

No 11-522-XIF au catalogue

**La série des symposiums internationaux  
de Statistique Canada - Recueil**

**Symposium 2005 : Défis  
méthodologiques reliés aux  
besoins futurs d'information**



2005



Statistique  
Canada

Statistics  
Canada

Canada

## Estimation d'un modèle sur la sélectivité de l'attrition, de l'emploi et des salaires à partir de l'Enquête sur la dynamique du travail et du revenu (EDTR)

Brahim Boudarbat<sup>1</sup> et Lee Grenon<sup>2</sup>

### RÉSUMÉ

Afin de comprendre les biais de sélection dans l'estimation d'un modèle lors de l'utilisation de microdonnées tirées d'une enquête longitudinale portant sur un panel, nous nous penchons sur un modèle composé de trois équations liées à la non-attrition / réponse, à l'emploi et aux salaires. Les trois équations sont corrélées librement. Le modèle est estimé à partir des microdonnées provenant de 22 990 individus ayant fourni suffisamment d'information lors de la première vague du second panel de l'EDTR, c'est-à-dire en 1996. Les résultats fournissent des preuves du caractère non aléatoire du comportement d'attrition. Les individus adoptant le comportement d'attrition ou les sortants [*attritors*] et les non-répondants seraient moins liés à leur emploi et proviendraient de populations à faible revenu. Nous trouvons une corrélation positive faible, quoique significative, entre la non-attrition et l'emploi. De plus, les résultats de l'équation liée aux salaires sont généralement surestimés lorsqu'on ne prend pas en considération la sélection découlant de la non-attrition et de l'emploi. Il semble que les salaires observés sont en moyenne supérieurs aux salaires qui le seraient si tous les répondants initialement choisis demeuraient dans l'échantillon.

Mots-clés : estimation de modèle, attrition, emploi, salaires.

### 1. INTRODUCTION

Cette étude examine la sélectivité de l'attrition lors de l'estimation d'un modèle pour un sous-échantillon spécifique de répondants longitudinaux. Nous nous penchons ici sur des répondants longitudinaux qui ont participé à des interviews portant sur le travail et qui étaient âgés de 16 à 64 ans lors de la première année du deuxième panel de l'Enquête sur la dynamique du travail et du revenu (EDTR), c'est-à-dire en 1996. Ce sous-échantillon a fourni un minimum d'information nécessaire à l'estimation du modèle dont nous nous servons. Il était constitué, dans une proportion appréciable, de personnes qui sont passées hors du champ de l'enquête ou qui sont devenues des non-répondants lorsque le panel a pris fin, c'est-à-dire en 2001. Les répondants sortent du champ de l'enquête lorsqu'ils déménagent à l'extérieur des provinces canadiennes, sont institutionnalisés ou décèdent. Les répondants hors du champ de l'enquête ne font pas partie de la population ciblée par l'EDTR et ne peuvent pas participer lors de l'année de référence. L'autre composante de l'attrition est la non-réponse, qui inclut les personnes qui ne peuvent être repérées ou jointes, et celles qui refusent absolument de participer. Les non-répondants font encore potentiellement partie de la population ciblée par l'enquête, mais n'y participent plus.

Lorsque des répondants choisis quittent l'échantillon, l'ensemble de données devient moins représentatif de la population d'où provient l'échantillon longitudinal si l'attrition est non aléatoire. Certaines études économétriques menées aux États-Unis et en Europe ont analysé l'effet de l'attrition au sein de données longitudinales sur l'estimation de modèles (voir par exemple le numéro spécial « Attrition in Longitudinal Surveys » du *Journal of Human Resources*, printemps 1998, vol. 33, n° 2). Ces travaux montrent généralement que le comportement des sortants et des participants sur le marché du travail est différent, même si le fait de ne pas tenir compte du biais de sélection a une influence minimale ou négligeable sur l'estimation. Dans cette étude, nous cherchons à vérifier si le résultat cohérent ci-dessus concernant l'effet de l'attrition sur l'estimation s'applique à l'EDTR. À cette fin, nous examinons un modèle structurel composé de trois équations librement reliées qui expriment la non-attrition / réponse, l'emploi et les salaires. La relation entre ces équations découle du

