'une famille de type « AUTRE », c'est à dire ni intacte, ni reconstituée, a une plus grande probabilité de décrocher qu'un individu provenant d'une famille intacte, soit la variable de référence. Si on oriente notre analyse en terme de ressources familiales disponibles pour l'enfant, une famille « AUTRE », qui correspond en grande partie à une famille monoparentale, génère un niveau de ressources plus faible qu'une famille intacte. Ce résultat revient systématiquement dans chaque étude analysé au tableau 1 qui s'intéresse à la question. Finalement, le nombre d'enfants est en relation positive avec la probabilité de décrocher. Notre résultat est peu robuste par manque de variance, mais il demeure tout de même que l'impact du nombre d'enfants est généralement statistiquement significatif à travers la littérature. Encore une fois, l'impact du nombre d'enfants s'explique économiquement par son influence décroissante sur le niveau de ressources disponibles pour chaque enfant. Cependant, l'étude de Haveman, Wolfe et Spaulding (1991) trouve que l'impact du nombre d'enfant n'est pas significativement différent de 0. Ce résultat peut aussi expliquer le manque de robustesse du nôtre.

Caractéristiques individuelles

Les variables représentant les caractéristiques individuelles sont toutes significatives au niveau 5%. D'abord, si on considère les variables de classement en math-français, on remarque que leur impact sur le décrochage est statistiquement différent du classement de référence « CLAS1030 » qui correspond au meilleur classement possible. Ainsi, chaque classement inférieur influence positivement la probabilité de décrocher et cette probabilité croît à mesure que le niveau de classement diminue (un classement 40-50 étant inférieur à un classement 30-35). Ce résultat semble assez robuste puisqu'on le retrouve pour chaque modèle estimé. Par conséquent, un étudiant qui excelle en mathématiques et en français a moins de chance de décrocher qu'un étudiant qui réussit moins bien. Cette variable capte en quelque sorte les aptitudes académiques de l'étudiant qui sont généralement, selon la littérature, inversement liés au décrochage. Par ailleurs, les aptitudes académiques de l'étudiant peuvent également être reflétées par le type d'encadrement scolaire requis pour l'étudiant en question. Pour notre échantillon, les étudiants sont regroupés selon trois types : « régulier-enrichi », « difficultés d'apprentissage » et « autres difficultés ». Par rapport à la variable de référence qui est « régulier-enrichi », les étudiants qui présentent des difficultés d'apprentissage ou autres ont une plus grande propension au décrochage. De plus, la probabilité de décrocher est plus élevée pour les étudiants qui ont des difficultés de types « autres », c'est à dire toutes formes de difficultés différentes des difficultés d'apprentissage (handicaps, troubles de comportement, etc.) Finalement, les résultats liés aux habiletés académiques de l'enfant sont ceux qui ressortent le mieux à travers chaque modèle.

L'indice de délinquance « QASTOT » et l'attitude face au jeu « GAMBLING » sont des variables de type individuel relativement uniques si on compare avec les différentes études présentes dans la littérature puisque les données de cette nature sont rarement disponibles. L'ajout de ces variables contribuent grandement à la particularité de notre étude. D'abord, on constate que l'indice de délinquance est positivement lié à la probabilité de décrocher mais, l'impact est relativement faible étant donné le faible coefficient estimé associé à la variable (0,007). Il est à noter que l'effet de la délinquance revient systématiquement pour tous les modèles où il est estimé. Par ailleurs, une forme quadratique a été testée pour la variable « QASTOT » afin de tenter de capter les comportements extrêmes, mais les résultats ne sont pas présentés car ils sont non-significatifs et donc inintéressants. Ensuite, pour la variable de l'attitude face au jeu, seul le coefficient du terme de premier degré est significatif à 10%. Effectivement l'estimation de l'effet du deuxième degré manque nettement de précision.

Cependant, à la lumière du résultat obtenu avec le modèle 2 par rapport au modèle de référence, la forme fonctionnelle de la variable « GAMBLING » semble se situer quelque part entre la forme linéaire et la forme quadratique. Toutefois, le manque de robustesse apparent des résultats reliés à cette dernière variable ne nous permet pas d'approfondir davantage.

Notre dernière analyse s'intéresse à l'impact du travail pendant les études sur la probabilité de décrocher. Puisque la variable dichotomique « TRAVSUJ » vaut 1 si l'individu n'a jamais travaillé pendant ses études, le coefficient négatif estimé implique qu'un individu qui n'a jamais travaillé pendant ses études a une probabilité plus faible de décrocher qu'un individu qui a eu au moins une expérience de travail. Puisque la notion de travail est définie de manière peu précise (elle n'inclut pas le nombre d'heures travaillées, le salaire, etc.), il est difficile de comparer ce résultat avec ceux apportés par la littérature. Cependant ce dernier peut s'expliquer économiquement par le coût d'opportunité plus élevé que représente le fait de poursuivre des études pour les étudiants qui ont une certaine expérience de travail. D'autant plus que ce coût peut être relativement important pour des étudiants issus d'un milieu défavorisé.

