

La qualité des données dans l'Enquête sur la dynamique du travail et du revenu (EDTR) de 2007

Jean-François Bastien

Division des méthodes d'enquêtes auprès des ménages
Édifice R.H. Coats, Ottawa, K1A 0T6

Table des matières

| | |
|--|----|
| 1. Introduction..... | 5 |
| 2. Composition/érosion de l'échantillon..... | 7 |
| 3. Erreurs d'échantillonnage..... | 9 |
| 4. Erreurs de couverture..... | 10 |
| 5. Taux de réponse..... | 13 |
| 6. Taux de permission d'accès au dossier fiscal..... | 17 |
| 7. Taux de couplage avec le dossier fiscal..... | 19 |
| 8. Taux d'imputation..... | 21 |
| 9. Arrondissement des données sur le revenu..... | 26 |

1. Introduction

L'Enquête sur la dynamique du travail et du revenu (EDTR) est une enquête longitudinale qui date de 1993. Elle se veut un moyen de mesurer les variations du bien-être économique des Canadiens ainsi que les facteurs touchant ces changements. La population cible est formée de toutes les personnes vivant au Canada, sauf les personnes suivantes : les résidents du Yukon, des Territoires du Nord-Ouest et du Nunavut, les résidents des réserves, les personnes institutionnalisées et le personnel militaire en casernement.

L'échantillon de l'EDTR est formé de 2 panels. Chaque panel demeure dans l'enquête pour 6 années consécutives et un nouveau panel y entre tous les 3 ans. En janvier suivant l'année de référence, les préposés aux entrevues téléphoniques prennent contact avec les ménages de l'échantillon de l'EDTR. Ils recueillent des données démographiques pour chaque membre du ménage et des données d'enquête complètes pour chaque membre de 16 ans et plus.

On a apporté un changement marquant à la collecte des données de l'EDTR au cours de l'année de référence 2004. Auparavant, on disait aux répondants à la fin de l'entrevue de janvier qu'on communiquerait à nouveau avec eux en mai pour obtenir des données sur leur revenu ainsi que sur certains types de dépenses. Depuis 2004, on pose toutes les questions au cours de la même entrevue, en janvier. Le répondant peut dès lors accorder à Statistique Canada la permission d'extraire toutes les données requises du fichier des données fiscales T1, ce qui élimine la nécessité de la deuxième partie de l'entrevue. On reportait la collecte des données sur le revenu à mai dans l'espoir que le répondant (venant de produire une déclaration de revenus) connaîtrait mieux les données requises.

Bien que conçue au départ comme une enquête longitudinale, l'EDTR a toujours conservé la capacité de produire des estimations transversales. Cet aspect transversal a pris une importance nouvelle avec l'annulation de l'Enquête sur les finances des consommateurs après l'année de référence 1997. L'EDTR est alors devenue la source première de données transversales sur le revenu du ménage et de la famille.

Toutes les personnes membres des ménages choisis pour l'EDTR au début de la première année de l'existence d'un panel font partie de l'échantillon longitudinal pour l'EDTR. À ce titre, ce sont ces particuliers qui font l'objet d'une observation longitudinale. Toute autre personne vivant dans un ménage avec une personne de l'échantillon longitudinal est appelée cohabitant. Les cohabitants vivant avec des personnes de l'échantillon longitudinal admissibles à l'échantillon transversal font aussi partie de l'échantillon transversal.

Pour plus de renseignements sur les concepts, les définitions et le plan de l'enquête, voir la publication de Statistique Canada intitulée « *Enquête sur la dynamique du travail et du revenu – Un aperçu de l'enquête* », ”,

<http://www.statcan.gc.ca/pub/75f0011x/75f0011x2008001-fra.htm>

Les enquêtes par sondage peuvent comporter des erreurs. Comme dans toutes ses enquêtes, Statistique Canada met un temps et un effort considérables à contrôler ces erreurs à chaque stade de l'EDTR. Mais il y a quand même des erreurs. Statistique Canada a pour politique de fournir des mesures de la qualité des données pour aider ses utilisateurs à bien interpréter les données. Le présent rapport résume les mesures de qualité qui ont pour objet de décrire la qualité des données de l'EDTR.

Le tableau ci-dessous illustre les principaux faits saillants des indicateurs de qualité pour l'année de référence 2007 pour le Canada.