---

<sup>1</sup> Brahim Boudarbat, École des Relations industrielles, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal (QC) Canada, H3C 3J7. Courriel : brahim.boudarbat@umontreal.ca

<sup>2</sup> Lee Grenon, Statistique Canada, British Columbia Inter-university Research Data Centre, Walter Koerner Library, bureau 202, 1958 Main Mall, Vancouver (C.-B.) Canada V6T 1Z2. Courriel : bcirdc@interchange.ubc.ca

---

Les auteurs tiennent à remercier Statistique Canada de leur avoir permis l'accès aux données de l'EDTR utilisées dans cette étude. Brahim Boudarbat est également reconnaissant de l'aide financière obtenue par l'entremise de la bourse « Mondialisation, révolutions technologiques et éducation » de l'INÉ du CRSHC lorsqu'il était étudiant post-doctoral à UBC.

fait que la situation d'activité est observée seulement chez les répondants à l'interview sur le travail et que les salaires sont observés seulement chez les répondants qui ont un emploi. Ainsi, le modèle permet de tester la sélectivité découlant de l'attrition à la fois dans l'équation sur l'emploi et dans l'équation sur les salaires. Notre modèle est élaboré dans la section 3. Ensuite, nous évaluons le modèle à partir de microdonnées provenant des 22 990 répondants en âge de travailler qui se trouvent dans notre sous-échantillon. Ce dernier est décrit à la section 2 et des paramètres structurels sont présentés à la section 4. La section 5 présente un court résumé accompagné du mot de la fin.

## 2. DONNÉES

Cette analyse concerne seulement les répondants longitudinaux qui ont accepté de fournir de l'information au cours de l'entrevue portant sur le travail afin de s'assurer qu'un minimum d'information soit disponible pour comparer les caractéristiques des répondants qui quittent l'échantillon au cours des années suivantes à celles de répondants qui sont toujours dans le champ de l'enquête et fournissent de l'information à la fin de la période de panel. Plus du cinquième de l'échantillon longitudinal qui était dans le champ de l'enquête et qui avait répondu à la première vague d'interviews sur le travail était hors du champ ou non-répondant aux interviews sur le travail lors de la dernière année de référence du panel<sup>3</sup>. Comme toutes les enquêtes de Statistique Canada, l'EDTR produit des poids d'enquête en utilisant des méthodologies de stratification a posteriori pour ajuster les poids afin de tenir compte de l'attrition et d'assurer la fiabilité de la qualité de données. L'estimation du modèle s'est faite à partir du poids longitudinal d'une personne pour l'année de référence 1996 (Lévesque et Franklin, 2000).

Les caractéristiques des sortants, des non-répondants et des répondants diffèrent et laissent entendre que l'attrition n'est peut-être pas aléatoire au sein de l'échantillon longitudinal de l'EDTR. Les répondants qui deviennent sortants ou non-répondants au cours de l'année suivante sont en moyenne plus jeunes, moins souvent mariés et plus souvent immigrants que les répondants. De plus, les répondants qui vivent dans une région urbaine ou qui déménagent au cours de l'année de référence  $t$  ont plus de chances de devenir des sortants au cours de l'année  $t+1$ . Par ailleurs, les sortants ont plus de chances que les répondants de vivre dans une région urbaine dont la population de leur secteur résidentiel est importante. L'attrition est également plus élevée lorsque les répondants vivent sans époux ou conjoint de fait. Il s'ensuit qu'en moyenne, le salaire et le revenu total du ménage sont plus faibles chez les répondants qui deviendront ensuite des sortants.

## 3. SPÉCIFICATION ÉCONOMÉTRIQUE

L'échantillon utilisé dans l'estimation du modèle élaboré ci-dessous est de 22 990 répondants longitudinaux qui font partie du champ de l'enquête, qui ont fourni suffisamment d'information au cours de l'interview sur le travail et qui avaient entre 16 et 64 ans en 1996. À la fin de 2001, 7 381 de ces répondants (ou 32 %) étaient au moins une fois devenus sortants ou non-répondants. Afin de faciliter l'estimation de notre modèle, nous considérons l'attrition comme un état sans retour, ce qui signifie que les répondants qui quittent le champ de l'enquête ou qui ne répondent pas à l'interview sur la population active au cours d'une vague sont considérés comme sortants lors des vagues subséquentes. Nous ne prenons pas en considération le fait que certains d'entre eux ont été transformés en répondants aux interviews sur le travail au cours des vagues suivantes<sup>4</sup>.