Modèle 4

Le modèle 4 a pour objectif de tester l'influence de la variable du niveau de délinquance (« QASTOT ») sur l'estimation de la variable de l'attitude face au jeu (« GAMBLING »). Il est généralement reconnu que les comportements de délinquance (vandalisme, usage des drogues) sont liés aux comportements de jeu pour de l'argent, et ce à travers une recherche commune de sensation forte traduite par des distorsions d'évaluation du risque (Zuckerman, 1994). L'intérêt est donc de vérifier si ces deux variables se nuisent mutuellement lors de l'estimation. D'abord, comparativement au modèle 3 qui sert de référence, l'estimation de ce nouveau modèle génère une fonction de vraisemblance plus faible donc a priori, le modèle semble moins bien spécifié. Ensuite, le principal changement se situe directement au niveau de ce qui était testé : la variable « GAMBLING » deviennent non-significatives. Finalement, les estimateurs des variables « NBENF84 » et « DIFFAPPR » perdent également de la précision. Par conséquent, à la lumière des résultats trouvés en estimant le modèle 4, le modèle semble mieux spécifié lorsqu'on inclut conjointement les variables « QASTOT » et « GAMBLING ».

Modèle 5

Les variables qui spécifient le type d'encadrement scolaire, soit « AUTRDIFF » et « DIFFAPPR », possèdent relativement un grand nombre de données manquantes (254). Le modèle 5 omet donc ces deux variables afin de pouvoir faire une estimation sur un plus grand nombre d'observations. Par ailleurs, le type d'encadrement scolaire peut être lié au classement en math-français et donc les estimateurs peuvent se nuire mutuellement. Cependant les indices de corrélation entre les types de variables sont faibles mais significatifs (inférieur à 0,4). Tel qu'attendu, le nombre d'observations est le plus élevé de tous avec 744. La vraisemblance est par conséquent incomparable avec les autres modèles. Pour ce qui est des changements sur les estimateurs, le plus grand nombre d'observations doit générer de la variance pour la variable qui représente la scolarité maternelle de 0 à 10 ans, car elle devient statistiquement différente de la variable de référence « SCOL1124 ». Ainsi, une scolarité maternelle inférieure à 11 ans (qui correspond à l'obtention d'un diplôme d'études secondaires) a un impact positif sur le décrochage par rapport à une scolarité supérieure ou égale à celle d'un DES. Par ailleurs, le nombre d'enfant n'a plus d'influence significative sur le décrochage. Finalement, l'estimation de la forme quadratique de la variable « GAMBLING » nous donne des résultats similaires à ceux du modèle 3.

Globalement, les 5 modèles estimés ne présentent pas de contradiction entre eux : il n'y a pas de changements brusques au niveau des estimations significatives. Certains modèles permettent de raffiner notre analyse pour ce qui est de l'impact de certaines variables alors que pour d'autres variables, l'effet estimé est constant d'un modèle à l'autre. Plus particulièrement, les variables de classement en math-français sortent significative à 5% pour chaque modèle. Tout comme les variables reliées à la délinquance et au type d'encadrement scolaire qui sortent aussi à 5 % pour tous les modèles où elles sont estimées. Pour ce qui est des autres constances, les variables « ANNAISSA » et « RECONSTI » ne sont jamais significativement différentes de 0 d'un modèle à l'autre. Les principales différences proviennent des contraintes amenées par chaque modèle. Ainsi, pour son grand nombre de variables significatives, le modèle 3 semble être celui qui est le mieux spécifié.

3.3 Conclusion

Ce travail s'intéresse à la question du décrochage scolaire et de ses déterminants socio-économiques. Par rapport aux travaux antérieurs sur la question au sein de la littérature, la particularité de cette étude se situe principalement au niveau de l'utilisation d'une banque de données riche et complexe. En effet, celle-ci permet de modéliser la problématique à l'aide de différentes mesures très précises provenant d'un échantillon de garçons montréalais issus de milieux défavorisés suivis pendant plusieurs années. Ces données proviennent d'une vaste étude longitudinale qui est citée sur le plan international pour son ampleur et sa qualité. Dirigée par le professeur Richard Tremblay du département de psychologie et de psychiatrie de l'Université de Montréal, cette étude a permis à différents chercheurs d'obtenir des résultats très intéressants en ce qui a trait entre autres au phénomène de la délinquance.

