

Mobilité intergénérationnelle des gains et du revenu des hommes au Canada : Étude basée sur les données longitudinales de l'impôt sur le revenu

par

Miles Corak* et Andrew Heisz**

N° 113

11F0019MPE N° 113

ISSN : 1200-5231

ISBN : 0-660-96033-8

Prix : 5 \$ l'unité; 25 \$, abonnement annuel

Analyse des entreprises et du marché du travail
24-H, immeuble R.-H.-Coats, Ottawa, K1A 0T6

*Statistique Canada (613) 951-9047

**Statistique Canada (613) 951-3748

Télécopieur : (613) 951-5403

Courrier électronique : coramil@statcan.ca

Octobre 1998

Une version antérieure du présent document a été distribuée, sous le titre «Unto the Sons: the Intergenerational Income Mobility of Canadian Men». Des ébauches du document ont également été présentées, dans le cadre de réunions de la *Society of Labor Economists*, de la *European Society for Population Economics* et de l'Association canadienne d'économie, ainsi qu'à l'occasion de colloques à l'Université Carleton, à l'Université Cornell, à l'OCDE, à Statistique Canada et à l'Université Syracuse. Les auteurs remercient Martin Browning, Kenneth Couch, Stephen Jenkins, Stephen Jones, Dean Lillard, Harry Paarsch, James Pesando, Geoff Rowe et Andreas Stich pour leur commentaires et suggestions. Ils remercient également Clive Loader pour le code S-plus qui a servi à l'analyse non paramétrique, Linda Standish et André St. Louis de Statistique Canada pour leurs contributions à la création de l'ensemble de données et Sophie Lefebvre pour son aide précieuse à la recherche.

Les vues exprimées dans ce document sont celles de l'auteur et ne sont pas nécessairement partagées par Statistique Canada.

Also available in English

Table des matières

<i>I. Introduction</i>	<i>1</i>
<i>II. Méthodologie et examen de la documentation</i>	<i>2</i>
<i>III. Résultats de la régression basée sur des spécifications linéaires</i>	<i>5</i>
<i>IV. Représentativité et exactitude des données</i>	<i>7</i>
<i>V. Non-linéarité</i>	<i>10</i>
<i>VI. Conclusion</i>	<i>14</i>
<i>Annexe</i>	<i>26</i>
<i>Bibliographie</i>	<i>29</i>

Résumé

Le but de la présente analyse est d'établir une estimation exacte du degré de mobilité intergénérationnelle du revenu au Canada. À partir des données sur l'impôt sur le revenu pour quelque 400 000 paires père-fils, nous en sommes arrivés à une élasticité intergénérationnelle des gains dont la valeur est d'environ 0,2. La mobilité des gains tend à être légèrement supérieure à celle du revenu; cependant, les méthodes non paramétriques révèlent une non-linéarité significative dans ces deux relations. Par ailleurs, la mobilité intergénérationnelle des gains est plus marquée à l'extrémité inférieure de la distribution qu'à son extrémité supérieure et elle affiche une forme en V inversé ailleurs dans la distribution. Dans le cas du revenu, la mobilité intergénérationnelle suit essentiellement le même profil, bien qu'elle soit beaucoup moins grande à l'extrémité supérieure de la distribution.

Classification JEL : D31, I132, J62

Mots-clés : Mobilité intergénérationnelle du revenu, répartition des revenus, régression non paramétrique

I. Introduction

Le but du présent document est d'examiner dans quelle mesure les personnes «héritent» de leur statut économique, c'est-à-dire d'évaluer le degré de la mobilité intergénérationnelle. Quel avantage économique y a-t-il à naître dans une famille riche? La pauvreté est-elle un état qui se transmet de générations à générations?

Les études sur l'estimation de la mobilité intergénérationnelle du revenu dans les pays de l'OCDE, qui sont de plus en plus nombreuses, s'appuient pour la plupart sur un modèle de régression linéaire. Ces études s'intéressent principalement aux biais potentiels associés aux différences qui existent (le plus souvent) entre le cycle de vie du père et celui du fils, ainsi qu'à l'erreur de mesure associée aux fluctuations transitoires du revenu et aux erreurs de déclaration. Bien que certaines théories ont trait précisément à la non-linéarité et que les analyses empiriques connexes basées sur les matrices de transition laissent croire à une non-linéarité des données, aucune étude n'a examiné en détail la possibilité qu'il y ait une erreur de spécification dans le modèle linéaire.

C'est là un des thèmes principaux de notre étude. Nous examinons la relation entre le revenu à l'âge adulte d'une cohorte d'hommes canadiens nés entre 1963 et 1966 et ayant vécu dans un ménage où le père était présent. L'analyse est basée sur les données administratives du régime fiscal canadien. Notre échantillon est donc très large, avec près de 400 000 paires père-fils. Ces données offrent un certain nombre d'avantages dont nous tentons de tirer profit, tout en cherchant à contourner leurs lacunes. Ainsi, les données administratives renseignent non seulement sur le montant du revenu, mais aussi sur sa composition. Il nous est donc possible d'examiner à la fois la mobilité intergénérationnelle des gains et du revenu. De plus, la large taille de notre échantillon nous permet d'utiliser des techniques non paramétriques pour étudier la non-linéarité dans l'étendue de la mobilité intergénérationnelle.

Nous commençons par faire un survol de la documentation existante, puis nous procédons à l'analyse proprement dite, en trois étapes. Premièrement, à partir de nos données, nous examinons les préoccupations soulevées dans les ouvrages antérieurs quant aux biais potentiels provenant des différences entre le cycle de vie des pères et celui des fils et à l'erreur de mesure associée au calcul du revenu permanent. Nous examinons ensuite les limites potentielles des données, c'est-à-dire la possibilité de biais dû à la sélection de l'échantillon, ainsi que la sous-déclaration des revenus. Enfin, nous analysons des profils non linéaires, d'abord en étudiant des matrices de transition par quartile, décile et percentile, puis en adoptant des techniques non paramétriques, nommément une version de l'estimation par le plus proche voisin.

Parmi les résultats que nous avons obtenus, mentionnons la grande mobilité intergénérationnelle entre les hommes au Canada. La corrélation intergénérationnelle moyenne entre les gains et le revenu marchand des pères et des fils est d'environ 0,2 – une valeur qui, jusqu'à récemment, était considérée comme faisant consensus aux États-Unis. L'utilisation de techniques non paramétriques a toutefois révélé une grande non-linéarité dans cette relation. Ainsi, la mobilité intergénérationnelle des gains est beaucoup plus forte à l'extrémité inférieure de la répartition des revenus qu'à son extrémité supérieure. En outre, l'élasticité entre les gains des pères et des fils, qui est d'environ 0,0 dans le percentile inférieur, augmente à 0,4 environ dans le percentile supérieur. L'élasticité suit toutefois une forme en V inversé dans les parties intermédiaires de la distribution. La mobilité du revenu présente un profil assez similaire, bien qu'elle soit beaucoup plus faible tout

en haut de la distribution. L'élasticité intergénérationnelle entre les revenus du père et du fils dans le percentile supérieur atteint presque 0,8.

L'utilisation d'un modèle des contraintes liées à l'emprunt peut aider à mieux comprendre certains de ces profils. Nous n'offrons cependant aucune interprétation de ces profils en regard de politiques particulières et laissons entendre que, pour ce faire, davantage de recherches devront être menées dans d'autres pays sur l'étendue et le profil de la non-linéarité dans la mobilité intergénérationnelle.

II. Méthodologie et examen de la documentation

La mobilité intergénérationnelle du revenu a été examinée sous deux angles complémentaires dans les études publiées jusqu'à maintenant : au moyen de l'estimation par les moindres carrés d'un modèle autorégressif qui établit un lien entre le revenu de l'enfant (à l'âge adulte) et le revenu du parent et par la construction d'une matrice de transition établissant un lien entre la position occupée par l'enfant et celle du parent dans la répartition des revenus.

Dans le premier cas, le revenu de l'enfant est exprimé comme une fonction (linéaire) du revenu du parent. Si Y représente le revenu permanent (généralement exprimé sous forme logarithmique), t la génération de l'enfant et $t-1$ la génération du parent, alors le processus type de génération de données prend la forme suivante :

$$Y_i(t) = \beta_0 + \beta_1 Y_i(t-1) + \varepsilon_i \quad (1)$$

où les données se situent au niveau individuel, i représente une paire père-fils et ε_i est une composante aléatoire dont la distribution correspond habituellement à $N(0, \sigma^2)$. Le terme constant représente la variation de revenu commune à la génération t , tandis que le coefficient β_1 indique dans quelle mesure les niveaux de revenu sont en accord avec ceux des parents; ce coefficient mesure donc l'étendue de la mobilité intergénérationnelle¹.

Obtenir une estimation exacte de β_1 est le but principal des études qui utilisent cette approche. Si la valeur de β_1 est inférieure à un, la répartition des revenus présente alors une régression vers la moyenne : tant que les parents dont le revenu se situe au-dessus (ou en dessous) de la moyenne auront des enfants dont le revenu se situera également au-dessus (ou en dessous) de la moyenne, l'écart par rapport à la moyenne ne sera pas aussi grand – un résultat peu surprenant dans les sociétés industrialisées avancées. Cependant, plus la valeur de β_1 est élevée (même si elle demeure inférieure à un), plus l'enfant est susceptible, à l'âge adulte, d'être dans la même situation économique que ses parents, donc plus la persistance des niveaux de revenu intergénérationnels est grande. Même de faibles valeurs de β_1 confèrent des avantages substantiels aux enfants des mieux nantis².

Bien que des modèles comme celui de l'équation (1) aient été utilisés avec des données de divers pays, les recherches les plus complètes sont celles qui ont été réalisées avec des données américaines. Altonji et Dunn (1991), Behrman et Taubman (1990), Mulligan (1997), Peters (1992), Shea (1997), Solon (1992) et Zimmerman (1992) en offrent quelques exemples récents; pour leur part, Becker et Tomes (1986) résument des études antérieures menées dans plusieurs pays ainsi qu'aux États-Unis. Selon ces derniers auteurs, 0,2 est la valeur de β_1 qui semble faire consensus. Ces résultats ont toutefois été contestés, certains alléguant qu'ils sont faussés par des erreurs de

mesure et des problèmes de sélection de l'échantillon. Ces problèmes ont été examinés par Atkinson, Maynard et Trinder (1983), Jenkins (1987) et Solon (1989,1992). Solon, par exemple, note qu'idéalement $Y_i(t)$ et $Y_i(t-1)$ dans l'équation (1) devraient être des mesures du revenu permanent; cependant, les chercheurs sont souvent forcés d'utiliser une mesure du revenu pour une année donnée, disons $\tilde{Y}_i(t-1) = Y_i(t-1) + v_i$, où v_i représente une perturbation transitoire du revenu. Il s'ensuit des erreurs dans les variables, qui font que le coefficient estimé ($\tilde{\beta}$) diffère du coefficient vrai par un facteur qui dépend du rapport de variances de v_i par rapport à Y_i , de sorte que $\tilde{\beta} (1 + \sigma_v^2 / \sigma_Y^2) = \beta_1$. Solon (1989) propose de gonfler la valeur de $\tilde{\beta}$ par un facteur de 1,3 à 1,8 lorsqu'un échantillon représentatif est utilisé. Il souligne toutefois qu'un grand nombre d'études sont basées sur des échantillons exagérément homogènes, de sorte que la variance empirique de Y sous-estime en fait la variance de la population et, de ce fait, augmente le biais dû à l'erreur de mesure. L'estimation par consensus (0,2) proposée par Becker et Tomes est basée sur des études sujettes à ces biais. Behrman et Taubman (1990), Lillard et Reville (1997), Mulligan (1997) et Solon (1992) corrigent les données en fonction de ces problèmes et estiment entre 0,4 et 0,5 la valeur de β_1 , les premiers proposant une valeur privilégiée aussi élevée que 0,6. Zimmerman (1992) propose lui aussi une valeur de 0,4, mais Peters (1992) - qui utilise les mêmes données - obtient une estimation qui se situe entre 0,1 et 0,2, tout comme Couch et Dunn (1997) qui, eux, utilisent les données de la PSID.

Cependant, l'étendue de la documentation n'est pas aussi vaste que le nombre d'études le laisse supposer. En effet, les données pour toutes les enquêtes américaines ne sont basées que sur deux enquêtes différentes (la *Panel Study of Income Dynamics* (PSID) et la *National Longitudinal Survey - NLS*) et sur des échantillons parfois très petits. Solon (1992), par exemple, utilise un échantillon de 348 paires père-fils tirées de la PSID, tandis que Zimmerman (1992) en utilise 876 de la NLS. De même, certains des échantillons utilisés par Cooper *et coll.* (1993) comptent moins de 100 observations. Couch et Dunne (1997) et Couch et Lillard (1997) soulignent que les règles d'exclusion utilisées pour la construction de l'échantillon d'analyse peuvent influencer considérablement sur les estimations obtenues. Or comme la plupart des chercheurs utilisent les mêmes ensembles de données, la petite taille des échantillons contribue -- étant donné les problèmes de spécification et de mesure -- à la grande diversité des estimations obtenues.

Les recherches menées dans les autres pays sont plus limitées, quoique leur nombre soit en hausse. Atkinson, Maynard et Trinder (1983), par exemple, présentent des résultats basés sur une seule ville britannique, selon lesquels la valeur de β_1 varie entre 0,4 et 0,45 (selon la façon dont le revenu est mesuré); pour leur part, Dearden, Machin et Reed (1997) en arrivent à une valeur qui se situe entre 0,4 et 0,6, avec un échantillon plus représentatif. Gustafsson (1994), et Björklund et Jänti (1997a), proposent quant à eux une valeur entre 0,2 et 0,25, à partir de données suédoises. Björklund et Jänti (1997b), Mulligan (1997, chapitre 7) et Solon (1997) analysent des enquêtes détaillées basées sur des données internationales; à notre avis, toutefois, l'aspect sans doute le plus frappant est le manque de résultats pour le Canada. À notre connaissance, la question de la mobilité intergénérationnelle a été peu étudiée par les économistes au Canada³. Dans une version préliminaire de la présente recherche (Corak et Heisz, 1995), nous avons obtenu une estimation légèrement inférieure à 0,2; cette étude avait toutefois peu tenu compte des problèmes de mesure inhérents aux données. Fortin et Lefebvre (1998) en arrivent à un résultat similaire, avec des données d'enquête.