Le modèle se concentre sur la corrélation possible entre trois variables : les salaires, la situation d'activité et le statut de répondant (attrition). Les salaires sont seulement observés chez les répondants ayant un emploi, et la situation d'activité chez les répondants qui fournissent de l'information à propos de leur emploi au cours de l'année de référence. Ainsi, les données sur la situation d'activité des non-répondants à l'interview sur le travail sont censurées (manquantes), de même que les salaires des non-répondants et des répondants qui n'ont pas d'emploi. Si cette censure à deux niveaux n'est pas aléatoire, les résultats basés sur des données observées risquent d'être entachés d'un biais de sélection. Afin d'évaluer le biais de sélection potentiel lors de l'estimation

---

<sup>3</sup> La qualité des données pour l'EDTR peut être évaluée au moyen d'un grand nombre de mesures de non-réponse et d'estimations de la variance (Michaud et Webber, 1995). Cette étude ne mesure pas la non-réponse au sein de l'échantillon; les analyses examinent plutôt la sélectivité de l'attrition au sein d'un modèle structurel pour un sous-échantillon spécifique de répondants.

<sup>4</sup> Le retour au sein de l'échantillon peut également être non aléatoire, ce qui pourrait contrebalancer ou intensifier l'effet de l'attrition.

de la situation d'activité et des salaires, nous proposons le modèle suivant. Les deux sources de sélection sont représentées par les équations à forme réduite (1) et (2) ci-dessous :

### 3.1 Critère de non-attrition et de réponse

$$a_{it}^* = Z_{i(t-1)}\theta + \varepsilon_{1it}, \quad i=1, \dots, n; \quad t=2, \dots, T_i \quad (1)$$

Dans l'équation ci-dessus  $i$  se rapporte aux individus et  $t$ , aux périodes de temps (c'est-à-dire les vagues de l'enquête). L'individu  $i$  est un répondant à la période  $t$  ( $a_{it} = 1$ ) si  $a_{it}^* \geq 0$  et est un sortant au cours de la période  $t$  ( $a_{it} = 0$ ) dans le cas contraire. Étant donné l'hypothèse que l'attrition est un état sans retour, si  $a_{it} = 0$ , alors  $a_{it} = 0$  pour tout  $t' > t$ . Comme l'information concernant la période en cours n'est pas disponible pour les individus qui quittent l'échantillon ou ne répondent pas, nous utilisons des variables retardées. La période initiale de l'analyse de l'attrition correspond à la seconde vague puisque tous les individus ont répondu lors de la première vague.

### 3.2 Critère d'emploi

$$e_{it}^* = X_{it}\alpha + \varepsilon_{2it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T_i \quad (2)$$

Le répondant  $i$  a un emploi ( $e_{it} = 1$ ) si  $e_{it}^* \geq 0$  et n'a pas d'emploi ( $e_{it} = 0$ ) si  $e_{it}^* < 0$ . Une personne n'a pas d'emploi si elle est au chômage ou ne fait pas partie de la population active.  $Z_{it}$  et  $X_{it}$  sont des vecteurs de covariables exogènes, alors que  $\varepsilon_{1it}$  et  $\varepsilon_{2it}$  sont des composantes aléatoires capturant des variables non observées.  $Z_{it}$  et  $X_{it}$  sont observées chaque fois que  $a_{it} = 1$ , et  $Z_{it}$  et  $X_{it}$  sont observées chez tous les individus dans l'échantillon. À l'instar de Zabel (1998), nous avons inclus des variables nominales de vague dans  $Z_{it}$  et  $X_{it}$  afin de tenir compte de la dépendance à la durée. Un changement monotone des coefficients pour les variables nominales de vague indique la présence d'une telle dépendance. Dans l'équation (1), une dépendance négative signifie que la probabilité d'attrition dans l'enquête augmente avec le temps, *ceteris paribus*. Autrement dit, la probabilité qu'une personne soit observée dans l'échantillon diminue avec le temps. Par contre, une durée positive sous-entend que les participants à l'enquête demeurent vraisemblablement dans le champ de l'enquête jusqu'à la fin de la période de référence du panel.