À l'aide d'un modèle *probit*, nous estimons par forme réduite la relation entre la probabilité de décrocher et différentes variables socio-économiques parmi lesquelles certaines ont été bâties à même la panoplie de mesures présentes dans la banque de données. Le survol des résultats importants de la section 3.4 dresse un bilan des variables les plus fréquemment utilisées à travers les études portant sur le même type de problématique. Malheureusement, certaines contraintes d'information nous empêchent d'inclure la totalité de ces variables courantes. Ainsi, il ne nous est pas possible d'estimer les effets de variables telles le revenu familial, le nombre de déménagement, le degré de pratique religieuse et l'implication des parents au niveau de la vie académique de l'enfant. Par ailleurs, le modèle économétrique ne permet pas de mesurer l'impact des effets non-observés. Cependant la particularité de nos données permet d'ajouter au modèle des déterminants pertinents propres à cette étude tels le niveau de classement en math-français, le type d'encadrement scolaire requis, le niveau de délinquance et l'attitude face au jeu.

En posant pour hypothèses que les variables indépendantes sont exogènes, que les différentes corrélations possibles n'ont pas d'influence sur le modèle et que les données manquantes sont aléatoires, nous obtenons plusieurs résultats intéressants. Puisque notre étude ne s'intéresse qu'à un groupe spécifique d'individus, les comparaisons avec la littérature demeurent plutôt symboliques.

D'abord la scolarité de la mère, par manque de variance, joue un rôle uniquement lorsque la régression est faite sur plusieurs observations : l'estimation devient alors en accord avec la littérature dans la mesure où la scolarité de la mère est inversement proportionnelle à la probabilité de décrocher. Par ailleurs, l'âge de celle-ci à la naissance de son premier enfant est en relation négative avec la variable dépendante, et ce de façon significative.

Ensuite nous notons dans un premier temps qu'à l'image de la plupart des conclusions admises sur le sujet, les familles monoparentales influencent positivement la probabilité de décrocher par rapport aux familles intactes ; ce résultat est relativement significatif d'un modèle à l'autre. Cependant les familles reconstituées ne se distinguent pas des familles intactes quant à leur impact sur le décrochage, ce résultat peut sans doute s'expliquer par la faiblesse relative du nombre de famille reconstituée. Dans un deuxième temps, nous observons qu'un nombre d'enfants élevé pour une même famille influence le décrochage, ce qui est en accord avec la théorie des ressources familiales. Toutefois ce résultat semble très peu robuste.

Les résultats qui concernent les caractéristiques individuelles sont multiples. En premier lieu, les conclusions quant aux variables sur les niveaux de classement en math-français coïncident avec la théorie sur les habiletés : les faibles niveaux de classement influencent davantage la probabilité de décrocher par rapport aux niveaux plus élevés. De plus, ce résultat semble être le plus robuste étant donné qu'il se retrouve systématiquement pour chacun des modèles estimés. Dans le même ordre d'idées, les étudiants présentant des difficultés d'apprentissage et autres ont plus de chances de décrocher par rapport aux étudiants dits de niveau « régulier ou enrichi ». Ensuite, pour ce qui est du travail pendant les études, les étudiants qui n'ont jamais travaillé ont une probabilité plus faible de décrocher. Ce résultat relativement robuste tend à défendre l'hypothèse du coût d'opportunité relié au fait d'entreprendre des études. Toutefois, les contraintes sur les données empêchant le raffinement de cette variable nuisent aux comparaisons avec la littérature. Finalement, la délinquance influence faiblement et positivement le décrochage alors que l'attitude face au risque, sous les formes fonctionnelles estimées, n'a qu'un impact plus ou moins significatif, et ce en dépit de sa colinéarité potentielle avec la délinquance. Une meilleure mesure de la variable de l'attitude face au risque pourrait toutefois améliorer son impact dans le modèle.

Les difficultés méthodologiques d'un tel travail se situent principalement au niveau des données manquantes et des problèmes d'endogénéité potentielle et de biais de sélection. D'abord, l'hypothèse des observations manquantes aléatoires peut être restrictive dans la mesure où les difficultés d'obtenir certains renseignements auprès d'un étudiant concernant des mesures d'ordre académiques peuvent être directement reliées au phénomène observé, soit le décrochage. Autrement dit, ce n'est pas nécessairement un hasard si le suivi de certains étudiants se perd dans le temps, cela peut plutôt traduire des effets de sélection.

Ensuite, pour bien comprendre les liens de causalité, il est important de bien distinguer les liens d'interdépendance entre les déterminants qui servent à expliquer la variable dépendante. Malheureusement, les déterminants estimés n'ont pas toujours le degré d'orthogonalité nécessaire pour assurer l'indépendance des effets estimés. Par exemple, très peu de chercheurs affirmeraient que la structure familiale est exogène. Alors comment des variables endogènes affectent d'autres variables endogènes? Ainsi, la négligence du fait que certaines variables non-observées peuvent influencer à la fois la structure familiale et les atteintes scolaires peut être problématique. Les conflits familiaux (et tout le stress qu'ils engendrent) peuvent aussi bien être à la source d'un divorce que de piètres résultats scolaires de la part de l'enfant. Les effets de sélection sont loin d'être impossibles et donc viennent brouiller les interprétations des estimations.. La négligence de la prise en compte de ce type de problème peut entraîner une perte de robustesse en ce qui concerne l'estimateur des moindres carrés ordinaires, dans la mesure où celui-ci sous-évalue la vraie variance et estime avec inexactitude les différences dans les variances.