La deuxième approche proposée pour l'étude de la mobilité intergénérationnelle est celle de la matrice de transition, qui complète l'utilisation de modèles autorégressifs. Cette approche consiste à répartir la population en groupes de taille égale, classés par rang de revenu, et à présenter la répartition des parents et des enfants selon ces groupes. La matrice de transition a l'avantage de donner une description plus détaillée de la mobilité intergénérationnelle. Peters (1992), de même que Dearden, Machin et Reed (1997), offrent des exemples types. Dans les deux études, environ le tiers des fils dont le père se situait dans le quartile inférieur ont accédé à la moitié supérieure de la répartition des revenus. Ces auteurs ont également constaté qu'il y avait moins de mobilité au haut et au bas de la distribution, les fils nés de pères aux deux extrémités de la répartition des revenus étant beaucoup plus susceptibles d'occuper la même position que leur père à l'âge adulte, que les fils dont les pères avaient un revenu se situant dans le deuxième ou troisième quartile. Ce phénomène peut toutefois constituer également un inconvénient. Comme l'ont fait remarquer Atkinson, Maynard et Trinder (1983, p. 83), le profil non linéaire pourrait refléter en partie des valeurs plafond et plancher, en haut et en bas de la matrice : en d'autres mots, il ne peut y avoir mobilité ascendante pour ceux qui se trouvent à l'extrémité supérieure à la naissance, pas plus qu'il ne peut y avoir mobilité descendante pour ceux qui occupent le quartile inférieur. Le degré d'immobilité au sommet et au bas est donc de ce fait exagéré. Ces limites ne se posent lors de l'utilisation de modèles de régression. En fait, la possibilité de non-linéarité a été soulignée par des chercheurs utilisant des modèles d'autorégression. Solon (1992) ajoute le carré du revenu du père à l'équation (1) mais constate, en grande partie à cause de la petite taille de l'échantillon, que cet ajout n'est pas significatif. De même, Mulligan (1997, p. 193) ne trouve aucune preuve de non-linéarité et suggère que cela pourrait être dû, en partie, au fait qu'il y a sous-représentation, dans la PSID, des personnes qui se situent au sommet de la répartition des revenus. À l'inverse, Behrman et Taubman (1990) - eux aussi utilisant la PSID - et Lillard (1998) et Peters (1992), qui ont utilisé la NLSY, constatent que les termes d'ordre supérieur sont statistiquement significatifs.

La non-linéarité est un volet important de la théorie de la dynamique du revenu intergénérationnel. Les principaux travaux dans ce domaine sont ceux de Becker et Tomes (1979, 1986), Becker (1991) Bénabou (1994) et Durlauf (1994, 1996). Becker et Tomes (1986) proposent un modèle de la mobilité intergénérationnelle basé sur un modèle du capital humain, qui reflète les décisions prises par les parents de dépenser leur revenu ou de l'investir dans leurs enfants. Ces auteurs incorporent dans ce modèle les contraintes du marché financier et de ce fait soulèvent la possibilité que la corrélation entre le revenu familial et le revenu de l'enfant à l'âge adulte soit plus forte dans les familles à faible revenu que dans celles à revenu élevé. Mulligan (1997) présente une description précise de ce modèle et de ses répercussions en regard de la non-linéarité. L'hypothèse proposée par Durlauf au sujet de la persistance intergénérationnelle de la pauvreté repose sur les effets de proximité associés à la qualité et au financement des études. Certains travaux empiriques ont également incorporé des éléments de ces modèles, pour inclure des variables additionnelles dans l'estimation de l'équation (1) ou pour stratifier des échantillons de façons spéciales; aucun consensus ne semble toutefois se dégager de ces travaux quant à la nature et à l'étendue des profils non linéaires⁴.

III. Résultats de la régression basée sur des spécifications linéaires

Solon (1992) et Zimmerman (1992) sont ceux qui se sont intéressés de plus près aux problèmes de mesure et aux aspects de la méthodologie susceptibles d'introduire un biais dans les estimations obtenues à partir de l'équation (1). Leurs travaux sont à l'origine de la réévaluation du degré de mobilité intergénérationnelle aux États-Unis et laissent croire que ce degré est beaucoup moindre qu'on ne le croyait initialement. C'est ce qui constitue le point de départ de notre analyse.

Contrairement aux travaux de ces chercheurs et à la plupart des études décrites dans la documentation, nous utilisons l'information extraite des dossiers fiscaux. Notre échantillon est très large et n'est pas sujet aux problèmes d'attrition ou aux erreurs de déclaration. Nous examinons une cohorte d'hommes qui étaient âgés de 16 à 19 ans en 1982, qui ont produit une déclaration de revenus entre 1982 et 1986 (alors qu'ils vivaient toujours à la maison) et dont le «père» vivait au foyer durant l'année de la déclaration⁵. Les couplages à l'intérieur de la famille, établis à partir des dossiers fiscaux, sont le résultat en partie de la construction du fichier familial T1 (T1FF) par Statistique Canada, lequel est décrit en annexe. Ce fichier permet notamment d'établir un lien entre les numéros d'assurance sociale (NAS) du père et du fils. Il nous est également possible, à partir des fichiers d'impôt sur le revenu, de remonter jusqu'en 1978 (première année pour laquelle l'information du T1 est disponible) pour connaître le revenu du père, et d'aller jusqu'en 1995 pour obtenir le revenu du fils, alors âgé entre 29 et 32 ans. Pour être retenus dans l'échantillon, les fils doivent avoir produit une déclaration de revenu en 1995. Nous nous sommes également inspirés de la méthode de Solon (1992), en retenant uniquement les fils aînés lorsque le même père était apparié à plusieurs fils. L'analyse est basée sur les gains et le revenu marchand total (défini comme le revenu avant impôt provenant de toutes les sources, incluant les gains, le revenu net d'un travail indépendant et le revenu de biens)⁶. Toutes les données sont exprimées en dollars constants de 1986, l'IPC servant de déflateur⁷.

La première question que nous examinons à la lumière de ces données est celle du biais associé aux mesures du revenu annuel, en regard d'une mesure du revenu permanent. Selon Solon (1992), l'utilisation de la moyenne pluriannuelle du revenu du père réduit l'importance du caractère transitoire du revenu et augmente le rapport signal-bruit. Le nombre maximal d'années utilisées pour établir la moyenne est de cinq (1978 à 1982).

Le tableau 1 présente les statistiques descriptives associées à chacune des deux mesures pour le sous-ensemble de nos données dans lequel le revenu moyen du père est supérieur ou égal à 1 \$ durant les cinq années à l'étude. En ce qui a trait aux gains, décrits au panel A, nous obtenons 339 951 paires père-fils au sein desquelles les gains moyens du père varient entre un dollar et plus de 1,829 millions de dollars (la moyenne étant de 31 388 \$). Nous obtenons par ailleurs 389 348 paires père-fils pour lesquelles le revenu marchand total est connu, le revenu moyen étant ici légèrement supérieur (35 586 \$).

L'équation d'estimation inclut à la fois l'âge du père et l'âge du fils, pour tenir compte du fait que ceux-ci ne sont pas observés au même stade durant leur vie. Le modèle que nous utilisons est donc le suivant :

$$Y_i(t) = \beta_0 + \beta_1 \bar{Y}_i(t-1) + \beta_2 \hat{\text{Age}} \text{ du fils} + \beta_3 \hat{\text{Age}} \text{ du fils}^2 + \beta_4 \hat{\text{Age}} \text{ du père} + \beta_5 \hat{\text{Age}} \text{ du père}^2 + \varepsilon_t \quad (2)$$

où \bar{Y}_i représente le logarithme naturel du revenu, parfois dont on a fait la moyenne sur plusieurs années consécutives, et où l'âge est mesuré en années⁸. Le tableau 2 présente les estimations de β_1 obtenues par régression des moindres carrés, pour les gains et le revenu marchand total. Chaque panel du tableau présente les résultats de la régression, obtenus en utilisant chacune des moyennes possibles du revenu du père : la première colonne indique les résultats basés sur les données annuelles, la deuxième présente les résultats établis à partir des moyennes sur deux années successives et ainsi de suite, jusqu'à l'estimation unique basée sur la moyenne quinquennale. La taille de l'échantillon varie pour chaque type d'estimation et elle est indiquée, avec l'erreur-type, sous l'estimation du coefficient. Le biais par défaut des estimations portant sur une seule année est évident : les estimations augmentent à mesure que le nombre d'années sur lequel est établie la moyenne augmente. En ce qui a trait aux gains pour le panel A, les estimations tendent graduellement vers 0,131 (valeur sur cinq ans). Le ratio entre ce coefficient et ceux basés sur des données annuelles varie de 1,15 à 1,3, ce qui correspond à peu près à la borne inférieure de l'intervalle proposé par Solon en l'absence de problèmes dus à un échantillon exagérément homogène. Les deux estimations basées sur les moyennes sur quatre ans se situent à deux erreurs-types de ce chiffre, tout comme les trois estimations basées sur les moyennes triennales. Un profil général similaire se dégage pour le revenu marchand total (voir panel B du tableau 1). Dans ce dernier cas, les estimations portant sur une seule année sont de 30 à 60 p.100 inférieures à celles basées sur la moyenne quinquennale. L'élasticité entre le revenu marchand du père et celui du fils, qui est de 0,194 dans le cas de la moyenne quinquennale, est supérieure à celle observée entre les gains. En résumé, ces conclusions semblent indiquer qu'il est nécessaire d'utiliser au moins la moyenne triennale du revenu du père pour l'analyse et qu'une période de cinq ans serait assez longue pour réduire le biais dû aux fluctuations transitoires du revenu.

La robustesse de ces conclusions en regard des règles utilisées pour la sélection de l'échantillon, la façon dont sont apportées les corrections en fonction des différences dans le cycle de vie, ainsi que le choix de la variable représentant le revenu du père sont évalués au tableau 3. Tous les résultats indiqués dans ce tableau sont basés sur des moyennes quinquennales de la mesure du revenu du père. Les quatre rangées étiquetées 1, dans chacun des panels du tableau 3, correspondent aux résultats figurant à la dernière colonne du tableau 2. Trois conclusions principales se dégagent des résultats indiqués. Premièrement, les règles de sélection de l'échantillon semblent influencer largement sur l'estimation de l'élasticité. Cet effet semble toutefois se limiter à l'inclusion ou non, avant le calcul de la moyenne, des personnes ayant un revenu nul (ou négatif) dans l'échantillon. Si le père doit avoir réalisé un revenu d'au moins un dollar durant chacune des cinq années sur lesquelles la moyenne est calculée (plutôt que d'utiliser une moyenne d'au moins 1 \$), l'élasticité entre les gains augmente alors de 0,131 à 0,228 (voir rangées 1 et 2 du panel A). L'élasticité varie peu au-delà de ce niveau, alors qu'il y a relèvement du seuil d'exclusion, et elle atteint 0,242 à un seuil de 3 000 \$.

En outre, l'utilisation du seuil de 1 \$ donne lieu à des estimations similaires de l'élasticité entre les gains et entre les revenus, qui s'établissent entre 0,23 et 0,24. Deuxième point, les différences entre le cycle de vie des pères et celui des fils ne semblent pas avoir un effet important. Si aucune correction n'est incluse dans le modèle de régression pour tenir compte de l'effet de l'âge, l'élasticité estimée ne diffère pas de celles obtenues par inclusion d'une valeur quadratique dans l'âge du père et du fils. Cette conclusion résiste aux autres spécifications, à l'inclusion de facteurs pour tenir compte de l'effet de l'état matrimonial du fils, ainsi qu'à l'utilisation d'un échantillon formé de l'ensemble des fils (plutôt que seulement des fils âgés). Lorsque l'échantillon se limite aux fils âgés de 32 ans, l'élasticité des gains est légèrement supérieure (0,140), tandis qu'elle est légèrement inférieure (0,117) lorsque l'échantillon se limite aux 29 ans. En revanche, l'élasticité du revenu marchand ne semble pratiquement pas touchée par ces considérations. Enfin, l'élasticité des gains du fils par rapport au revenu marchand total du père a une valeur estimative de 0,192, ce qui équivaut essentiellement à l'élasticité entre les revenus⁹.

La sensibilité des résultats sur les gains aux règles de sélection de l'échantillon ressort également du panel D du tableau, qui met en contraste les estimations par la méthode des moindres carrés et les régressions par quantile vers la moyenne. Les résultats sur les gains ne résistent pas au changement de la méthode d'estimation ; par contre, les résultats sur le revenu demeurent, eux, essentiellement inchangés. La méthode des moindres carrés et la régression par quantile donnent des résultats qui sont beaucoup plus près l'un de l'autre, lorsque le père doit avoir réalisé un revenu d'au moins un dollar durant chacune des cinq années -- les valeurs obtenues sont alors de 0,228 contre 0,206.

En résumé, lorsqu'on tient compte de certaines des préoccupations soulevées dans la documentation — à savoir les biais associés à l'erreur de mesure, les corrections en fonction du cycle de vie et les règles d'exclusion pour déterminer l'échantillon — notre meilleure estimation de l'élasticité entre les gains du père et ceux du fils dans l'économie canadienne se chiffre à environ 0,2 — cette valeur pouvant être légèrement supérieure ou inférieure selon les règles d'exclusion utilisées pour la construction de l'échantillon.

IV. Représentativité et exactitude des données

Reste toutefois à déterminer dans quelle mesure les données administratives se comparent aux données d'enquête. Il semble y avoir deux volets connexes à cette question. Le premier a trait à la représentativité de notre échantillon. Même si notre ensemble de données est vaste, cela n'écarte pas pour autant les risques de biais de sélection. D'une part, pour établir un lien père-fils, il faut que le fils ait produit une déclaration de revenus alors qu'il vivait toujours dans la maison familiale, à une période ou l'autre entre 1982 et 1986. Cela laisse sous-entendre que le fils a un NAS et que ce numéro peut être apparié à celui du père. Sont exclus de notre analyse les fils qui vivaient dans un ménage où le père était absent, qui ont quitté la maison familiale avant de produire une déclaration de revenus ou qui n'ont pas produit de déclaration en 1995. Par conséquent, il est probable que les membres de la cohorte qui n'ont pas été actifs sur le marché du travail durant leur adolescence, qui ont quitté la maison familiale avant d'aller sur le marché du travail, qui vivaient dans une famille monoparentale dirigée par une femme ou qui ont peut-être été peu actifs à l'âge adulte¹⁰ soient sous-représentés dans notre ensemble de données. Dans tous ces cas, il est plausible que les données sous-représentent les membres de la cohorte qui sont sujets à avoir un faible revenu à l'âge adulte. Nos règles de sélection excluent également les immigrants dont les pères sont arrivés au pays après 1986.