### 3.3 Équation des salaires

Les salaires sont donnés par l'équation suivante :

$$y_{it} = W_{it}\beta + \varepsilon_{3it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T_i \quad (3)$$

où  $y_{it}$  est le logarithme du salaire horaire<sup>5</sup>,  $W_{it}$  est un vecteur de covariables exogènes et  $\varepsilon_{3it}$  est une composante aléatoire. Le modèle structurel est donné par les équations (1), (2) et (3). Ce modèle est séquentiel étant donné que la variable nominale  $e_{it}$  est observée seulement si  $a_{it} = 1$  (l'individu est dans le champ de l'enquête et répond), et  $y_{it}$  est observé seulement si  $a_{it} = 1$  et  $e_{it} = 1$  (l'individu est dans le champ de l'enquête, répond et a un emploi<sup>6</sup>) (voir Maddala, 1983, p. 273-283, pour d'autres exemples sur les critères de sélectivité multiples).

<sup>5</sup> Dans l'estimation empirique, nous considérons le salaire horaire composite pour tous les emplois rémunérés occupés par le répondant durant l'année  $t$ .

<sup>6</sup> Veuillez prendre note que nous estimons l'équation (3) en utilisant des salaires horaires, qui ne sont donnés que pour les travailleurs rémunérés. Cependant, les travailleurs rémunérés incluent les travailleurs non rémunérés. Ces derniers sont donc négligés dans l'équation (3).

Idéalement, il faudrait estimer le modèle en considérant que les termes aléatoires  $\varepsilon_{jit}, j=1,2,3, t=1, \dots, T_i$  sont librement corrélés pour un même individu. Cependant, une telle démarche exigerait le calcul de probabilités conjointes à partir d'une distribution variable  $3 \times T_i$ , ce qui présente des problèmes sur le plan pratique. Afin de faciliter l'estimation du modèle, nous adopterons la démarche du modèle à effets aléatoires (voir ci-dessous). Nous estimons aussi le modèle de manière constante sur deux étapes en suivant la démarche suggérée par Ham (1982). Cette dernière démarche, qui représente un prolongement de la règle de l'estimateur à deux étapes (pour une sélection) proposée par Heckman (1979), est plus attrayante en termes de calcul que la méthode de maximum de vraisemblance et produit des estimations de paramètres constantes. La première étape fait appel à une estimation conjointe des équations de sélection (1) et (2) en utilisant des données du panel<sup>7</sup>. Ensuite, des termes de correction fondés sur les estimations de paramètres obtenus sont calculés et insérés dans l'équation des salaires (3) pour prendre en compte le biais de sélection.

### 3.4 Étape 1 : Équations des sélections

Dans le but de simplifier le calcul des probabilités conjointes, nous adoptons le modèle à effets aléatoires, qui pose :

$$\varepsilon_{1it} = u_{1i} + v_{1it} \quad \varepsilon_{2it} = u_{2i} + v_{2it} \quad \varepsilon_{3it} = u_{3i} + v_{3it} \quad (4)$$

où  $u_{1i}, u_{2i}$  et  $u_{3i}$  désignent les effets spécifiques individuels qui sont censés être librement corrélés, mais indépendants de  $Z_{it}, X_{it}$  et  $W_{it}$ , ainsi que de  $v_{jit}$  pour  $j=1,2,3$  et  $t=1, \dots, T_i$ . Nous supposons également que les termes d'erreur  $v_{jit}$  sont distribués indépendamment parmi les individus et dans le temps. De plus,  $v_{jit}$  sont mutuellement indépendants. Par conséquent, les corrélations entre les équations (1), (2) et (3) sont données par les corrélations entre les effets spécifiques individuels  $u_{1i}, u_{2i}$  et  $u_{3i}$ . Soit  $u_i^* = (u_{1i}, u_{2i}, u_{3i})'$  si  $u_i^*, \varepsilon_{1it}, \varepsilon_{2it}$  et  $\varepsilon_{3it}$  sont indépendants. Nous supposons que  $u_i^*$  suit une distribution normale à trois variables :