Afin d'améliorer ce travail, il pourrait être intéressant d'appliquer d'une part, un modèle probit bivarié avec biais de sélection pour tester l'hypothèse des données manquantes aléatoires. En effet, celui-ci pourrait estimer la probabilité de décrocher compte tenu le fait que l'individu soit conservé dans l'échantillon. Ainsi, l'hypothèse ne serait pas rejetée si l'application d'un tel modèle ne changeait pas les résultats d'estimations initiaux. D'autre part, on pourrait également appliquer un test d'exogénéité à notre modèle pour encore une fois tester l'hypothèse qui s'y rattache. Finalement, la banque de données présente certaines lacunes en ce qui a trait à la mesure de certaines variables économiques telles l'attitude face au risque et les préférences pour le temps. Ainsi, l'étude pourrait bénéficier d'une démarche d'économie expérimentale auprès des individus de l'échantillon pour améliorer ces mesures économiques pertinentes dans l'étude du décrochage scolaire.

Bibliographie

Becker, Gary S., "Human Capital and the Personal Distribution of Income" (Ann Arbor: University of Michigan Department of Economics, 1967).

Becker, Gary S., "Human Capital" (New York: Columbia University Press for the National Bureau of Economic Research, second edition, 1975).

Chen S. "Is Investing in College Education Risky?" Working Paper. State University of New York at Albany (2002).

Dagenais M., Montmarquette C., Viennot-Briot N. "Dropout, School Performance and Working while in School : An Econometric Model with Heterogeneous Groups". Working Paper, Scientific Series, CIRANO, 2001.

D'Amico, R. and P. Baker, "The Nature and Consequences of High School Employment", in *Pathways to the Future*, P. Baker et al.(editors), Columbus : Ohio State University, volume 4, (1984),.1-49.

Eckel C., Johnson C., Montmarquette C. "Will the Working Poor Invest in Human Capital? A Laboratory Experiment". Working Paper, Social Research And Demonstration Corporation, 2002.

Ginther D., Pollak R. "Does Family Structure Affect Children's Educational Outcomes?" Working Paper, Department of Economics Washington University, 2000.

Griliches, Z. "Economic Data Issues." In Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, Vol. 3. Amsterdam : North Holland,1986.

Haitovsky Y. "Missing Data in Regression Analysis. *Journal of the Royal Statistical Society*". Series B (Methodological), Volume 30, Issue 1, 1968, 67-82.

Haveman R., Wolfe B. "The Determinants of Children's Attainments: A Review of Methods and Findings". *Journal of Economic Literature*, Volume 33, Issue 4, 1995, 1829-1878.

Henderson V., Mieszkowski P. et Sauvageau Y. "Peer Group Effects and Educational Production Functions." *J. Public Econ.* 10 (August 1978):97-106.

Houle R., Ouellet L. "The Influence of Socio-Economic Factors on Private Demand for University". *The Canadian Journal of Economics*, Volume 15, Issue 4, 1982, 693-705.

Marsh, H.B., "Employment During High School : Character Builder or a Subversion of Academic Goals?", *Sociology of Education* 64, (1991), 172-189.

Ruhm, C.J., "Is High School Employment Consumption or Investment?", *Journal of Labor Economics* 15, (1997), 735-776.

Santos R. "Effects of imputation on regression coefficients". In *Proc. Survey Res. Meth. Sect., Am. Statist. Assoc.*, 1981, pp.140-5.

Steinberg, L. and S.M. Dornbusch, "Negative Correlates of Part-Time Employment During Adolescence : Replication and Elaboration", *Developmental Psychology* 27, (1991), 304-313.

Stephenson, S.P., "In-School Labour Force Status and Post-School Wage Rates of Young Men", *Applied Economics*, 13, (1981), 279-302.

Summers, Anita A., et Wolfe, Barbara L. "Do Schools Make a Difference?" *A.E.R.* 67 (September 1977) : 639-52.

Zuckerman Marvin. "Behavioral expression and biosocial bases of sensation seeking". New York: Cambridge University Press, 1994.