Le deuxième volet concerne l'exactitude des gains et des revenus déclarés. Contrairement aux données d'enquête, il existe, dans un système d'impôt progressif, un incitatif manifeste à déclarer un revenu inférieur aux autorités fiscales. Cependant, ce qui importe surtout ici – compte tenu de nos objectifs – c'est de savoir si ces incitatifs sont très forts, s'ils varient le long de la répartition des revenus et s'ils ont évolué au fil des ans, ayant de ce fait une incidence différente sur les pères et les fils. Il est difficile d'évaluer dans quelle mesure les revenus sont non déclarés ou sous-déclarés, mais il y a lieu de croire que les pères dans notre échantillon aient pu être davantage tentés de le faire que leurs fils. Durant la fin des années 70 jusqu'au milieu des années 80, il existait jusqu'à 10 tranches d'imposition; en 1988, le régime fiscal a été modifié et il ne compte plus que trois tranches d'imposition. Quelle que soit leur position occupée dans la répartition des revenus, les pères ont donc sans doute été plus enclins à déclarer des revenus légèrement moindres que ne l'ont été leurs fils, ce qui laisse sous-entendre que la mobilité intergénérationnelle sera surestimée.

Il convient de souligner toutefois que les problèmes de cette nature ne sont pas uniques à nos données. Faisant référence au problème de sélection de l'échantillon, Solon (1992, p. 398) souligne que l'érosion de l'échantillon a sans doute nui à la représentativité de la PSID pour les analyses intergénérationnelles, les personnes à faible revenu et à revenu élevé étant plus susceptibles de quitter l'enquête. Ainsi, l'échantillon qu'il utilise pour son analyse ne correspond pas tout à fait à 60 p. 100 de la cohorte initiale des fils. L'information sur le revenu extraite des enquêtes n'est pas elle non plus à l'abri des problèmes de qualité. En fait, ce sont ces préoccupations qui ont amené Statistique Canada à coupler les répondants d'enquête à leurs dossiers fiscaux, ainsi qu'à utiliser les données fiscales comme source d'information sur le revenu pour l'*Enquête sur la dynamique du travail et du revenu*, une enquête longitudinale canadienne similaire à la PSID aux États-Unis. Michaud *et coll.* (1995) comparent l'information sur le revenu déclarée par un groupe de répondants aux données indiquées sur leurs dossiers fiscaux et constatent que les réponses à l'enquête ont tendance à omettre certains montants du revenu et à sous-déclarer le revenu provenant de certaines sources, principalement le revenu d'un travail indépendant et les revenus d'intérêt et de dividendes. Par contre, le montant moyen du revenu salarial déclaré lors de l'enquête est essentiellement le même que celui indiqué dans les fichiers d'impôt et, bien qu'environ 2,4 p. 100 des répondants à l'enquête ont déclaré un revenu salarial alors qu'aucune donnée correspondante ne figurait dans leur fichier d'impôt, exactement la même proportion n'a pas déclaré de revenu à l'enquête même si cette information apparaissait dans leur fichier d'impôt (Michaud *et coll.* 1995, tableaux 6 et 7).

Afin de pouvoir évaluer l'étendue de ces problèmes, nous présentons au tableau 4 une comparaison entre nos données administratives et celles obtenues dans le cadre d'une enquête représentative menée à l'échelle nationale, l'*Enquête sur les finances des consommateurs* (EFC). Il s'agit d'une enquête transversale, menée chaque année en mai, et où les répondants doivent indiquer le niveau et la composition de leur revenu durant l'année précédente. Nous avons prélevé un échantillon parmi tous les hommes qui étaient âgés de 29 à 32 ans durant l'année d'enquête 1995, calculé une mesure des gains et du revenu marchand total la plus près possible des définitions utilisées pour notre analyse, puis nous avons comparé la distribution ainsi obtenue à trois échantillons provenant des données administratives : l'ensemble des déclarants en 1995 qui se situent dans la même catégorie d'âge, le sous-ensemble de ces déclarants qui ont pu être couplés à un père et, enfin, le sous-ensemble représentant les aînés parmi les appariements multiples au même père (et qui forment la base de notre échantillon d'analyse).

Selon l'EFC, il y avait 1 015 million d'hommes âgés de 29 à 32 ans en 1995. Dans ce groupe d'âge, 907 137 hommes ont produit une déclaration de revenus en 1995 et 497 242 ont pu être couplés à un père. (Ce dernier chiffre inclut les personnes qui ont déclaré des revenus nuls ou négatifs ou dont les pères ont déclaré des revenus nuls ou négatifs). En d'autres mots, la condition qui consiste à tenter d'apparier le père et le fils par le biais des fichiers d'impôt, de manière à connaître le revenu du père lorsque le fils est à un âge assez jeune, nous permet de retracer environ 55 % des déclarants de notre cohorte et environ 49 % de l'ensemble de la cohorte¹¹. Deux facteurs peuvent donc causer l'introduction d'un biais de sélection dans notre échantillon d'analyse : premièrement, ce ne sont pas toutes les personnes qui produisent une déclaration de revenus et, deuxièmement, les déclarants ne sont pas tous appariés à leur père.

Nos données tendent à surestimer légèrement le revenu médian de la population. Selon l'EFC, les gains médians de cette population sont de 18 700 \$, soit près de 800 \$ de plus que les gains médians des déclarants; ils sont toutefois de 2 750 \$ à 2 900 \$ inférieurs à ceux des déclarants avec père et des déclarants les plus âgés avec père. Un profil similaire se dégage dans le cas du revenu marchand. Cependant, les différences les plus marquées entre les données d'enquête et les données administratives se remarquent aux extrémités de la répartition des revenus. Le percentile inférieur et même le quintile inférieur ont tendance à être plus élevés dans l'EFC, alors que le quintile et le percentile supérieurs sont plus faibles. Les données d'enquête sous-représentent clairement les personnes qui se situent au sommet de la répartition des revenus : à titre d'exemple, le revenu maximal selon l'EFC est de 157 000 \$, alors qu'il est supérieur à quatre millions de dollars dans les fichiers d'impôt. Parallèlement, les fichiers d'impôt permettent de retracer beaucoup plus de personnes ayant des gains et des revenus minimums. La plus grande queue de la distribution, en particulier dans sa extrémité supérieure, contribue à la plus grande inégalité transversale des données fiscales, comme en témoigne l'écart-type du logarithme des gains et du revenu¹².

Il est difficile pour nous d'évaluer l'impact de l'exclusion des non-déclarants de l'analyse; nous pouvons en revanche évaluer l'incidence de l'impossibilité d'apparier la totalité des déclarants à leur père. Pour étudier cela, nous divisons le nombre total des déclarants âgés de 29 à 32 ans en 1995 en deux groupes, ceux appariés à un père (et donc disponibles pour l'analyse) et ceux non appariés à un père (et donc exclus de l'analyse). Afin de déterminer si le biais potentiel de sélection a une incidence sur l'estimation de la valeur de β_1 , nous utilisons un modèle de régression, avec sélection d'échantillon par la méthode d'estimation à deux degrés de Heckman. La règle de sélection est modélisée en utilisant une équation des probits. Comme cela requiert une estimation du maximum de vraisemblance, nous choisissons à peu près un échantillon au 1/10 en retenant seulement les personnes dont le NAS se termine par le chiffre 5. Les résultats détaillés obtenus à partir de ce modèle des probits sont disponibles sur demande; en résumé, toutefois, disons que les personnes dans l'échantillon ont tendance à être plus âgées, sont beaucoup moins susceptibles de vivre à Toronto ou à Montréal et, aussi dans une moindre mesure, en Colombie-Britannique et sont plus susceptibles d'être mariées que célibataires, vivant en union de fait ou divorcées. La sous-représentation des grandes régions métropolitaines dans notre échantillon s'explique du fait que la plupart des immigrants ont tendance à s'installer à Toronto, Montréal ou Vancouver; or comme nous l'avons mentionné précédemment, ces immigrants ne sont pas inclus dans l'analyse à moins d'être arrivés au pays avant 1986 et que leur père était alors sur le marché du travail. Les autres résultats sont en accord avec notre hypothèse a priori selon laquelle il y a sous-représentation, dans notre échantillon, des personnes sujettes à avoir un faible revenu.

Les résultats du deuxième degré de l'estimation sont présentés au tableau 5. Sont également présentés, pour fins de comparaison, les résultats d'une régression des moindres carrés basée sur le même échantillon au 10%. Les gains et les revenus des pères sont mesurés sur la base de la moyenne quinquennale. Les résultats de la régression des moindres carrés, obtenus à partir de cet échantillon plus petit, ne sont pas statistiquement différents de ceux indiqués au tableau 2, pour ce qui est de l'élasticité des gains et des revenus du fils par rapport à ceux du père. Plus important encore, les résultats de cette régression sont essentiellement les mêmes que ceux obtenus par la méthode des moindres carrés. Alors que le rapport inverse de Mill est négatif et statistiquement significatif, seules les estimations du coefficient pour l'âge du fils diffèrent, d'un modèle à l'autre¹³.

Nous en concluons que nos règles de sélection de l'échantillon n'introduisent pas un biais majeur dans les résultats. Cela n'évalue pas cependant l'incidence de l'exclusion des non-déclarants. Peut-être devrions-nous insister ici sur le fait que Fortin et Lefebvre (1998), qui utilisent une enquête représentative et une méthodologie différente de la nôtre, arrivent essentiellement aux mêmes résultats. Leur échantillon est formé de 2 500 à 3 400 paires père-fils tirées de l'*Enquête sociale générale* de 1986 et 1994. Cette enquête ne contient pas d'information sur le revenu des parents, mais elle recueille des données sur leur niveau de scolarité et leur profession. Ces auteurs utilisent l'information sur le revenu par profession provenant du recensement du Canada, pour estimer le revenu d'emploi des parents. Leurs méthodes par variables instrumentales donnent une élasticité intergénérationnelle entre les pères et les fils qui se chiffre à 0,157 (avec une erreur-type de 0,034), pour la cohorte née entre 1955 et 1969 (Fortin et Lefebvre, 1998, tableau 4.4). Ces auteurs présentent également une quantité de résultats pour d'autres cohortes ainsi que différents instruments et, dans tous les cas, ils arrivent à un résultat de l'ordre de 0,2. En d'autres mots, ces auteurs arrivent essentiellement aux mêmes résultats que nous, pour un groupe d'hommes à peu près du même âge que notre cohorte, mais en utilisant des données représentatives à l'échelle nationale et des techniques économétriques différentes (lesquelles sont probablement incohérentes vers le haut). Cela semble venir corroborer notre conclusion selon laquelle les biais de sélection ne semblent pas avoir un effet significatif sur nos résultats.

V. Non-linéarité

Ce qui se dégage de ces conclusions—à savoir qu'il existe une grande mobilité intergénérationnelle au Canada et que celle-ci est peut-être encore plus marquée qu'aux États-Unis—ressort encore davantage lorsqu'on examine les matrices de transition comme celles décrites précédemment. Pour établir ces matrices, nous utilisons l'échantillon basé sur une moyenne quinquennale des gains du père et une moyenne triennale des gains du fils (1993 à 1995), tous deux corrigés en fonction des effets du cycle de vie¹⁴. Une matrice de transition par quartile est présentée au premier panel du tableau 6. Les personnes nées de pères dont le revenu se situe dans la queue de la distribution des revenus sont certainement plus susceptibles d'avoir elles aussi un revenu qui se situe dans la queue de la distribution. Environ 35 % des fils nés de pères se situant dans le quartile supérieur avaient eux aussi des gains dans le quartile supérieur et environ 33 % de ceux nés de pères dans le quartile inférieur se situaient dans le même quartile, à l'âge adulte. Les mêmes proportions, au bas de la distribution des revenus, s'établissent respectivement à 33 % et 42 %. À l'inverse, la partie intermédiaire de la distribution du revenu se caractérise par une mobilité presque parfaite.

Une désagrégation plus fine révèle de nettes asymétries aux extrémités de la répartition des revenus, un phénomène qu'illustre la matrice de transition par décile au panel B du tableau. Dix-huit pour cent des personnes dont les pères avaient un revenu dans le décile supérieur de la distribution se trouvaient dans le même décile, alors que près de 16 % de celles vivant dans une famille dans le décile inférieur restent au bas de la distribution. Parallèlement, la probabilité de reculer d'un décile, pour les personnes au sommet, est à peu près la même que celle d'accéder à un décile supérieur pour les personnes au bas de la distribution.

En plus de représenter la non-linéarité du processus de mobilité, les probabilités plus élevées de non-déplacement pour les personnes qui se situent tout en haut ou tout en bas de la distribution des revenus du père sont également reliées aux valeurs planchers et plafonds dans les matrices de transition. Si l'effet de plancher et de plafond se limitait, disons à la matrice de transition par quartile — c'est-à-dire si le profil non linéaire était dû uniquement au fait que les personnes au sommet ne peuvent progresser davantage vers le haut alors que celles tout en bas ne peuvent descendre davantage — on s'attendrait alors à ce que les probabilités de transition n'affichent des pics significatifs que dans les déciles supérieur et inférieur. En fait, la situation varie quelque peu. Des probabilités nettement plus élevées apparaissent pour les quartiles supérieur et inférieur, et aussi pour les déciles inférieur et supérieur. En d'autres mots, la désagrégation plus détaillée (c'est-à-dire l'utilisation de déciles) atténue vraiment les pics. Parallèlement, toutefois, les neuvième et deuxième déciles demeurent plus élevés que les déciles voisins. Le même phénomène se produit avec les percentiles. La figure 1 illustre une série de profils de la matrice de transition par percentile de gains. On remarquera que les pics les plus aigus se trouvent dans les percentiles supérieur et inférieur, ce qui dénote un effet de plancher et de plafond. Cependant, les fils dont les pères se situent dans le 95^e percentile ont tendance à grimper encore davantage dans la répartition des revenus. Cela laisse croire que le processus sous-jacent pourrait être non linéaire. Ce profil peut avoir son homologue à l'extrémité inférieure de la distribution des revenus. À titre d'exemple, les personnes dont le père se situe dans le 5^e percentile ont une tendance, quoique plus faible, à reculer encore davantage dans la distribution des revenus. Les percentiles révèlent en général que la probabilité de mobilité ascendante est plus faible pour les fils issus de milieux à faible revenu, alors que la probabilité de mobilité ascendante est plus élevée pour les fils de milieux à revenu élevé. Par exemple, les probabilités de transition à l'intérieur du quartile inférieur de la distribution des revenus, pour les fils dont le père se situe au 5^e et au 10^e percentiles, sont généralement supérieures à l'intervalle de confiance de deux écarts-types de l'un ou l'autre côté des 10 %, alors que ces probabilités sont généralement égales ou inférieures, pour ce qui est du passage au quartile supérieur. Inversement, ceux dont le père se situe dans le 95^e percentile sont plus susceptibles d'appartenir au quintile supérieur.