$$u_i^* \sim N(0, \Sigma), \quad (5)$$

$$\text{où } \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{pmatrix}.$$

L'attrition et la non-réponse sont aléatoires et il n'y a pas de biais de sélection dans les estimations de l'équation d'emploi (équation 2) si les déterminants individuels non observés de l'emploi ne sont pas corrélés avec les déterminants non observés d'attrition et de non-réponse, (c'est-à-dire si  $\sigma_{12} = 0$ ). De la même manière, il n'y a pas de sélectivité découlant de l'attrition lors de l'estimation de l'équation de salaire si  $u_{1i}$  et  $u_{3i}$  ne sont pas corrélées (c'est-à-dire si  $\sigma_{13} = 0$ ). Telle qu'elle est décrite ci-dessus, la première étape de notre méthode nécessite l'estimation conjointe des équations (1) et (2). Pour cette raison, la contribution individuelle de la fonction de vraisemblance conditionnelle à  $u_i = (u_{1i}, u_{2i})'$  est :

$$L_i(u_i) = \prod_{t=1}^{T_i} L_{it}(u_i), \quad (6)$$

où

$$L_{it}(u_i) = \{Pr(a_{it} = 0 | u_i)\}^{1-a_{it}} \times \left[ \{Pr(a_{it} = 1, e_{it} = 0 | u_i)\}^{(1-e_{it})} \times \{Pr(a_{it} = 1, e_{it} = 1 | u_i)\}^{e_{it}} \right]^{a_{it}} \quad (6.1)$$

pour  $t \geq 2$ ,

<sup>7</sup> Ham (1982) n'utilise que des données transversales.

$$L_{it}(u_i) = L_{i1} = \{Pr(e_{it} = 0 | u_i)\}^{(1-e_{it})} \times \{Pr(e_{it} = 1 | u_i)\}^{e_{it}} \quad (6.2)$$

pour  $t=1$ .

Étant donné que tous les individus sont dans le champ de l'enquête et ont répondu à l'interview sur le travail à la période  $t = 1$ , la contribution d'un individu à la fonction de vraisemblance dépend seulement de sa situation d'activité à cette période. Puisque  $\varepsilon_{1it}$  et  $\varepsilon_{2it}$  sont indépendantes en fonction de  $u_i$ , l'équation (6.1) se simplifie à :

$$L_{it}(u_i) = \{Pr(a_{it} = 0 | u_i)\}^{1-a_{it}} \times [Pr(a_{it} = 1 | u_i) (\Pr(e_{it} = 0 | u_i))^{1-e_{it}} (\Pr(e_{it} = 1 | u_i))^{e_{it}}]^{a_{it}} \quad (6.3)$$

En ce qui concerne les problèmes d'identification, nous supposons que  $v_{1it}$  et  $v_{2it}$  suivent une distribution  $N(0, I)$ . Par la suite, l'apport inconditionnel d'un individu à la fonction de vraisemblance est :

$$L_i = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} L_i(u_i) g(u_i) du_{1i} du_{2i} \quad (7)$$

où  $g(\cdot)$  est la fonction de densité conjointe de  $u_{1i}$  et  $u_{2i}$ . Finalement, les estimations de vraisemblance maximale complètes des paramètres en (1) et (2) avec l'hypothèse d'homoscédasticité sont obtenus en maximisant le logarithme de la fonction de vraisemblance :

$$\log(L) = \sum_{i=1}^n \log(L_i). \quad (8)$$

Étant donné que la fonction en (8) fait appel à une intégration à deux dimensions, l'optimisation directe n'est généralement pas envisageable. Nous utiliserons plutôt un maximum de vraisemblance simulée. Remarquez que la fonction en (7) est une espérance ( $L_i = E_{u_i} [L_i(u_i)]$ ), dont on peut obtenir une approximation par une moyenne simulée :

$$L_{is} = \frac{1}{R} \sum_{r=1}^R L_i(u_{ir}) \quad (9)$$

où  $u_{ir}$ ,  $r=1, \dots, R$ , sont  $R$  tirages de la distribution à double variable de  $u_i$ .  $u_{1i}$  et  $u_{2i}$  peuvent être exprimées comme combinaisons linéaires de deux  $N(0, I)$  indépendantes,  $\eta_{1i}$  et  $\eta_{2i}$  :  $u_{1i} = s_1 \eta_{1i} + s_2 \eta_{2i}$  et