Annexe

Tableau 5 - *Distribution de la variable GAMBLING*

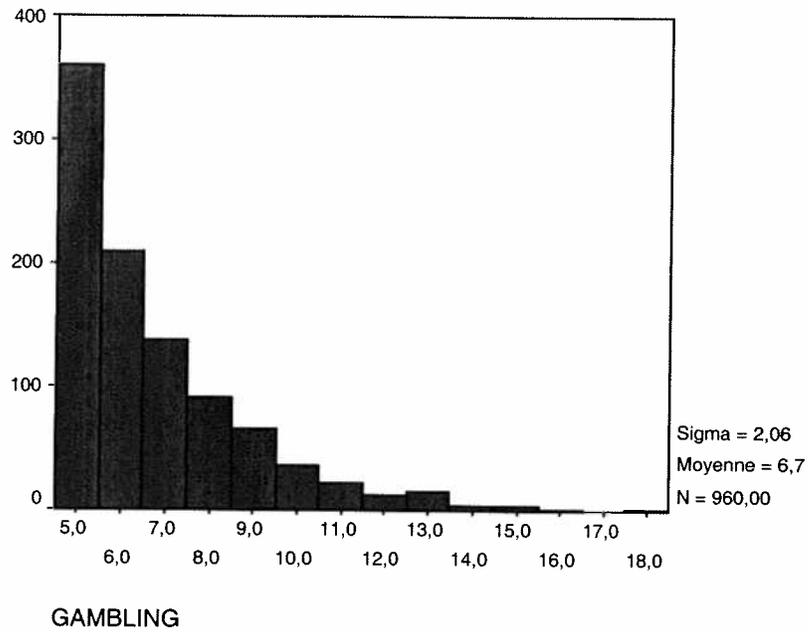


Tableau 6 - *Distribution de la variable QASTOT*

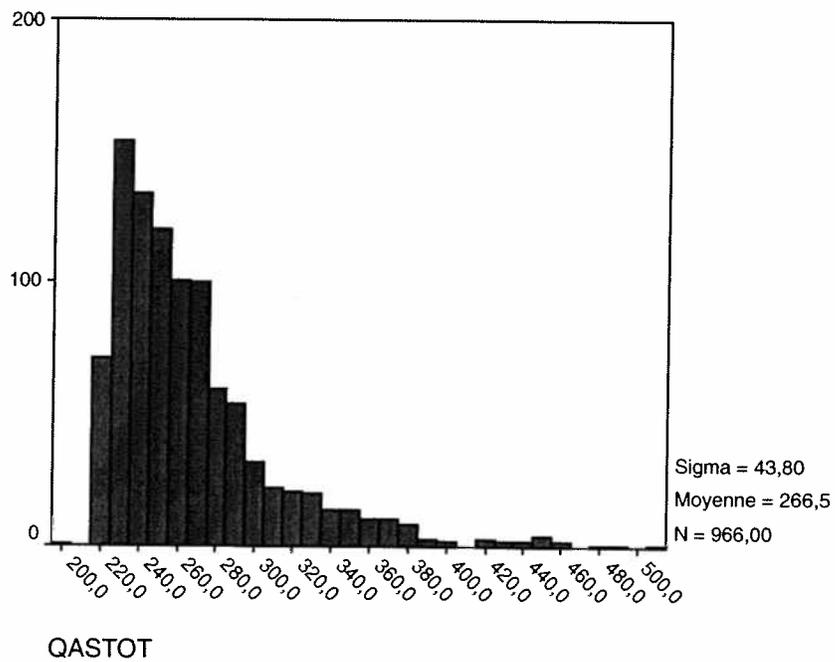


Tableau 7 - Distribution de la variable CLASTOT

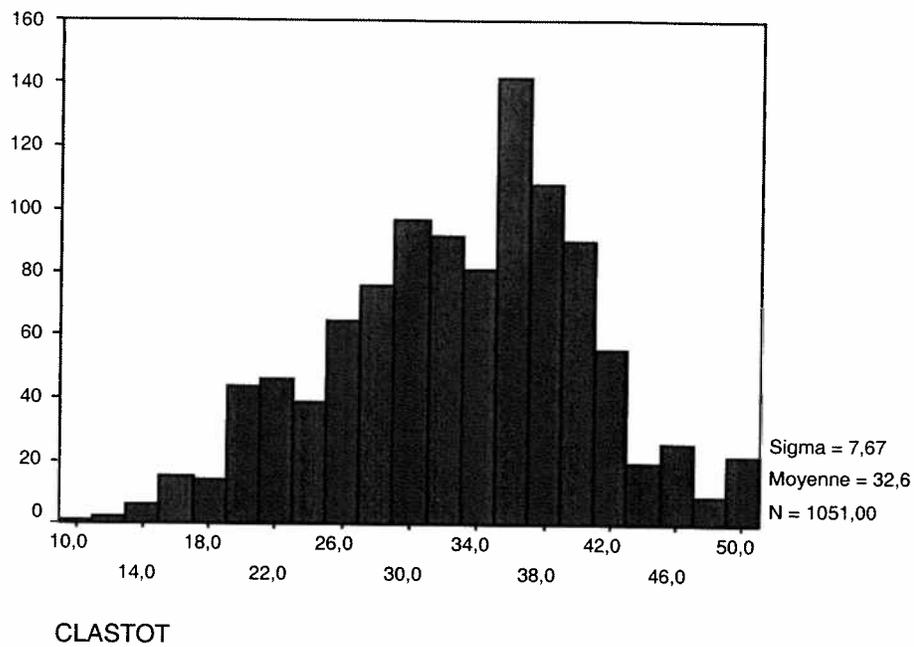


Tableau 8 - Distribution de la variable NBENF84

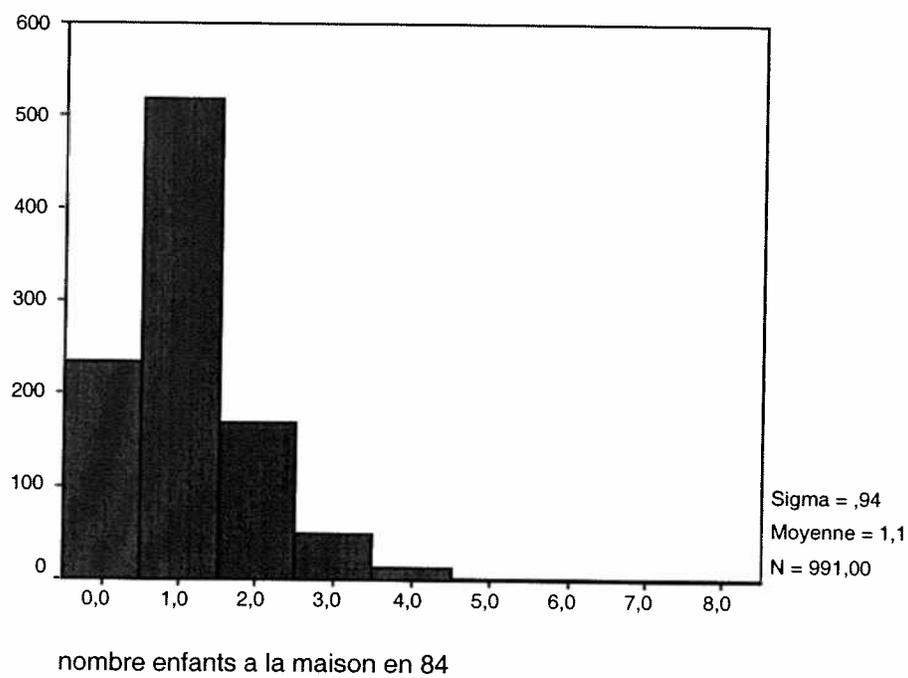


Tableau 9 - Distribution de la variable SCOLMERE

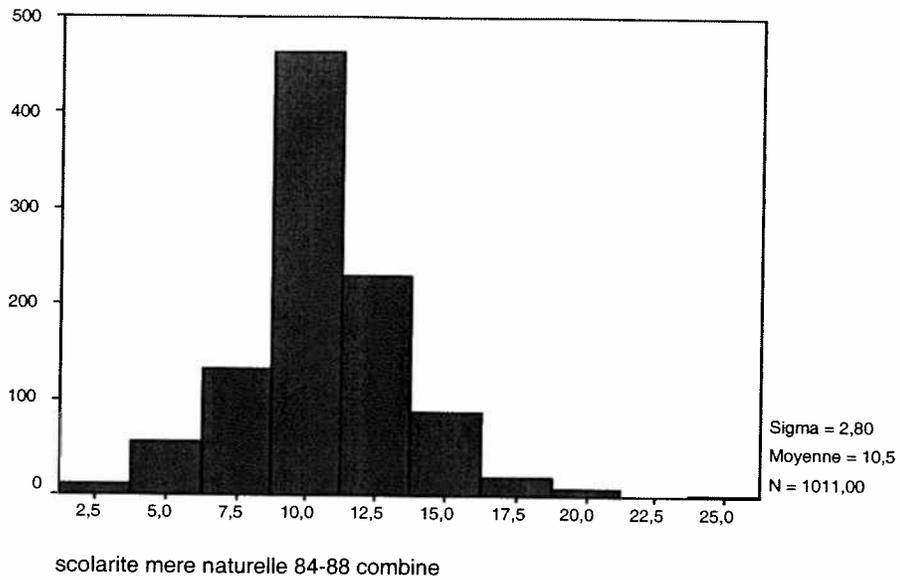


Tableau 10 - Distribution de la variable SCOLPERE

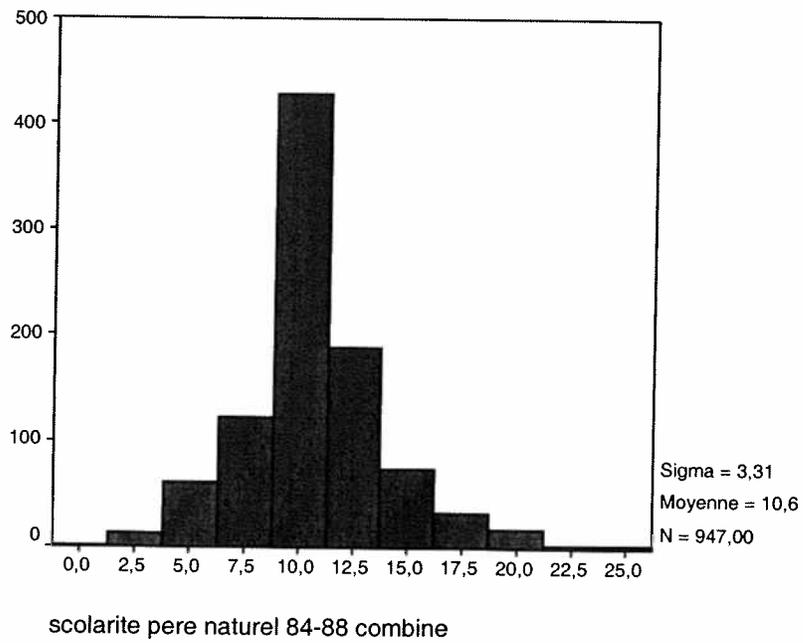


Tableau 11 - *Distribution de la variable AMEREPRE*

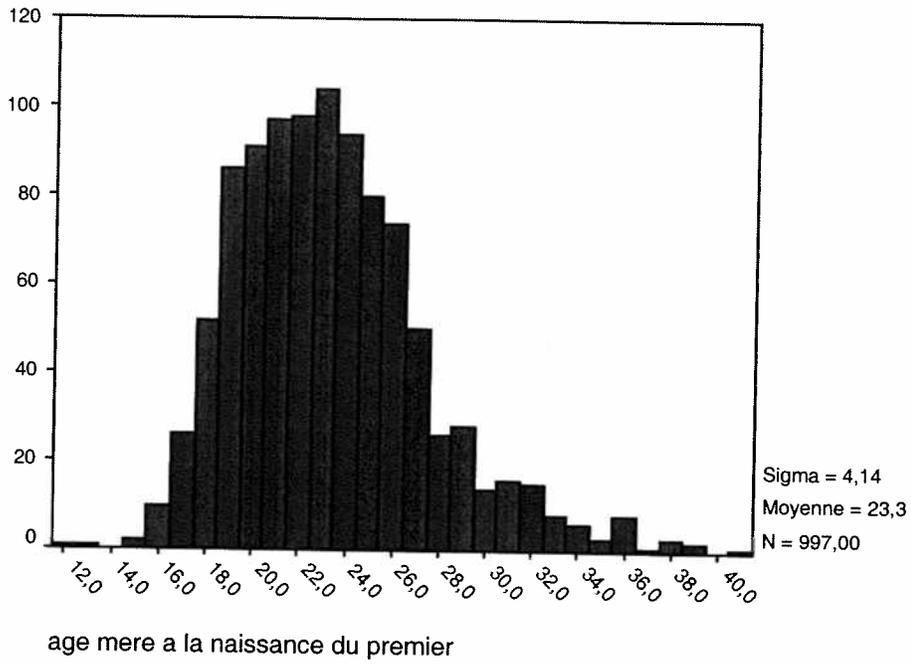