Ces résultats font ressortir la nécessité de réévaluer les résultats de la régression en tenant compte de la non-linéarité. Nous utilisons des techniques non paramétriques, à la fois pour évaluer la validité du modèle loglinéaire et étudier la nature des non-linéarités possibles. En termes plus précis, nous utilisons les estimateurs du plus proche voisin basés sur les régressions des moindres carrés à pondération locale¹⁵. Il faut pour ce faire que la relation entre les variables indépendantes et dépendantes (dites lisses) soit continue, mais aucune hypothèse explicite n'est formulée quant à la forme fonctionnelle qui les relie. En bref, la procédure consiste à définir un bloc d'observations autour d'une valeur particulière de la variable indépendante, disons $Y_o(t-1)$. Dans le cas des estimateurs du plus proche voisin, ce bloc est défini comme une fraction du nombre total d'observations (N) et est désigné intervalle de mesure (α). Chacune des αN observations à

l'intérieur de ce bloc est pondérée par des facteurs de voisinage, $w(Y_i)$, de sorte que les observations qui se rapprochent le plus de $Y_o(t-1)$ obtiennent le poids le plus élevé. Une régression des moindres carrés généralisés est ensuite estimée, en utilisant un polynôme de la variable explicative. Les coefficients estimés à partir de cette régression sont utilisés pour prévoir la valeur de la variable dépendante à $Y_o(t-1)$ et nous obtenons ainsi un point sur la courbe lisse. Une nouvelle valeur de $Y_o(t-1)$ et ses observations proches sont ensuite choisies et le processus est répété¹⁶.

Il y a trois choix à faire concernant la modélisation : la fonction de pondération, le degré du polynôme et l'intervalle de mesure. Selon Cleveland et Loader (1995), le choix de la fonction de pondération n'est pas crucial, à la condition de respecter certaines caractéristiques générales, et nous avons retenu leur raisonnement en utilisant la fonction tricubique¹⁷. Notre principale préoccupation concerne la première dérivée de la relation fonctionnelle entre le revenu du père et celui du fils. Comme cette valeur est déterminée à partir des estimations du degré du polynôme, nous utilisons une forme fonctionnelle relativement souple, en l'occurrence la forme cubique. Des polynômes d'ordre supérieur ajouteraient sans doute peu à l'estimation de la première dérivée, mais ils augmenteraient en revanche le fardeau de calcul; cependant, des polynômes d'ordre inférieur peuvent omettre une partie de la courbure de la courbe lisse, au risque de provoquer une erreur de spécification de la première dérivée.

Des trois, le choix de α est le plus important. Ce choix traduit le compromis inhérent aux régressions locales, à savoir entre le biais et l'efficacité. Un intervalle de mesure plus restreint signifie un biais moins important, mais une erreur-type plus élevée. Comme nous cherchons à obtenir une idée de la non-linéarité dans la courbe lisse et à évaluer le caractère adéquat de la spécification linéaire utilisée dans la documentation, nous sommes plus enclins à privilégier une estimation plus efficace, au risque d'introduire un biais. Nous utilisons une version modifiée du critère d'information d'Akaike (AIC) et l'inspection graphique pour choisir la valeur la mieux appropriée de α ¹⁸.

Pour réduire le fardeau de calcul, nous utilisons un échantillon environ au 1/10 de nos données, en reprenant les mêmes critères de sélection que ceux mentionnés précédemment¹⁹. La figure 2 présente un diagramme de dispersion des données sur les revenus et les gains. Il s'agit de données brutes, exprimées sous forme de logarithmes naturels du revenu (corrigées en fonction des effets du cycle de vie). Il existe manifestement une corrélation positive entre le revenu du père et celui du fils. (La ligne droite représente l'ajustement par les moindres carrés.) Cependant, il semble également y avoir concentration de points de données pour les très faibles niveaux de gains exprimés sous forme logarithmique et ajustés selon l'âge (inférieurs à environ -8) pour le père, le fils et les deux. Pour un faible segment de la population, les gains du père exprimés sous forme logarithmique et ajustés selon l'âge sont très faibles et il n'existe aucune corrélation apparente avec les revenus du fils. De même, un petit nombre de fils gagnent un salaire minimal, en dépit du revenu de leur père. Ce profil est également présent, quoique dans une moindre mesure, dans le cas du revenu marchand total. Ces points irréguliers expliquent pourquoi les résultats des régressions entre les gains (tableau 2) sont moins robustes que ceux obtenus par les régressions entre revenus. Nous croyons que ces observations ne fournissent pas une approximation exacte des gains ou des revenus permanents²⁰. Dans ce qui suit, nous avons supprimé les observations pour lesquelles la valeur logarithmique ajustée selon le cycle de vie est inférieure à -6.

Les deux panels de la figure 3 présentent un graphique de la courbe lisse qui établit un lien entre les gains du père et ceux du fils, ainsi que de la première dérivée connexe. La valeur optimale de α (pour réduire au minimum le AIC) est de 0,84. Cette valeur est utilisée pour établir la courbe lisse représentée par le panel A. (La ligne droite est l'ajustement par les moindres carrés.) Le deuxième panel présente l'élasticité correspondante. La valeur prévue à partir du modèle de régression loglinéaire se situe clairement à l'extérieur de l'intervalle de confiance de 95 % de l'estimation du plus proche voisin (représentée par les lignes pointillées), ce qui laisse croire qu'il s'agit d'une erreur de spécification dans la relation entre les gains du père et ceux du fils. De toute évidence, l'élasticité établie à partir du modèle non paramétrique n'est pas constante. Dans ce panel, les lignes pointillées verticales représentent les percentiles inférieur et supérieur de la distribution des gains des pères. Aux parties inférieures de la distribution, l'élasticité est nulle, voire négative, puis elle augmente pour atteindre une valeur de près de 0,4 au sommet. Sur la majeure partie de la distribution, l'élasticité présente une forme en V inversé, atteignant un sommet d'environ 0,3 près de la médiane²¹.

La figure 4 illustre les élasticités calculées selon trois modèles. Les résultats présentés au panel B de la figure 3 y sont représentés par la ligne EE, alors que la ligne MM représente l'élasticité entre le revenu marchand total du père et celui du fils et la ligne EM désigne l'élasticité entre les gains du fils et le revenu marchand total du père. Dans chaque cas, l'intervalle est défini d'après le AIC (pour EM, l'intervalle est de 0,84 et il est de 0,76 pour MM). Dans tous les cas, l'élasticité présente une forme en V inversé entre le 1^{er} et le 99^e percentiles, mais l'élasticité entre les revenus est supérieure aux autres valeurs, sur l'ensemble de la distribution. Cette élasticité diffère largement de ces autres valeurs au percentile supérieur et même en dessous et atteint une valeur de près de 0,8 au sommet, ce qui est plus du double de l'élasticité entre les gains ou de l'élasticité entre les gains et les revenus. Un profil similaire, quoique moins prononcé, apparaît au bas de la distribution des revenus.

Nous avons davantage confiance dans nos résultats au sommet de la distribution des revenus que dans ceux au bas de la distribution. Comme nous l'avons indiqué précédemment, les règles de sélection utilisées pour construire nos données donnent lieu à une sous-représentation des personnes susceptibles d'être à faible revenu. Bien que ce biais de sélection n'influe pas sur les résultats des moindres carrés obtenus par le modèle linéaire, on ne peut en dire autant du modèle non paramétrique. On ne peut donc pas tirer de conclusions définitives quant à la nature de l'élasticité intergénérationnelle du revenu sous le premier percentile et, à notre avis, la nature du processus qui détermine la mobilité intergénérationnelle du revenu tout en haut et tout en bas de la distribution exige généralement une analyse plus poussée, car les estimateurs non paramétriques ont tendance à être moins fiables aux extrémités des distributions.

Cela étant dit, le profil en V inversé semble être une conclusion robuste qui, avec les profils observés dans l'extrémité supérieure de la queue, peut être interprété en regard du modèle des contraintes d'emprunt. Comme l'expliquent Becker et Tomes (1986) et Mulligan (1997), ce modèle suppose une élasticité non linéaire entre les gains des parents et ceux des enfants. Les parents investissent dans le capital humain de leurs enfants, jusqu'au point où le rendement marginal entre l'investissement dans le capital humain et l'investissement financier est le même; au-delà de ce point, les investissements se font sous forme de transferts financiers. Les parents forcés d'emprunter peuvent être incapables d'atteindre le montant optimal d'investissement dans le capital humain, de sorte que la corrélation entre les gains des parents et ceux des enfants sera plus élevée

qu'elle ne le serait autrement. Cependant, le montant optimal d'investissement dans le capital humain dépend également des capacités des enfants. Les parents les plus susceptibles d'être forcés d'emprunter sont les parents à faible revenu dont les enfants ont de grandes aptitudes. S'il est vrai également que les aptitudes varient en fonction du revenu du parent — plus le revenu du parent est élevé, plus les aptitudes (en moyenne) des enfants le seront également — alors le profil en V inversé que nous avons observé n'est pas plausible. Sur la moitié inférieure de la distribution, l'élasticité augmente, parce que le revenu des parents est relativement bas et que les aptitudes des enfants augmentent; la contrainte d'emprunt devient alors obligatoire pour une proportion de plus en plus grande de la population. Sur la partie supérieure de la distribution, par contre, l'élasticité diminue parce que le revenu des parents devient peu à peu suffisamment élevé pour financer les études postsecondaires et donc la contrainte d'emprunt s'atténue, même si les aptitudes des enfants continuent d'augmenter. À mesure que le revenu des parents augmente, les transferts financiers aux enfants deviennent de plus en plus importants, comme en témoigne le fait que l'élasticité entre les revenus des pères et des fils est beaucoup plus forte que celle entre leurs gains.

VI. Conclusion

Notre analyse du degré de mobilité intergénérationnelle des revenus chez les hommes au Canada nous amène à conclure que l'élasticité intergénérationnelle entre les gains et les revenus du père et du fils se situe en moyenne à 0,2 - une estimation qui, jusqu'à tout récemment, était considérée comme la valeur faisant consensus aux États-Unis. Cette conclusion résiste à plusieurs mesures et questions méthodologiques soulevées dans la documentation, incluant l'erreur de mesure associée aux fluctuations transitoires du revenu et les différences entre le cycle de vie des pères et celui des fils. Cependant, l'élasticité entre les gains peut être sensible aux règles d'exclusion utilisées pour définir l'échantillon d'analyse. Comme nous utilisons les données administratives associées au régime fiscal canadien, nous envisageons également la possibilité que le biais de sélection de l'échantillon puisse influencer sur les résultats. En fait, nous avons constaté que nos données ont tendance à sous-représenter les personnes en marge du marché du travail, mais cela n'a aucune incidence sur l'estimation de l'élasticité intergénérationnelle des revenus du père par rapport à ceux du fils. Parallèlement, toutefois, nous avons observé également une non-linéarité significative dans la relation entre les revenus du père et ceux du fils. L'utilisation de méthodes non paramétriques nous porte à croire que les modèles loglinéaires supposent une erreur de spécification du processus de génération des données. Sur la moitié inférieure de la distribution des revenus du père, l'élasticité entre les gains du père et ceux du fils augmente de 0 à 0,3 environ, puis diminue à 0,1 dans la moitié supérieure. Elle augmente à nouveau tout en haut de la distribution des revenus, pour atteindre une valeur d'environ 0,4 dans le percentile maximal. L'élasticité entre le revenu marchand total suit une courbe similaire mais est légèrement supérieure et atteint presque 0,8 au sommet de la distribution. Ce profil peut être interprété en regard d'un modèle de contraintes d'emprunt, selon lequel les parents investissent dans le capital humain de leurs enfants, si l'on présume qu'il y a corrélation positive entre les aptitudes de l'enfant et le revenu des parents.

Mulligan (1997) souligne également que les politiques peuvent avoir une incidence sur le degré de mobilité intergénérationnelle, que l'on pense à l'impôt progressif sur la succession, à l'impôt progressif sur le revenu ou aux subventions pour les études collégiales et universitaires. On peut s'attendre à ce que tous ces facteurs aient une incidence sur les investissements dans le capital humain que font les parents à l'égard de leurs enfants. Si le degré de mobilité intergénérationnelle ne varie pas d'un pays à l'autre, alors il se peut que les différences entre ces politiques aient peu

d'impact. Cependant, cela n'est peut-être pas le cas si des facteurs sociaux plus larges associés aux valeurs familiales et aux structures sociales interviennent également. Le Canada est sans doute un pays particulièrement approprié à étudier et à comparer aux nombreuses études réalisées aux États-Unis car, comme Card et Freeman (1993) l'ont fait remarqué, ces deux pays sont très similaires sur le plan du marché du travail, des traditions culturelles et des niveaux de vie. Ces similitudes font naître la possibilité de distinguer plus clairement les effets qui résultent des différences entre les politiques et les institutions de ces deux pays sur leur marché du travail respectif. Entre autres différences, mentionnons l'absence d'impôt sur la succession au Canada, le fait que les taux d'imposition sur le revenu y soient généralement plus élevés et plus progressifs qu'aux États-Unis et que le Canada est généralement reconnu pour son meilleur accès à un enseignement postsecondaire de qualité.

Les résultats de notre spécification loglinéaire n'offrent aucune indication voulant que l'élasticité intergénérationnelle des gains ou des revenus soit supérieure à 0,25, celle-ci étant une estimation nettement inférieure aux valeurs de 0,4 à 0,5 obtenues par de nombreux chercheurs ayant utilisé des données américaines. Ces résultats, combinés à des renseignements plus détaillés sur les institutions, pourraient aider à mieux comprendre l'impact de ces politiques sur la mobilité intergénérationnelle. Notre analyse fait ressortir l'utilité des données administratives à cette fin, en soulignant la possibilité que le modèle loglinéaire soit une erreur de spécification de la relation entre les gains du père et ceux du fils; du même coup, elle met en évidence la nécessité d'adopter des techniques économétriques souples pour étudier les non-linéarités potentielles dans cette relation.