$$u_{2i} = s_3 \eta_{2i} \quad (10)$$

$s_1$ ,  $s_2$  et  $s_3$  sont trois coefficients inconnus qui doivent être estimés. Remarquez que  $u_{1i}$  et  $u_{2i}$  sont indépendantes si  $s_2 = 0$ . Enfin, les paramètres en (1) et (2), qui incluent  $s_1$ ,  $s_2$  et  $s_3$ , sont obtenus en maximisant le logarithme de la vraisemblance simulée<sup>8</sup> :

$$\log(L_s) = \sum_{i=1}^n \log(L_{is}). \quad (11)$$

Un échantillon de  $u_i$  est construit de la manière suivante. Premièrement, nous tirons deux échantillons indépendants de dimension  $R$  d'une  $N(0, I)$ . Ensuite, un échantillon de  $u_i$  est obtenu en utilisant les formules

<sup>8</sup> Voir Gourieroux et Monfort (1996) ainsi que Train (2002) pour une discussion et une démonstration statistique. Voir aussi Green (2002) pour des applications du maximum de vraisemblance simulée.

en (10). Gouriéroux et Monfort (1996) montrent que si  $\sqrt{n}/R \rightarrow 0$  et que  $R$  et  $n \rightarrow \infty$ , l'estimateur du maximum de vraisemblance simulée et l'estimateur de maximum de vraisemblance réelle sont asymptotiquement équivalents. Dans l'application empirique, nous utilisons  $R = 50$ <sup>9</sup>.

### 3.5 Étape 2 : Équation des salaires ajustée en fonction de la sélection

Dans la deuxième étape, nous estimons l'équation des salaires corrigée en fonction de la sélection. L'espérance de  $y_{it}$ , conditionnelle à la réponse et au fait d'occuper un emploi (tout en ne tenant pas compte de la corrélation entre les observations)<sup>10</sup> est :

$$E\left(y_{it} \mid a_{it}^* \geq 0, e_{it}^* \geq 0\right) = W_{it}\beta - \sigma_{13}\lambda_{1t} - \sigma_{23}\lambda_{2t} \quad (12)$$

Les formules de  $\lambda_{1t}$  et  $\lambda_{2t}$  pour  $t \geq 2$  sont données par Ham (1982). Pour la période  $t = 1$ , il n'y a qu'une seule source de sélection, soit la situation d'activité;  $\lambda_{2t}$  est simplement l'inverse du ratio de Mills. Les estimations de paramètres de la première étape sont utilisées pour former des estimations constantes  $\hat{\lambda}_{1t}$  et  $\hat{\lambda}_{2t}$  à partir de  $\lambda_{1t}$  et  $\lambda_{2t}$ . Par la suite, nous estimons  $\beta$ ,  $\sigma_{13}$  et  $\sigma_{23}$  en effectuant la régression par les moindres carrés ordinaires (MCO) groupée à partir de l'échantillon sélectionné (comme le propose Woolridge, 2001, chapitre 17, pour un modèle avec un seul critère de sélection) :

$$y_{it} = Z_{it}\beta - \sigma_{13}\hat{\lambda}_{1it} - \sigma_{23}\hat{\lambda}_{2it} + \varepsilon_{3it}^*, \quad (13)$$

où  $\varepsilon_{3it}^* = \varepsilon_{3it} + \sigma_{13}(\hat{\lambda}_{1it} - \lambda_{1it}) + \sigma_{23}(\hat{\lambda}_{2it} - \lambda_{2it})$ .

Enfin, les estimations constantes des écarts-types des pentes MCO sont obtenues en utilisant les formules de Ham (1982).