Tableau 12 - *Distribution de la variable APEREPRE*

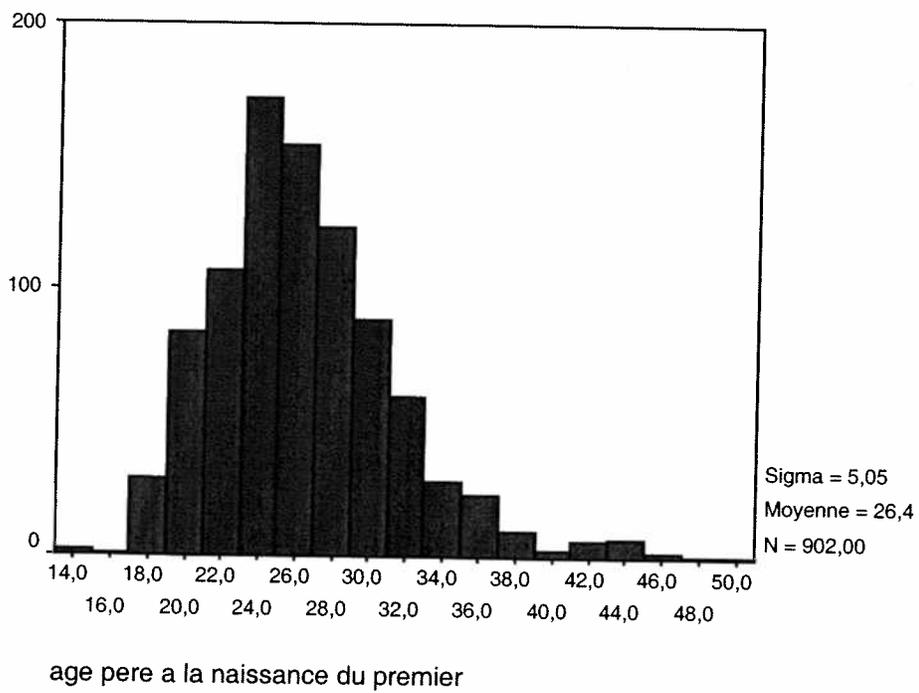


Tableau 13 - *Fréquences de la variable CLASTOT*

	Fréquenc	Pourcentage	Pourcentage		Fréquenc	Pourcentage	Pourcentage
	e	valide	cumulé		e	valide	cumulé
10,00	1	,1	,1	31,25	4	,4	41,1
12,00	2	,2	,3	31,66	1	,1	41,2
13,00	2	,2	,5	31,66	2	,2	41,4
13,32	1	,1	,6	31,68	1	,1	41,5
13,34	2	,2	,8	31,75	2	,2	41,7
14,00	1	,1	,9	31,92	1	,1	41,8
15,00	7	,7	1,5	32,00	28	2,7	44,4
16,00	6	,6	2,1	32,25	1	,1	44,5
16,66	1	,1	2,2	32,34	1	,1	44,6
16,68	1	,1	2,3	32,50	26	2,5	47,1
17,00	2	,2	2,5	32,91	2	,2	47,3
17,50	3	,3	2,8	33,00	28	2,7	50,0
18,00	7	,7	3,4	33,32	1	,1	50,0
18,34	2	,2	3,6	33,34	6	,6	50,6
19,00	14	1,3	4,9	33,75	9	,9	51,5