Tableau 1

STATISTIQUES DESCRIPTIVES : GAINS ET REVENU MARCHAND TOTAL

	Moyenne	Écart-type	Minimum	1 ^{er} percentile	25 ^e percentile	Médiane	75 ^e percentile	99 ^e percentile	Maximum
A. Total des gains	Nombre d'observations=339 951								
Gains du fils en 1995	25 219	20 960	1	128	14 504	24 345	33 512	68 667	3 692 884
Gains du père 1978-1982	31 388	22 246	1	1,2	20 470	29 750	39 597	93 518	1 829 229
Logarithme des gains du fils en 1995	9 796	1 239	0	4 852	9 582	10 100	10 420	11 137	15 122
Logarithme des gains du père 1978-1982	9 981	1 459	0	0 182	9 927	10 301	10 587	11 446	14 419
B. Revenu marchand total	Nombre d'observations=389 348								
Revenu marchand du fils en 1995	25 386	24 158	1	314	13 842	23 822	33 345	77 324	4 161 596
Revenu marchand du père 1978-1982	35 586	35 275	1	2 627	22 148	31 372	42 327	130 539	8 105 156
Logarithme du revenu marchand du fils en 1995	9 820	1 085	0	5 749	9 535	10 078	10 415	11 256	15 241
Logarithme du revenu marchand du père 1978-1982	10 268	0 738	0	7 874	10 005	10 354	10 653	11 779	15 908

Tableau 2

EFFET DES FLUCTUATIONS TRANSITOIRES SUR LES ESTIMATIONS PAR LES MOINDRES CARRÉS DE L'ÉLASTICITÉ INTERGÉNÉRATIONNELLE DES GAINS ET DES REVENUS

	Une année	seule	Moyenne sur deux ans	Moyenne sur trois ans	Moyenne sur quatre ans	Moyenne sur cinq ans
A. Logarithme des gains du père						
1978	0,114 (0,003) [339 558]		0,106 (0,003) [343 551]			
1979	0,110 (0,003) [340 824]		0,113 (0,003) [342 433]	0,123 (0,003) [341 527]	0,126 (0,003) [341 181]	
1980	0,101 (0,002) [341 283]		0,120 (0,003) [341 843]	0,123 (0,003) [342 038]	0,129 (0,003) [341 044]	0,131 (0,003) [339 951]
1981	0,100 (0,002) [342 901]		0,127 (0,003) [337 793]	0,135 (0,003) [337 748]		
1982	0,104 (0,003) [333 498]					
B. Logarithme du revenu marchand total du père						
1978	0,147 (0,003) [400 216]		0,168 (0,003) [397 559]			
1979	0,146 (0,003) [402 078]		0,172 (0,003) [398 871]	0,187 (0,003) [393 965]	0,191 (0,003) [391 352]	
1980	0,140 (0,003) [403 347]		0,167 (0,003) [401 065]	0,180 (0,003) [396 174]	0,184 (0,003) [394 086]	0,194 (0,003) [389 348]
1981	0,133 (0,003) [405 380]		0,155 (0,003) [403 799]	0,173 (0,003) [399 026]		
1982	0,119 (0,002) [406 075]					

() indique les erreurs-types d'après l'estimateur hétéroscédatique robuste de White; [] indique la taille de l'échantillon

Tableau 3

ÉLASTICITÉ INTERGÉNÉRATIONNELLE POUR DIVERS ÉCHANTILLONS ET SPÉCIFICATIONS :
GAINS ET REVENU MARCHAND DU PÈRE ET DU FILS

	Gains		Revenu marchand total	
	Élasticité	Erreur-type	Élasticité	Erreur-type
A. Règles de sélection de l'échantillon				
1. Revenu moyen sur cinq ans ≥ 1 \$	0,131	(0,003)	0,194	(0,003)
2. Revenu durant chacune des cinq années ≥ 1 \$	0,228	(0,006)	0,236	(0,004)
3. Revenu durant chacune des cinq années ≥ 100 \$	0,227	(0,003)	0,231	(0,003)
4. Revenu durant chacune des cinq années $\geq 1\ 000$ \$	0,237	(0,003)	0,239	(0,002)
5. Revenu durant chacune des cinq années $\geq 3\ 000$ \$	0,242	(0,003)	0,236	(0,002)
B. Ajustements en fonction du cycle de vie				
1. Valeur quadratique dans l'âge du père et du fils	0,131	(0,003)	0,194	(0,003)
2. Aucune correction en fonction de l'âge	0,131	(0,003)	0,193	(0,003)
3. Variables fictives pour l'âge du fils	0,131	(0,003)	0,194	(0,003)
4. Valeur quartique dans l'âge du père et du fils	0,132	(0,003)	0,196	(0,003)
5. Fils aîné né en 1963	0,140	(0,006)	0,197	(0,006)
6. Fils aîné né en 1966	0,117	(0,007)	0,197	(0,008)
7. Ensemble des frères et soeurs	0,129	(0,003)	0,196	(0,003)
8. Correction en fonction de l'état matrimonial	0,134	(0,003)	0,193	(0,003)
C. Choix de la variable explicative				
1. Gains	0,131	(0,003)		
2. Revenu total marchand	0,192	(0,004)		
D. Méthode d'estimation				
Revenu moyen sur cinq ans ≥ 1 \$				
1. Moindres carrés	0,131	(0,003)	0,194	(0,003)
2. Régression vers la médiane	0,068	(0,001)	0,192	(0,002)
Revenu durant chacune des cinq années ≥ 1 \$				
3. Moindres carrés	0,228	(0,006)	0,236	(0,004)
4. Régression vers la médiane	0,206	(0,002)	0,225	(0,002)

Tableau 4

COMPARAISON ENTRE LES STATISTIQUES SUR LE REVENU TIRÉES DE L'ENQUÊTE SUR LES FINANCES DES CONSOMMATEURS ET DES FICHIERS D'IMPÔT SUR LE REVENU HOMMES ÂGÉS DE 29 À 32 ANS (GAINS ET REVENU MARCHAND TOTAL DE 1995, MESURÉS EN DOLLARS DE 1986)

	Nombre	Moyenne	Écart-type	Asymétrie	Aplatissement	Minimum	Percentile inférieur	Quintile inférieur	Quartile inférieur	Médiane	Quartile supérieur	Quintile supérieur	Percentile supérieur	Maximum
A. GAINS														
1. Total des salariés														
EFC	1 015 000	19 619	15 142	0,7	0,7	0	0	0	6 861	18 727	29 963	44 944	59 925	137 828
Total des déclarants	907 137	19 107	20 142	28,8	3 109,6	0	0	0	4 818	17 948	29 007	45 130	62 665	3 692 884
Déclarants avec père	497 242	21 853	20 594	28,9	3 273,5	0	0	0	8 738	21 479	31 572	47 345	65 956	3 692 884
Déclarants plus âgés	448 196	21 987	21 027	30,0	3 339,9	0	0	0	8 805	21 622	31 738	47 545	66 431	3 692 884
2. Gains > 0														
EFC	860 000	23 141	13 746	0,8	1,4	0,75	698	2 996	12 734	22 472	31 461	46 442	62 468	137 828
Total des déclarants	769 089	22 537	20 031	34,2	3 726,9	1	1	1 588	11 161	21 336	30 954	46 650	65 135	3 692 884
Déclarants avec père	439 453	24 728	20 219	34,3	3 967,0	1	85	2 572	13 939	23 801	33 025	48 515	68 040	3 692 884
Déclarants plus âgés	396 053	24 881	20 695	35,4	4 007,9	1	86	2 572	14 059	23 963	33 196	48 747	68 453	3 692 884
3. Logarithme des gains														
EFC	860 000	9,79	0,89	-2,1	8,7	-0,29	6,55	8,01	9,45	10,02	10,36	10,75	11,04	11,83
Total des déclarants	769 089	9,58	1,49	-3,9	20,6	0	7,37	7,57	9,32	9,97	10,34	10,75	11,08	15,12
Déclarants avec père	471 280	9,76	1,28	-4,3	26,3	0	4,44	7,85	9,54	10,08	10,41	10,79	11,13	15,12
Déclarants plus âgés	396 053	9,77	1,28	-4,3	26,3	0	4,5	7,85	9,55	10,08	10,41	10,79	11,13	15,12
B. REVENU MARCHAND														
1. Revenu total														
EFC	1 015 000	22 131	15 908	1,5	6,5	-22 548	0	0	11 158	21 034	30 592	47 926	72 420	157 303
Total des déclarants	907 137	20 593	22 835	35,0	3 700,5	-328 069	0	0	7 528	18 985	29 607	46 572	69 584	4 161 596
Déclarants avec père	497 242	23 637	23 893	36,3	3 850,9	-149 981	0	0	11 441	22 419	32 279	49 105	74 834	4 161 596
Déclarants plus âgés	448 196	23 794	24 479	37,2	3 873,9	-149 981	0	0	11 530	22 558	32 445	49 369	75 509	4 161 596
2. Revenu > 0														
EFC	954 000	23 552	15 329	1,7	7,7	0,75	749	3 314	12 446	22 362	31 461	48 554	72 584	157 303
Total des déclarants	826 944	22 652	22 891	37,9	4 006,8	1	40	1 818	10 699	20 834	30 727	47 512	71 790	4 161 596
Déclarants avec père	471 280	24 979	23 810	38,5	4 110,8	1	288	3 038	13 443	23 391	32 921	49 707	76 090	4 161 596
Déclarants plus âgés	424 884	25 140	24 421	39,4	4 116,9	1	288	3 046	13 533	23 536	33 095	49 996	76 869	4 161 596
3. Logarithme du revenu														
EFC	954 000	9,79	0,91	-2,10	9,40	-0,29	6,62	8,11	9,43	10,02	10,36	10,79	11,19	11,97
Total des déclarants	826 944	9,60	1,37	-3,67	20,28	0	3,69	7,51	9,28	9,94	10,33	10,77	11,18	15,24
Déclarants avec père	471 280	9,80	1,10	-3,56	22,84	0	5,66	8,02	9,51	10,06	10,40	10,81	11,24	15,24
Déclarants plus âgés	424 884	9,80	1,10	-3,55	22,68	0	5,66	8,02	9,51	10,07	10,41	10,82	11,25	15,24

Nota : EFC désigne l'Enquête sur les finances des consommateurs de 1995. Toutes les statistiques tirées de cette enquête sont basées sur les données pondérées d'un échantillon de 2 587 hommes âgés de 29 à 32 ans.

Total des déclarants fait référence à l'information fiscale sur tous les hommes de 29 à 32 ans qui ont produit une déclaration de revenus (formule T1) pour l'année d'imposition 1995.

Déclarants avec père fait référence aux données fiscales sur les hommes de 29 à 32 ans qui ont été apparés à un père en 1982.

Déclarants plus âgés désignent un sous-échantillon des déclarants avec père formé des aînés de tous les enfants d'un même père.

Tableau 5

RÉGRESSIONS DES MOINDRES CARRÉS ET DE SÉLECTION DE L'ÉCHANTILLON DE L'ÉLASTICITÉ DES GAINS ET DES REVENUS DU FILS PAR RAPPORT AUX GAINS ET AUX REVENUS DU PÈRE

	Gains		Revenu marchand total	
	Non corrigés en fonction du biais de sélection de l'échantillon	Corrigés en fonction du biais de sélection de l'échantillon	Non corrigés en fonction du biais de sélection de l'échantillon	Corrigés en fonction du biais de sélection de l'échantillon
Terme constant	0,848 (6,45)	14,8 (6,99)	-0,479 (5,19)	14,7 (5,86)
Logarithme des gains/revenus du père	0,145 (0,005)	0,145 (0,005)	0,194 (0,008)	0,199 (0,008)
Âge du fils	0,416 (0,422)	-0,391 (0,457)	0,468 (0,340)	-0,415 (0,383)
Âge du fils au carré	-0,00612 (0,007)	0,00625 (0,007)	-0,00685 (0,006)	0,00657 (0,006)
Âge du père	0,0225 (0,012)	0,0259 (0,012)	0,0170 (0,0094)	0,0214 (0,0094)
Âge du père au carré	-0,00025 (0,0001)	-0,00027 (0,0001)	-0,00018 (0,0001)	-0,00021 (0,0001)
λ		-0,925 (0,067)		-1,01 (0,055)
R ² corrigé	0,0303	0,0365	0,0205	0,0309
Nombre d'observations	33 660		38 637	

Nota : Les échantillons utilisés sont formés de toutes les paires père-fils dans lesquelles le NAS du fils (aîné) se termine par le chiffre 5 et qui ne renferment aucune valeur manquante pour les variables utilisées pour l'équation de sélection des probits. Les résultats des probits sont indiqués en annexe. λ est l'inverse du rapport de Mill calculé à partir de ces résultats. Les gains et les revenus du père sont définis par le logarithme de la moyenne annuelle entre 1978 et 1982. Cette moyenne doit être égale ou supérieure à un pour être incluse dans l'échantillon.

() indique l'erreur-type de l'estimation.

Tableau 6

MATRICES DE TRANSITION DES GAINS PÈRE-FILS

A. Matrice de transition par quartile

		Gains du fils			
		<i>Supérieur</i>	<i>Troisième</i>	<i>Deuxième</i>	<i>Inférieur</i>
Gains du père	<i>Supérieur</i>	0,345	0,248	0,205	0,202
	<i>Troisième</i>	0,271	0,269	0,241	0,219
	<i>Deuxième</i>	0,212	0,263	0,273	0,252
	<i>Inférieur</i>	0,172	0,220	0,281	0,327