## 4. RÉSULTATS EMPIRIQUES

### 4.1 Équations de non-attribution et d'emploi

Au début, nous remarquons que le coefficient estimé  $s_2$  sur  $\eta_{2i}$  dans l'équation (10) est statistiquement significatif à un niveau de 5 %, ce qui signifie que les déterminants non observés individuels de non-attribution et d'emploi,  $u_{1i}$  et  $u_{2i}$ , sont corrélés. La corrélation entre ces deux termes est estimée à 0,45, mais la corrélation estimée entre l'ensemble des termes aléatoires,  $\varepsilon_{1it}$  et  $\varepsilon_{2it}$ , est seulement de 0,03, bien que statistiquement significative à un niveau de 5 %. Par conséquent, l'estimation séparée des équations de non-attribution et d'emploi n'introduira que des biais très limités<sup>11</sup>. Les résultats suggèrent également que l'effet spécifique individuel est plus important sur l'emploi que sur le comportement de non-attribution.

Nous en venons également à la conclusion que les variables qui augmentent la participation au marché du travail et (ou) qui réduisent la mobilité (par exemple, l'éducation, le fait d'être marié, l'état de santé) augmentent aussi la probabilité que l'individu demeure dans le champ de l'enquête et qu'il réponde. Dans le même esprit, le fait

<sup>9</sup> Nous avons initialement estimé le modèle en utilisant  $R = 30$ . L'augmentation du nombre de tirages de 30 à 50 n'entraîne que peu de changements dans les résultats. Néanmoins, étant donné la complexité de la fonction de vraisemblance et la taille importante de l'échantillon, l'estimation du modèle est exigeante sur le plan du calcul, même avec un petit nombre de tirages.

<sup>10</sup> Nous n'avons pas trouvé d'étude dans les travaux qui faisait la correction pour deux sources de sélection pour l'utilisation des données de panel. Woolridge (2001, chapitre 17) présente un cas avec un critère de sélection.

<sup>11</sup> Nous avons estimé les équations de non-attribution et d'emploi séparément et nous avons obtenu des estimations très semblables à celles obtenues lorsque les équations sont estimées conjointement.

d'être immigrant, et particulièrement un immigrant d'un groupe minoritaire visible, réduit significativement à la fois la probabilité d'avoir un emploi et la probabilité de demeurer un répondant dans l'échantillon. Par ailleurs, une personne qui a déménagé au cours d'une année (signe de mobilité géographique) a plus de risques d'être non-répondant ou hors du champ de l'enquête lors de l'année suivante. Les résultats indiquent aussi que les femmes sont moins susceptibles d'être sortantes, mais qu'elles ont moins souvent un emploi que les hommes. Dans un autre ordre d'idées, le fait qu'un individu soit étudiant n'affecte pas la participation à l'enquête, bien que son statut d'étudiant réduise la probabilité qu'il ait un emploi. Il est cependant intéressant de constater que l'âge, le statut de propriétaire et la taille de la famille n'ont aucun effet significatif sur l'attrition. Voici un aspect important que nos résultats mettent en évidence : la hausse du revenu familial réduit la probabilité d'attrition. Ce résultat concorde avec MaCurdy, Mroz et Gritz (1998), qui ont découvert que les personnes qui quittent le *National Longitudinal Survey of Youth* (NLSY) sont, dans une mesure disproportionnée, issues de la population à faible revenu.

Comme elles ne changent pas de manière monotone, les estimations de coefficients pour les variables nominales de vague ne font état d'aucune dépendance à la durée, qu'elle soit positive ou négative. La probabilité d'attrition est la plus élevée en 2000 et la plus faible en 1998. Ce résultat concorde avec les statistiques descriptives.

## 4.2 Les équations de salaire

En interprétant les résultats, nous nous concentrons sur l'analyse des biais survenus parce qu'on ne tient pas compte des sélections de non-attrition et d'emploi lors de l'estimation de l'équation de salaire, plutôt que sur l'analyse des effets des covariables sur le niveau de salaire. Il est utile de noter que, même lorsque la corrélation entre les équations de non-attrition et d'emploi est fixée à zéro, il n'y a pratiquement aucun effet sur les estimations corrigées des équations de salaire; réalité qui confirme la faible interdépendance entre les deux statuts. Le résultat le plus original que nous avons constaté est que les estimations de coefficients sur les termes de correction sont négatives et hautement significatives, permettant d'entrevoir le caractère non aléatoire de la non-attrition et de l'emploi. À partir de l'équation (13), nous pouvons interpréter le signe négatif de ces estimations comme un indice de la corrélation positive du salaire avec la non-attrition et l'emploi. L'ampleur du biais de sélection découlant de l'emploi est plus importante que celle découlant de l'attrition. L'écart de salaires entre l'échantillon sélectionné (disponible) et un échantillon tiré au hasard qui aurait des caractéristiques observées identiques est estimé à 9,65 % à cause de la sélection basée sur la non-attrition, contre 13,51 % à cause de la sélection basée sur l'emploi<sup>12</sup>.