19,50	1	,1	5,0	34,00	29	2,8	54,2
19,75	1	,1	5,1	34,25	3	,3	54,5
20,00	28	2,7	7,8	34,50	2	,2	54,7
21,00	14	1,3	9,1	34,59	1	,1	54,8
21,25	3	,3	9,4	34,75	2	,2	55,0
21,66	1	,1	9,5	35,00	74	7,0	62,0
21,66	1	,1	9,6	35,41	2	,2	62,2
21,68	2	,2	9,8	35,50	3	,3	62,5
21,75	1	,1	9,9	35,84	1	,1	62,6
22,00	13	1,2	11,1	36,00	36	3,4	66,0
22,50	11	1,0	12,2	36,25	15	1,4	67,5
23,00	9	,9	13,0	36,50	1	,1	67,6
23,32	3	,3	13,3	36,66	5	,5	68,0
23,34	1	,1	13,4	36,68	4	,4	68,4
23,75	4	,4	13,8	36,75	1	,1	68,5
23,83	1	,1	13,9	37,00	25	2,4	70,9
24,00	21	2,0	15,9	37,50	31	2,9	73,8
25,00	37	3,5	19,4	37,75	2	,2	74,0
26,00	14	1,3	20,7	38,00	31	2,9	77,0
26,25	9	,9	21,6	38,34	8	,8	77,7
26,66	2	,2	21,8	38,50	1	,1	77,8