B. Matrice de transition par décile

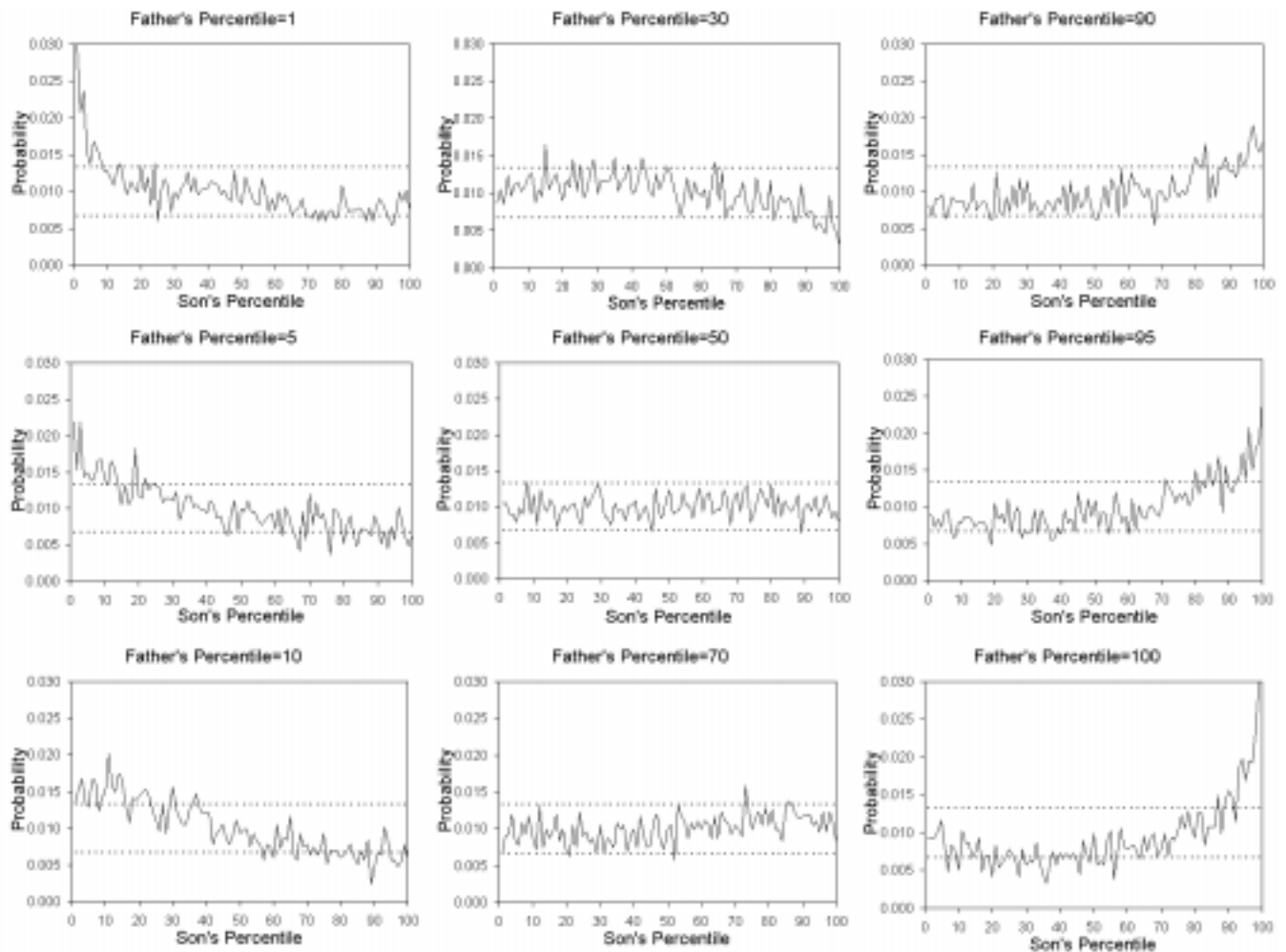
		Gains du fils									
		<i>Sup.</i>	<i>9^e</i>	<i>8^e</i>	<i>7^e</i>	<i>6^e</i>	<i>5^e</i>	<i>4^e</i>	<i>3^e</i>	<i>2^e</i>	<i>Inférieur</i>
Gains du père	<i>Sup.</i>	0,180	0,132	0,112	0,096	0,087	0,082	0,075	0,076	0,078	0,084
	<i>9^e</i>	0,148	0,128	0,116	0,102	0,095	0,083	0,082	0,083	0,084	0,079
	<i>8^e</i>	0,122	0,123	0,116	0,107	0,102	0,094	0,090	0,085	0,082	0,079
	<i>7^e</i>	0,108	0,114	0,111	0,107	0,105	0,098	0,093	0,093	0,088	0,085
	<i>6^e</i>	0,093	0,104	0,110	0,109	0,107	0,103	0,099	0,096	0,091	0,087
	<i>5^e</i>	0,083	0,096	0,106	0,109	0,104	0,106	0,104	0,104	0,095	0,094
	<i>4^e</i>	0,071	0,085	0,095	0,106	0,110	0,112	0,110	0,109	0,104	0,099
	<i>3^e</i>	0,066	0,078	0,085	0,096	0,107	0,116	0,116	0,117	0,111	0,109
	<i>2^e</i>	0,060	0,070	0,076	0,086	0,098	0,109	0,122	0,121	0,130	0,127
	<i>Inf.</i>	0,069	0,070	0,074	0,083	0,087	0,097	0,110	0,117	0,137	0,158

Nota : Données corrigées en fonction du cycle de vie, pour un échantillon de 334 018 observations père-fils. Les gains des fils correspondent à la moyenne calculée sur trois ans (1993 à 1995) et normalisée selon l'âge, conformément à la méthode décrite dans le texte; les gains des pères correspondent à la moyenne sur cinq ans (1978 à 1982), eux aussi normalisés selon l'âge. L'indice d'immobilité pour la matrice de transition par quartile est de 0,304. Dans le cas de la matrice par décile, il est de 0,126.

Source : Calculs des auteurs basés sur les données administratives de l'impôt sur le revenu, Statistique Canada

Figure 1

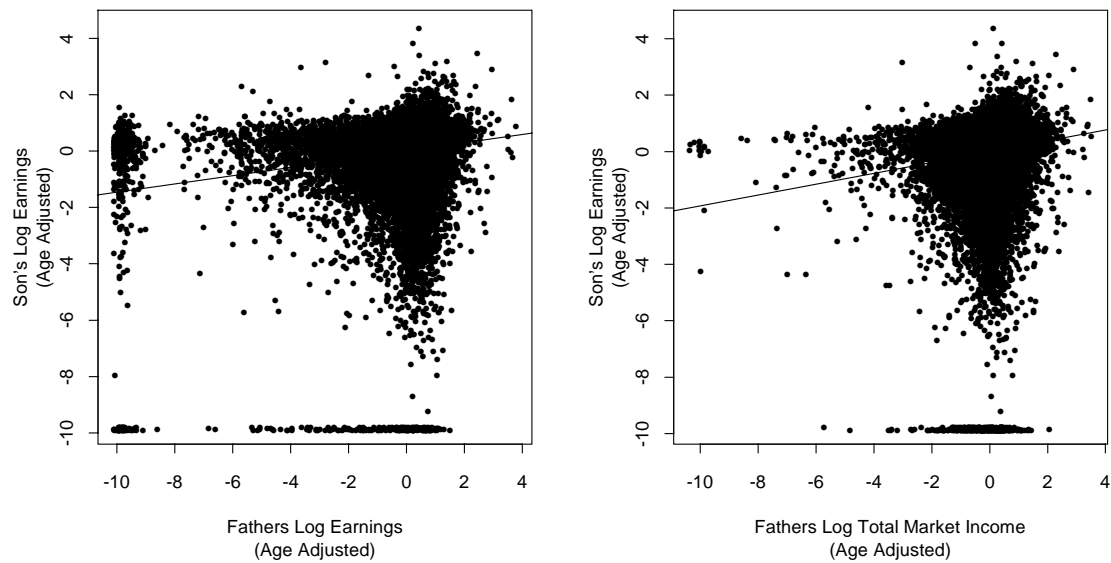
PROBABILITÉ DU PERCENTILE DES GAINS DU FILS COMPTE TENU DU PERCENTILE DES GAINS DU PÈRE



Nota : Les revenus du père et du fils sont corrigés selon l'âge, tel qu'indiqué dans le texte. La taille de l'échantillon est de 334 018. Le 1^{er} percentile dans le graphique du coin gauche supérieur (percentile du père =1) est hors-limite à 0,0329. Les 99^e et 100^e percentiles dans le graphique du coin droit inférieur (percentile du père =100) sont eux aussi hors-limite à 0,0371 et 0,0895.

Figure 2

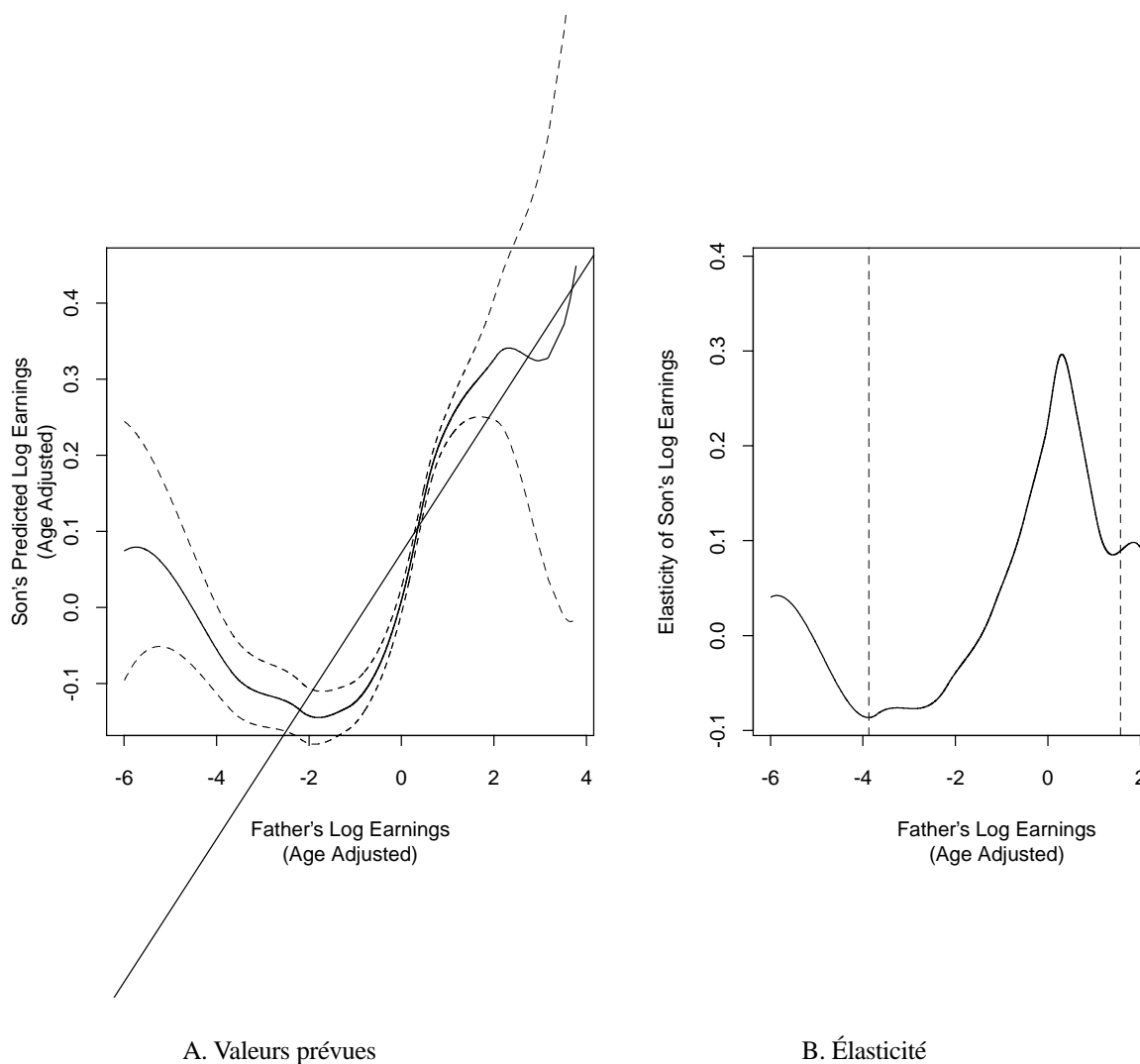
RAPHIQUE DE DISPERSION DU LOGARITHME DES GAINS DU FILS PAR RAPPORT AU LOGARITHME DES GAINS ET DU REVENU MARCHAND DU PÈRE



Nota : Les données sont corrigées selon l'âge et exprimées en logarithme, les lignes droites représentant les estimations par les moindres carrés. Le panel A compte 33 660 observations et la pente (erreur-type) de la droite des moindres carrés est égale à 0,146 (0,0047). Au panel B, le nombre d'observations est de 36 039 et la pente (erreur-type) de la droite des moindres carrés est de 0,193 (0,0094).

Figure 3

ESTIMATIONS DU PLUS PROCHE VOISIN DE LA RELATION ENTRE LES GAINS DU PÈRE ET CEUX DU FILS



Nota : Le panel A représente la relation prévue entre le logarithme des gains du père et ceux du fils corrigés en fonction du cycle de vie, par la technique d'estimation décrite dans le texte, avec un intervalle de mesure de 0,84. Les lignes pointillées se situent à deux erreurs-types de la relation estimée et la ligne droite représente l'estimation par les moindres carrés. Le panel B représente l'élasticité estimée, les lignes pointillées verticales représentant les 1^{er} et 99^e percentiles du logarithme des gains, corrigés en fonction de l'âge du père.

Figure 4

ÉLASTICITÉ DES GAINS ET DU REVENU MARCHAND TOTAL DU FILS PAR RAPPORT AUX GAINS ET AU REVENU MARCHAND TOTAL DU PÈRE



Nota : La ligne MM représente l'élasticité entre le revenu marchand total du fils et celui du père; la ligne EM fait référence à l'élasticité entre les gains du fils et le revenu marchand total du père et la ligne EE représente l'élasticité entre les gains du père et du fils. Tous les résultats sont basés sur l'estimation du plus proche voisin décrite dans le texte. Pour MM, l'intervalle de mesure est de 0,76; pour EM et EE, il est de 0,84. Les lignes pointillées verticales représentent les 1^{er} et 99^e percentiles du revenu marchand total du père, corrigé selon l'âge et exprimé en logarithme.

Annexe

Élaboration des données et définitions

Le couplage des numéros d'assurance sociale (NAS) des pères et des fils se fait à partir du fichier familial T1 (T1FF). Le NAS est un identificateur personnel unique et le T1FF est un ensemble de données établi à partir des dossiers T1 qui ont été traités dans le but d'apparier les membres de chaque famille de déclarants. (Les formulaires T1 sont les principales déclarations annuelles de revenus produites par les particuliers au Canada et le T1FF réunit l'univers des déclarants.) Diverses stratégies d'appariement sont utilisées pour identifier les membres de la famille et des processus d'imputation sont utilisés pour inclure également les membres non déclarants de la famille et compléter l'information manquante. Les couples (incluant les couples époux-épouse et les unions de fait) sont liés à partir des codes NAS du répondant et de celui de son (sa) conjoint(e) lorsque celui-ci est indiqué sur le formulaire T1, et aussi à partir de l'information sur le nom et l'adresse. L'appariement des enfants aux parents est fait à partir des champs de nom et d'adresse. Harris et Lucaci (1994) décrivent plus en détail la construction du T1FF.

Des paires père-fils ont été extraites du T1FF pour plusieurs années (1982 à 1986), afin d'améliorer la couverture de notre échantillon et de réduire l'ampleur du biais de sélection de l'échantillon. L'algorithme est présenté à la figure A-1. Dans les cas où un même enfant était apparié à différents pères, à différentes années, nous avons retenu l'appariement père-fils le moins récent. Seuls les pères et les fils pour lesquels aucune donnée n'a pas été imputée ont été retenus. (Le père n'est pas nécessairement le père biologique, mais doit être considéré plutôt comme l'homme qui est chef du ménage.) Les fils se limitent à ceux nés entre 1963 et 1966. Trois catégories de fils ont été exclues, parce que le lien père-fils n'a pu être établi: il s'agit des fils qui n'ont pas produit de déclaration de revenus entre 1982 et 1986 alors qu'ils vivaient toujours à la maison; de ceux qui ont produit une déclaration de revenus mais qui ont été appariés à une famille sans père et enfin de ceux qui ont produit une déclaration mais qui n'ont pas été appariés à une famille.