Si nous ne tenons pas compte des deux sources de sélection, les effets de plusieurs covariables sur les salaires sont surestimés. Par exemple, l'ajustement au titre de la sélection génère un écart homme-femme de 17,4 %, contre 23,1 % en l'absence d'un tel ajustement. De manière similaire, l'écart entre un non-immigrant et un immigrant qui n'est pas membre d'un groupe minoritaire visible diminue de 4,6 % à 2,2 %, et de 20,6 % à 12,1 % entre un non-immigrant et un immigrant qui fait partie d'un groupe minoritaire visible, après l'ajustement au titre de la sélection. De plus, les retours aux études sont surestimés, puisqu'ils diminuent de manière significative après la correction au titre de la sélection.

## 5. LE MOT DE LA FIN

Les estimations structurelles fournissent des preuves du caractère non aléatoire du comportement d'attrition. Le coefficient de corrélation entre les composantes aléatoires des équations de non-attrition et d'emploi est positif et significatif, quoique faible. Les estimations de l'équation de salaire indiquent des biais de sélection significatifs découlant à la fois de la non-attrition et de l'emploi. La plupart des coefficients sur les covariables sont surestimés lorsque la sélection n'est pas prise en considération. Nous concluons également que nos estimations de modèle surestiment les salaires lorsqu'ils ne sont pas ajustés, puisque les salaires dans l'échantillon longitudinal disponible sont probablement plus élevés que les salaires qui seraient observés si tous les répondants sélectionnés à l'origine demeuraient dans l'échantillon jusqu'à la fin du panel. De manière similaire, nous trouvons qu'une hausse du revenu familial réduit la probabilité d'attrition, ce qui pourrait entraîner une plus grande surestimation du revenu familial.

---

<sup>12</sup> L'écart est calculé en multipliant le négatif du coefficient de sélection par la valeur moyenne du terme de correction.



## RÉFÉRENCES

- Gourieroux, C. et Monfort, A. (1996), "Simulation Based Econometric Methods," Oxford University Press, New York, 1996.
- Green, W. H. (2002), "Convenient Estimators for the Panel Probit Model: Further Results," *Department of Economics, Stern School of Business, New York University*, Revised, October, 2002
- Ham, J.C. (1982), "Estimation of a Labour Supply Model with Censoring Due to Unemployment and Underemployment," *The Review of Economic Studies* Vol. 49, No. 3 (Jul., 1982), pp. 335-354.
- Heckman, J. (1979). "Sample selection bias as a specification error," *Econometrica* 47 : 153-161.
- Levesque, I. et Franklin, S. (2000). "Pondération longitudinale et transversale de l'Enquête sur la dynamique du travail et du revenu". No. au catalogue 75F0002MIF2000004. Ottawa: Statistique Canada.
- MaCurdy, T., Mroz, T. et Gritz, R.M. (1998), "An Evaluation of the National Longitudinal Survey of Youth," *Journal of Human Resources*, Vol. 33(2), pp. 345-436.
- Maddala, G.S. (1983), "Limited Dependent and Qualitative variables in Econometrics," Cambridge University Press, 1983.
- Michaud, S. et Webber M. (1994). "La mesure de la non-réponse dans une enquête longitudinale : l'expérience de l'Enquête sur la dynamique du travail et du revenu". No. au catalogue 75F0002MIF1994016. Ottawa: Statistique Canada.
- Wooldridge, J. M. (2001), "Econometric Analysis of Cross Section and Panel Data," M. Cambridge, Mass: The MIT Press, 2001.
- Zabel, J. E. (1998), "[An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior](#)" *The Journal of Human Resources*, Vol. 33, No. 2 (printemps 1998), pp. 479-506.

## TABLEAUX ET BIBLIOGRAPHIE

On peut obtenir la version intégrale du manuscrit en faisant la demande auprès des auteurs.