26,68	1	,1	21,9	38,75	10	1,0	78,8
26,75	2	,2	22,1	39,00	28	2,7	81,4
27,00	16	1,5	23,6	39,25	2	,2	81,6
27,09	1	,1	23,7	39,32	1	,1	81,7
27,50	21	2,0	25,7	39,50	2	,2	81,9
27,75	1	,1	25,8	40,00	55	5,2	87,2
28,00	15	1,4	27,2	40,25	1	,1	87,3
28,32	5	,5	27,7	40,41	1	,1	87,3
28,34	3	,3	28,0	41,00	13	1,2	88,6
28,34	4	,4	28,4	41,25	2	,2	88,8
28,75	10	1,0	29,3	41,66	3	,3	89,1
29,00	22	2,1	31,4	41,68	1	,1	89,2
29,32	1	,1	31,5	41,75	1	,1	89,2
29,59	1	,1	31,6	42,00	10	1,0	90,2
29,75	1	,1	31,7	42,25	1	,1	90,3
30,00	68	6,5	38,2	42,50	25	2,4	92,7
30,41	1	,1	38,2	43,00	6	,6	93,2
30,50	1	,1	38,3	43,32	2	,2	93,4
30,75	1	,1	38,4	43,50	2	,2	93,6
30,84	1	,1	38,5	43,75	2	,2	93,8
31,00	23	2,2	40,7