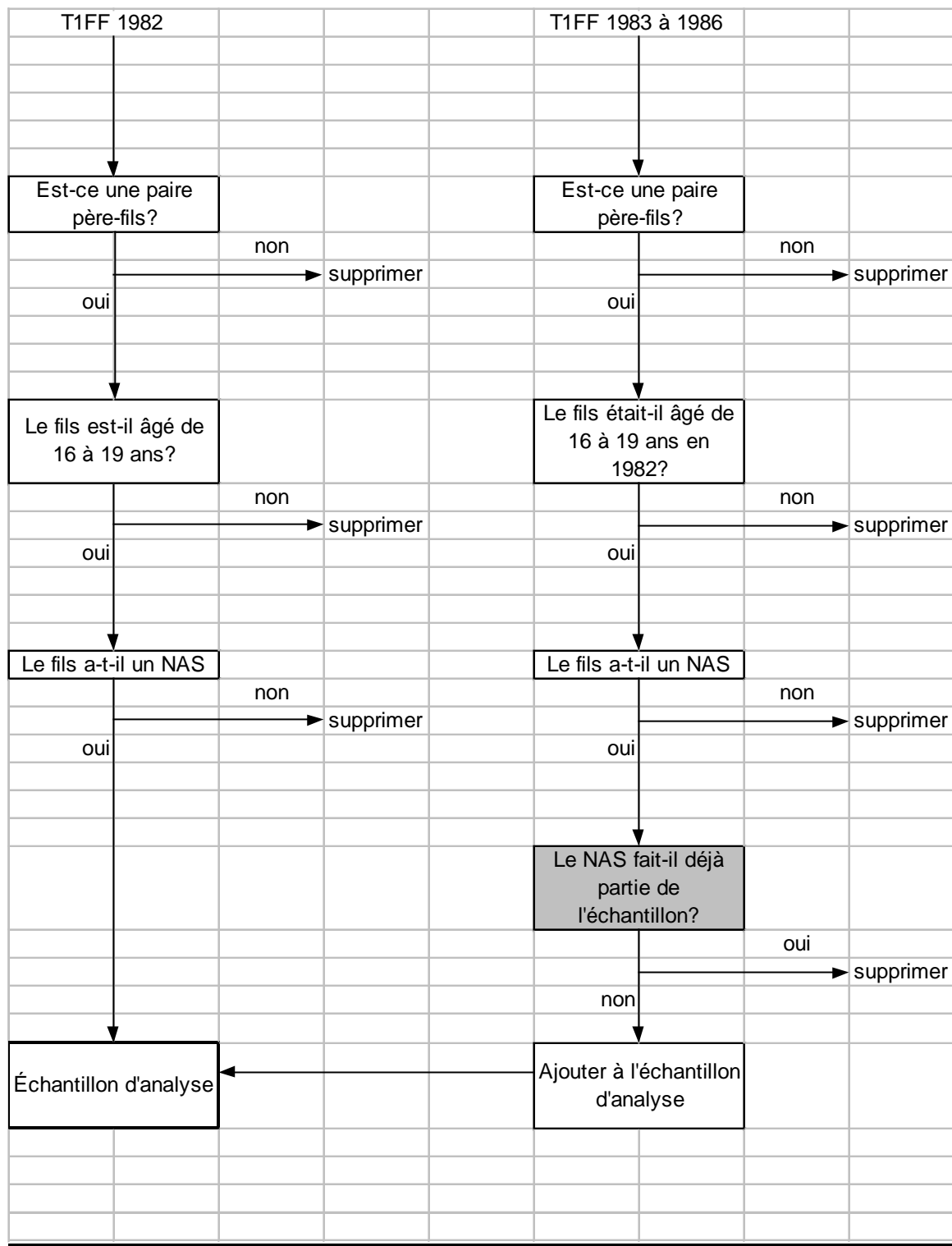
Grâce à ces paires père-fils établies d'après le NAS, nous pouvons obtenir l'information sur le revenu à partir des formulaires T1 produits par le père entre 1978 et 1982 et du formulaire T1 de 1995 pour le fils. Dans le cadre de notre étude, il nous a été possible d'apparier les relevés T1 de 1995 de 497 242 fils aux données T1 d'un père, pour au moins une des années comprises entre 1978 et 1982. La restriction de l'échantillon aux fils aînés, dans les cas où plusieurs fils étaient appariés au même père, a produit un échantillon de 448 196.

Notre analyse est basée sur deux mesures du revenu, les gains et le revenu marchand total. Les gains incluent les gains figurant sur les T4 et ceux déclarés directement par les déclarants. (Les T4 sont produits par l'employeur et doivent être joints au formulaire T1. Les commissions gagnées par les travailleurs salariés sont incluses. En 1995, le revenu minimum pour la production d'un T4 était de 500 \$, en dollars courants. Il était avant cela de 250 \$, toujours en dollars courants.) Les gains incluent également les «autres revenus d'emploi», définis comme étant la somme des allocations de formation des adultes, des subventions de recherche nettes, des pourboires et gratifications, des redevances provenant d'un ouvrage ou d'une invention et des jetons de présence. Ces gains sont déclarés directement sur le T1. (Nous avons comparé les données brutes des T4 produits par les employeurs aux gains déclarés par les particuliers sur le T1 et constaté que la concordance est presque parfaite. La seule exception se remarque tout au bas de la distribution des revenus et est attribuable au revenu

minimal qui détermine s'il y a ou non production d'un T4.) Le «revenu marchand total» inclut le revenu provenant des gains, le revenu net d'un travail indépendant, le revenu de biens et autres revenus. (Les revenus de biens ont quatre composantes : [1] les dividendes versés par des sociétés canadiennes; [2] le revenu net de location; [3] les gains ou les pertes en capital nets; [4] les revenus d'intérêt et autres investissements). Certaines composantes des autres revenus d'emploi, notamment les redevances et les jetons de présence, ne reflètent pas le revenu provenant d'une relation employeur-employé et devraient sans doute être inclus dans le revenu marchand. Cependant, comme il nous est impossible de séparer ces composantes, nous avons jugé préférable de les ajouter aux gains. (Cette distinction n'a toutefois pas d'incidence sur nos résultats. Les estimations par les moindres carrés basées sur une mesure des gains sans les autres revenus d'emploi sont égales à celles basées sur la somme des deux mesures.) Ceux qui le désirent peuvent s'adresser aux auteurs pour obtenir plus de détails sur les définitions utilisées.

Figure A-1

ALGORITHME UTILISÉ POUR LA CRÉATION DES COUPLAGES PÈRE-FILS



Bibliographie

- ALTMAN, N. S. (1992). "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression." *The American Statistician*. Vol. 46, No. 3 (August), pp. 175-185.
- ALTONJI, Joseph G. and Thomas A. Dunn (1991). "Relationships among the Family Incomes and Labor Market Outcomes of Relatives." *Research in Labor Economics*. Vol. 12, pp.269-310.
- ATKINSON, Anthony B. (1981). "On Intergenerational Income Mobility in Britain." *Journal of Post Keynesian Economics*. Vol. 3 (Winter), pp. 194-218.
- _____, A.K. Maynard, and C.G. Trinder (1983). *Parents and Children: Incomes in Two Generations*. London: Heinemann Educational Books.
- BECKER, Gary S. (1991). *A Treatise on the Family*. Enlarged Edition. Cambridge Massachusetts: Harvard University Press.
- _____ and Nigel Tomes (1986). "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics*. Vol. 4, No.3 pt.2, pp. S1-S39.
- _____ (1979). "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy*. Vol. 87, No. 6 pp. 1153-89.
- BEHRMAN, Jere R. And Paul Taubman (1990). "The Intergenerational Correlation between Children's Adult Earnings and their Parents' Income: Results from the Michigan Panel Survey of Income Dynamics." *Review of Income and Wealth*. Series 36, No.2 (June), pp.115-27.
- _____ (1986). "Intergenerational Mobility in Earnings in the U.S.: Some Estimates and a Test of Becker's Intergenerational Endowments Model." *Review of Economics and Statistics*. Vol. 67, No. 1 (February), pp. 144-51.
- BÉLAND, François (1987). "A Comparison of the Mobility Structure of Francophones and Anglophones in Quebec: 1954, 1964, 1974." *Canadian Review of Sociology and Anthropology*. Vol. 24, No. 2 pp. 232-251.
- BÉNABOU, Roland (1994). "Human capital, inequality, and growth: A local perspective." *European Economic Review*. Vol. 38, pp. 817-26.
- BOYD, Monica, John Goyder, Frank E. Jones, Hugh A. McRoberts, Peter C. Pineo, John Porter (1981). "Status Attainment in Canada: Findings of the Canadian Mobility Study." *Canadian Review of Sociology and Anthropology*. Vol. 18, No. 5 pp. 657-73.
- BJORKLUND, A.A. and M. Jäntti (1997a). "Intergenerational Income Mobility in Sweden Compared to the United States." *American Economic Review*. Vol. 87, No. 5 (December), pp.1009-18.
- _____ (1997b). "Intergenerational Mobility of Economic Status: Is the United States Different?" Paper presented to the AEA meetings, New Orleans.

- BROOKS-GUNN, Jeanne, Greg J. Duncan, Pamela Kato Klebanov, Naomi Sealand (1993). "Do Neighborhoods Influence Child and Adolescent Development?" *American Journal of Sociology*. Vol. 99, No. 2 (September), pp. 353-95.
- CARD, David and Richard B. Freeman (1993). "Introduction." In David Card and Richard B. Freeman (editors). *Small Differences that Matter: Labor Markets and Income Maintenance in Canada and the United States*. Chicago: University of Chicago Press.
- CLARK, Collen (1997). "1995 Labour Force Income Profiles Reconciliation With Other Statistics Canada Data." Small Area and Administrative Data Division, Statistics Canada. Unpublished mimeo.
- CLEVELAND, William S. and Susan J. Devlin (1988). "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association*. Vol. 83, No. 403, Applications and Case Studies. (September), pp. 596-610.
- _____, Susan J. Devlin, and Eric Grosse (1988). "Regression by Local Fitting: Methods, Properties, and Computational Algorithms." *Journal of Econometrics*. Vol. 37, Annals No. 1. (January), pp. 87-114.
- _____ (1993). *Visualizing Data*. Summit, New Jersey: Hobart Press.
- _____ and Clive Loader (1995). "Smoothing by Local Regression: Principles and Methods." Unpublished mimeo. AT&T Bell Laboratories.
- COOPER, Suzanne J., Steven N. Durlauf, and Paul A. Johnson (1993). "On the Evolution of Economic Status Across Generations." Paper presented to the Annual Meeting of the American Statistical Association, San Francisco California.
- CORAK, Miles and Andrew Heisz (1998). "How to Get Ahead in Life: Some Correlates of Intergenerational Income Mobility in Canada." In Miles Corak (editor). *Labour Markets, Social Institutions, and the Future of Canada's Children*. Ottawa: Statistics Canada, Catalogue No. 89-553.
- _____ (1995). "The Intergenerational Income Mobility of Canadian Men." *Canadian Business Economics*, Vol. 4, No. 1 (Fall), pp. 59-69.
- CORCORAN, Mary, Roger Gordon, Deborah Laren, Gary Solon (1992). "The Association Between Men's Economic Status and Their Family and Community Origins." *Journal of Human Resources*. Vol. 27, No. 4 (Fall), pp. 575-601.
- COUCH, Kenneth A. and Thomas A. Dunn (1997). "Intergenerational Correlations in Labor Market Status: A Comparison of the United States and Germany." *Journal of Human Resources*. Vol. 22, No.1 (Winter), pp. 210-32.
- COUCH, Kenneth A. and Dean R. Lillard (1997). "Sample Selection Rules and the Intergenerational Correlation of Earnings." *Labour Economics*. Forthcoming.

- CREESE, Gillian, Neil Guppy, and Martin Meissner (1991). *Ups and downs on the ladder of success: Social mobility in Canada*. Statistics Canada, Catalogue No. 11-612E, No. 5.
- DEARDEN, Lorraine, Stephen Machin, and Howard Reed (1997). "Intergenerational Mobility in Britain." *Economic Journal*. Vol. 107, No. 1 (January), pp. 47-66.
- DURLAUF, Steven N. (1996). "A Theory of Persistent Income Inequality." *Journal of Economic Growth*. Vol. 1, No. 1 pp. 75-94.
- _____ (1994). "Spillovers, stratification, and inequality." *European Economic Review*. Vol. 38, pp. 836-845.
- FORTIN, Nicole M. and Sophie Lefebvre (1998). "Intergenerational Income Mobility in Canada." In Miles Corak (editor). *Labour Markets, Social Institutions, and the Future of Canada's Children*. Ottawa: Statistics Canada, Catalogue No. 89-553.
- FOURNIER, Éline, George Butlin, and Philip Giles (1995). "Intergenerational change in the education of Canadians." In *Dynamics of Labour and Income: 1994 Report*. Statistics Canada, Catalogue No. 75-201E.
- GOLDBERGER, Arthur S. (1989). "Economic and Mechanical Models of Intergenerational Transmission." *American Economic Review*. Vol. 79, No. 3 (June), pp. 504-13.
- GUSTAFSSON, Björn (1994). "The Degree and Pattern of Income Immobility in Sweden." *Review of Income and Wealth*. Series 40, No.1 (March), pp. 67-86.
- HÄRDLE, Wolfgang (1990). *Applied nonparametric regression*. Cambridge U.K.: Cambridge University Press.
- _____ (1991). *Smoothing Techniques With Implementation in S*. New York: Springer-Verlag.
- HARRIS, Shelly and Daniela Lucaciu (1994). "An Overview of the T1FF Creation." LAD Reports, Reference Number 94-24-01 v1.2. Small Areas and Administrative Data Division, Statistics Canada.
- HAVEMAN, Robert and Barbara Wolfe (1994). *Succeeding Generations: On the Effects of Investments in Children*. New York: Russell Sage Foundation.
- HILL, Martha S. and Greg J. Duncan (1987). "Parental Family Income and the Socioeconomic Attainment of Children." *Social Science Research*. Vol. 16, pp. 39-73.
- HUNGERFORD, Thomas L. (1993). "U.S. Income Mobility in the Seventies and Eighties." *Review of Income and Wealth*. Series 39, No.4 (December), pp. 403-17.
- ISAJIW, Wsevolod W., Aysan Sev'er, and Leo Driedger (1993). "Ethnic Identity and Social Mobility: A Test of the 'drawback model'." *Canadian Journal of Sociology*. Vol. 18 No.2, pp. 177-96.

- JENKINS, Stephen (1987). "Snapshots versus Movies: 'Lifecycle Biases' and the Estimation of Intergenerational Earnings Inheritance." *European Economic Review*. Vol. 31, pp. 1149-58.
- LILLARD, Dean R. (1998). "Nonlinearities in Intergenerational Mobility in Education and Earnings." Paper presented to the meetings of the American Economics Association, Chicago Illinois.
- LILLARD, Lee A. and Robert T. Reville (1997). "Intergenerational Mobility in Earnings and Occupational Status." Paper presented to the meetings of the American Economics Association, Chicago Illinois.
- McROBERTS, Hugh and Kevin Selbee (1981). "Trends in Occupational Mobility in Canada and the United States: A Comparison." *American Sociological Review*. Vol. 46 (August), pp. 406-21.
- McROBERTS, Hugh (1980). "An Income Attainment Model for Native-Born Canadian Male Wage Earners." In *Reflections on Canadian Incomes*. Ottawa: Economic Council of Canada.
- MICHAUD, Sylvie, David Dolson, Donna Adams, and Martin Renaud (1995). "Combining Administrative and Survey Data to Reduce Respondent Burden in Longitudinal Surveys." Statistics Canada, SLID Research Paper Series No. 95-19.
- MITCHELL, Barbara A., Andrew V. Wister, and Thomas K. Burch (1989). "The Family Environment and Leaving the Parental Home." *Journal of Marriage and the Family*. Vol. 51 (August), 605-13.
- MULLIGAN, Casey B. (1997). *Parental Priorities and Economic Inequality*. Chicago: University of Chicago Press.
- PETERS, Elizabeth H. (1992). "Patterns of Intergenerational Mobility in Income and Earnings." *Review of Economics and Statistics*. Vol. 74, No. 3 pp.456-66.
- SHEA, John (1997). "Does Parents' Money Matter?" NBER Working Paper No. 6026.
- SOLON, Gary (1997). "Intergenerational Mobility in the Labor Market." Forthcoming in the *Handbook of Labor Economics*.
- _____ (1992) "Intergenerational Income Mobility in the United States." *American Economic Review*. Vol. 82, No.3 (June), pp. 393-408.
- _____ (1989). "Biases in the Estimation of Intergenerational Earnings Correlations." *Review of Economics and Statistics* .Vol. 71, No.1 (February), pp. 172-74.
- _____, Marianne E. Page and Greg J. Duncan (1997). "Correlations between Neighboring Children in Their Subsequent Educational Attainment." Unpublished mimeo.
- WANNER, Richard A. and Bernadette C. Hayes (1996). "Intergenerational Occupational Mobility Among Men in Canada and Australia." *Canadian Journal of Sociology*. Vol. 21 No. 1, pp. 43-76.

ZIMMERMAN, David J. (1992). "Regression Toward Mediocrity in Economic Stature." *American Economic Review*. Vol. 82, No. 3 (June), pp. 409-29.

Notes en fin de document

¹ Le coefficient de corrélation entre le logarithme du revenu du père et du fils (ρ) est égal à β_1 si le degré d'inégalité (dans la distribution transversale des revenus) ne change pas d'une génération à l'autre; dans les autres cas, $\beta_1 = \rho[\sigma(Y_i(t)) / \sigma(Y_i(t-1))]$, où les Y_i sont mesurées en logarithmes et σ représente l'écart-type des variables entre parenthèses. Mulligan (1997, p. 164-170) précise que la variation d'inégalité dans le temps varie en fonction de la valeur de β_1 .

² À titre d'exemple, si les niveaux de revenu dans la relation qui précède sont exprimés en logarithmes naturels, alors β_1 représente l'élasticité du revenu d'un enfant par rapport à celui de ses parents. Au Canada, en 1981, le rapport entre le revenu moyen des hommes (travaillant à temps plein, toute l'année) du quintile supérieur et celui des hommes du quintile inférieur était de 3,84. Si l'on utilise ce chiffre, l'avantage au plan du revenu dont bénéficie la personne née d'un père dont le revenu se situe dans le quintile supérieur par rapport à une autre dont le père occupe le quintile inférieur, pour différentes valeurs de β_1 , s'établit comme suit :

β_1	0,1	0,2	0,3	0,4	0,5
Avantage sur le plan du revenu	14 %	31 %	50 %	71 %	96 %

³ Même s'il existe de nombreuses études sociologiques, celles-ci portent sur la «mobilité sociale» et s'intéressent aux changements dans le statut professionnel des parents et des enfants. Citons par exemple McRoberts et Selbee (1981), Béland (1987), Isajiw, Sev'er et Driedger (1993), Creese, Guppy et Meissner (1991), Fournier, Butlin, Giles (1995) et Wanner et Hayes (1996). Boyd *et coll.* (1981) examinent la corrélation entre le niveau de scolarité et le statut professionnel des pères, et de leurs fils et filles stratifiés selon l'origine ancestrale et la langue. McRoberts (1980) établit un lien entre ces caractéristiques de base et le revenu des fils, dans une analyse dont l'esprit se rapproche sans doute le plus de celles décrites dans le texte.

⁴ Haveman et Wolfe (1994) et Peters (1992) ajoutent une série de caractéristiques sur les antécédents familiaux, en s'appuyant sur la théorie de Becker. Hill et Duncan (1987), Corcoran, Gordon, Laren et Solon (1992), Brooks-Gunn, Duncan, Klabanov et Sealand (1993), Solon, Page et Duncan (1997) de même que Corak et Heisz (1998) étudient eux aussi les effets des antécédents familiaux ainsi que les caractéristiques du voisinage. Cooper, Durlauf et Johnson (1993) stratifient leur échantillon par quartier, pour tenter de découvrir les secteurs qui se caractérisent par une pauvreté ou une richesse permanentes. Ils ont constaté une plus grande persistance du revenu chez les familles qui habitent dans les quartiers riches et pauvres que chez les familles à revenu moyen. Précisons toutefois que les conclusions de ces chercheurs sont basées sur une définition assez large de quartier et sur un échantillon assez petit.

⁵ Le «père» n'est pas nécessairement le père biologique, mais devrait plutôt être perçu comme l'homme qui est le chef du ménage. Les beaux-pères, par exemple, sont inclus. Pour être inclus dans notre analyse, le père doit être né entre 1908 et 1952 inclusivement.

⁶ Nous avons également calculé une mesure des gains qui inclut les gains d'un emploi rémunéré et le revenu net d'un travail indépendant. L'analyse de ces données est disponible sur demande.

⁷ Certains pourraient prétendre que notre cohorte d'hommes est encore trop jeune et donc que nous ne mesurons pas leur revenu permanent de façon exacte : entre 29 et 32 ans, certains membres de l'échantillon ne font que commencer à profiter des larges investissements dans le capital humain qui ont été faits alors qu'ils étaient dans la vingtaine et leur revenu continuera d'augmenter. Cependant, l'âge de cette cohorte n'est pas en désaccord avec l'âge utilisé dans les études aux États-Unis; nous avons néanmoins tenté d'évaluer très sommairement cette question en examinant la dynamique du revenu sur une période de dix ans, dans un groupe d'hommes âgés de 29 à 32 ans en 1985, soit le même groupe d'âge que celui que nous avons utilisé pour notre étude, mais dix ans plus tôt. En 1995, environ 24 % de ces personnes occupaient le même décile, dans la distribution des revenus, qu'en 1985 et environ 55 % se trouvaient encore dans le même décile ou avaient augmenté ou régressé d'un décile. À titre de comparaison, 27 % d'un groupe d'hommes de dix ans plus leur aînés (39 à 42 ans) en 1985 se situaient toujours dans le même décile de revenu en 1995, alors que 59 % occupaient toujours le même décile ou avaient régressé ou augmenté d'un décile. Ces données portent à croire que le profil de la dynamique du revenu des personnes vers la fin de la vingtaine ou le début de la trentaine ne diffère pas trop de celui d'une cohorte formée de personnes âgées vers la fin de la trentaine ou le début de la quarantaine, c'est-à-dire une cohorte que l'on peut considérer comme bien engagée dans la réalisation de son revenu viager.

⁸ Lorsqu'on fait la moyenne du revenu, l'âge moyen durant l'intervalle équivalent est utilisé.

⁹ Les coefficients de corrélation associés à ces élasticités, qui ont été calculés suivant la méthode décrite en 1, sont tous inférieurs à 0,2 et varient peu. Les coefficients de corrélation associés à la rangée 1 du tableau sont de 0,154 pour les gains et de 0,132 pour le revenu et ils varient respectivement d'environ 0,12 à 0,17 et de 0,11 à 0,19 pour les autres scénarios.

¹⁰ Selon Mitchell, Wister et Burch (1989), l'âge moyen pour quitter la maison est d'environ 20 ans pour les aînés des parents biologiques, les hommes ayant tendance à quitter la maison environ neuf mois plus tard que les femmes. Bien que cela pourrait laisser croire que la sous-représentation dans nos données n'est pas trop importante, il convient de préciser que leur analyse fait référence à un vaste éventail de personnes qui ont quitté le domicile familial entre les années 50 et le milieu des années 80 et donc qui n'est peut-être pas tout à fait comparable à notre cohorte.

¹¹ Dans le T1FF, ce problème du sous-dénombrement est résolu en imputant les enfants à certaines familles, en fonction d'une règle basée sur l'âge de la mère et son état matrimonial. Aucune donnée imputée n'est utilisée dans notre analyse. Clark (1997) offre une comparaison plus détaillée entre les données administratives basées sur l'EFC et les T1, en faisant ressortir certaines différences conceptuelles et méthodologiques entre les deux sources. Cet auteur précise que la différence entre les chiffres de population et d'emploi est due à l'exclusion des non-déclarants des données fiscales et au fait que les gains moyens de ce groupe ont dû s'établir à environ 1 500 \$, en 1995.

¹² Il est un fait bien connu que les revenus déclarés lors d'un sondage sont arrondis, cela témoignant de la préférence des répondants pour les nombres entiers et leur tendance à déclarer des chiffres qui sont des multiples de dix, cent ou mille. Les erreurs d'arrondissement peuvent être plus importantes à l'extrémité supérieure de la distribution des revenus. Environ 54 % des répondants à l'EFC déclarent leurs gains sous forme de multiples de 100 \$, 46 %, de multiples de 1 000 \$ et 23 %, de multiples de 10 000 \$. Seulement 12,6 % des déclarants plus âgés avec père ont utilisé un multiple de 100 \$ et

11,7 % ont utilisé un multiple de 1 000 \$ ou de 10 000 \$. Les différences entre les deux sources sont encore plus marquées lorsqu'on utilise le revenu marchand : 46 % des répondants à l'EFC ont indiqué un multiple de 100 \$, contre seulement 5,6 % chez les déclarants plus âgés; de même, 37 % et 14 % ont utilisé respectivement des multiples de 1 000 \$ et de 10 000 \$, comparativement à 4,6 % des déclarants.

¹³ Nous avons également réalisé une analyse similaire, en utilisant les gains du fils comme variable dépendante et le revenu marchand total du père comme variable indépendante, ce qui nous donne une estimation des moindres carrés (et erreur-type) de 0,191 (0,009). L'estimation corrigée de Heckman était de 0,194 (0,009).

¹⁴ On a fait la moyenne du revenu des fils sur trois ans, car les fluctuations transitoires peuvent entraîner une surévaluation de la mobilité lorsque celle-ci est représentée dans une matrice de transition, même si elles n'introduisent pas de biais dans les résultats de la régression. En fait, les résultats ne changent pas même lorsqu'on utilise uniquement les revenus de 1995. Les données utilisées sont les résidus de la régression $Y_i = \gamma_0 + \gamma_1 \hat{Age} + \gamma_2 \hat{Age}^2$, où i représente la paire père-fils et Y est le niveau de revenu (exprimé en dollars constants de 1986).

¹⁵ Voir Altman (1992) et Härdle (1990,1991) pour une introduction aux techniques non paramétriques. Notre approche est basée sur les travaux de Cleveland et Devlin (1988), Cleveland, Devlin et Grosse (1988), Cleveland (1993, p. 94-101) et notamment ceux de Cleveland et Loader (1995). Notre analyse utilise le programme LOCFIT écrit par Loader, qui est disponible à <http://cm.bell-labs.com/cm/ms/departments/sia/project/locfit/index.html>.

¹⁶ L'algorithme est en fait plus complexe qu'il n'apparaît ici et le nombre de régressions à exécuter n'est pas aussi élevé que le nombre d'observations. Voir Cleveland, Devlin et Grosse (1988) pour plus de détails à ce sujet.

¹⁷ En termes plus précis, si $u = Y_i(t-1) - Y_o(t-1)$, nous voulons une fonction qui atteint sa valeur maximale à $u=0$, diminue lentement vers 0 à mesure que u augmente en valeur absolue et diffère de zéro dans un intervalle borné. La fonction de pondération tricubique est : $w(u) = \begin{cases} (1-|u|^3)^3 & |u| < 1 \\ 0 & |u| > 1 \end{cases}$

¹⁸ Les références citées en 15, en particulier celles de Cleveland et de ses co-auteurs, montrent que le choix de α par un moyen mécanique, en s'appuyant sur un indicateur comme le AIC, mènera souvent à des choix inexacts. Le choix de α ne doit pas être fait indépendamment des objectifs de l'analyse. De plus, comme Härdle (1990) l'a souligné, le fait de minimiser un tel critère en regard du choix de la courbe lisse ne mène pas nécessairement à un choix optimal pour la première dérivée de cette courbe. C'est pourquoi le AIC ne nous sert qu'à guider le choix de α . Une règle souvent utilisée consiste à adopter une valeur de 0,7.

¹⁹ La sélection de l'échantillon se fait en choisissant tous les fils dont le numéro d'assurance sociale se termine par le chiffre 5, à la condition que leurs gains ou revenus soient supérieurs ou égaux à un et que la moyenne quinquennale de la mesure correspondante pour le père soit elle aussi supérieure ou égale à l'unité.

²⁰ Nous avons examiné de plus près ces valeurs aberrantes et avons constaté que, bien que les gains étaient très faibles, bon nombre de ces personnes avaient des revenus totaux plus élevés (revenu marchand total incluant les paiements de transfert du gouvernement). Un groupe, en marge du marché du travail, touchait des prestations d'assurance-chômage et d'aide sociale. D'autres touchaient des indemnités d'accident du travail ou étaient titulaires de bourses ou de subventions ou d'autres prestations, ce qui témoignait (dans certains cas) d'une participation à des études postsecondaires. Certains pères ont également déclaré des revenus provenant du Régime de pensions du Canada ou du Régime de rentes du Québec. L'utilisation d'un seuil d'exclusion de -8 ne modifie pas les résultats.

²¹ D'autres valeurs de α ont aussi été examinées. Une valeur de 0,7 ne donne pas des résultats très différents, sauf peut-être à l'extrémité supérieure de la queue où l'élasticité atteint 0,5, alors qu'elle n'est que légèrement inférieure à 0,4 lorsque la valeur de α est fixée à 0,84. Une valeur de 0,9 semble causer un lissage excessif des données, ce qui a pour effet d'éliminer la hausse d'élasticité observée lorsque le logarithme des gains corrigés selon l'âge est supérieur à deux.