Tableau 1.1 Principaux indicateurs de qualité de l'EDTR pour le Canada en 2007

| Indicateur | Statistique |
|--|--------------------------------------|
| Taille d'échantillon longitudinal (individu) <ul style="list-style-type: none"> • Panel 4 • Panel 5 | 29 088 32 703 |
| Taille d'échantillon transversal (individus longitudinaux) admissible et cohabitants <ul style="list-style-type: none"> • Panel 4 • Panel 5 | 32 519 35 789 |
| Coefficient de variation <ul style="list-style-type: none"> • Revenu total médian | 0,6 % |
| Taux de glissement au niveau des personnes <ul style="list-style-type: none"> • Panel 4 • Panel 5 | 17,0 % 12,3 % |
| Taux de glissement au niveau des ménages <ul style="list-style-type: none"> • Panel 4 • Panel 5 | 16,3 % 12,1 % |
| Taux de réponse <ul style="list-style-type: none"> • Transversal, personne • Transversal, ménage • Longitudinal, personne <ul style="list-style-type: none"> ○ Panel 4 ○ Panel 5 | 69,6 % 71,8 % 68,9 % 77,3 % |
| Taux de permission <ul style="list-style-type: none"> • Panel 4 • Panel 5 | 91,3 % 85,5 % |
| Taux de couplage avec le dossier fiscal (NAS trouvé) | 94,9 % |
| Taux d'imputation pour les personnes <ul style="list-style-type: none"> • Imputation totale • Imputation partielle | 2,5 % 22,4 % |
| Taux d'imputation pour les ménages <ul style="list-style-type: none"> • Imputation partielle | 38,9 % |

2. Composition/érosion de l'échantillon

Bien que conçue au départ comme enquête longitudinale, l'EDTR a aussi la capacité de produire des estimations transversales. Chaque personne ne faisant pas partie de l'échantillon longitudinal qui vit avec un répondant de l'échantillon longitudinal est aussi appelée à participer à l'enquête. Cette personne est dite cohabitante. Les tableaux 2.1 et 2.2 montrent la composition de l'échantillon de l'EDTR par province et par région métropolitaine de recensement (RMR), respectivement, en nombre de répondants de l'échantillon longitudinal, de répondants de l'échantillon longitudinal qui ne sont pas admissibles à l'échantillon transversal (comme les personnes décédées, les personnes institutionnalisées et celles qui ont déménagé à l'extérieur du pays), et les cohabitants répondants.

La couverture de l'échantillon transversal de l'EDTR est maintenue par l'ajout de cohabitants chaque année. La seule exception est le cas des immigrants qui arrivent après le début d'un panel et avant le début du suivant, et qui forment de nouveaux ménages, ce qui introduit une légère sous-représentation. L'échantillon longitudinal, toutefois, est sujet à érosion. L'érosion est la perte graduelle de répondants chaque année pendant la vie du panel. Le tableau 2.3 montre le statut de répondant pour les personnes choisies au départ dans l'échantillon longitudinal. Le tableau montre aussi que l'échantillon longitudinal des répondants est formé des répondants faisant partie du champ de l'enquête, des personnes qui ont déménagé au Yukon, dans les Territoires du Nord-Ouest ou au Nunavut, des personnes qui ont déménagé à l'extérieur du Canada, des personnes institutionnalisées et des personnes décédées.

Tableau 2.1 Composition de l'échantillon de l'EDTR par province en 2007

| Province | Taille de l'échantillon longitudinal | | Échantillon longitudinal des personnes non admissibles à l'échantillon transversal ¹ | | Cohabitants | | Taille de l'échantillon transversal | |
|-----------------------|--------------------------------------|---------------|---|--------------|-------------|-------------|-------------------------------------|---------------|
| | Panel 4 | Panel 5 | Panel 4 | Panel 5 | Panel 4 | Panel 5 | Panel 4 | Panel 5 |
| Terre-Neuve | 1 218 | 1 462 | 96 | 44 | 182 | 165 | 1 304 | 1 583 |
| Île-du-Prince-Édouard | 816 | 899 | 55 | 24 | 158 | 116 | 919 | 991 |
| Nouvelle-Écosse | 1 971 | 1 963 | 146 | 66 | 376 | 276 | 2 201 | 2 173 |
| Nouveau Brunswick | 1 672 | 1 874 | 126 | 59 | 355 | 227 | 1 901 | 2 042 |
| Québec | 5 557 | 6 002 | 392 | 181 | 1 209 | 916 | 6 374 | 6 737 |
| Ontario | 8 331 | 9 232 | 567 | 362 | 1 561 | 1 214 | 9 325 | 10 084 |
| Manitoba | 1 983 | 2 246 | 161 | 94 | 401 | 289 | 2 223 | 2 441 |
| Saskatchewan | 2 033 | 2 405 | 168 | 105 | 402 | 329 | 2 267 | 2 629 |
| Alberta | 2 510 | 3 315 | 143 | 102 | 641 | 575 | 3 008 | 3 788 |
| Colombie-Britannique | 2 663 | 3 008 | 192 | 109 | 526 | 422 | 2 997 | 3 321 |
| Autre | 334 | 297 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 29 088 | 32 703 | 2 046 | 1 146 | 5811 | 4529 | 32 519 | 35 789 |

0 Zéro absolu ou valeur arrondie à zéro

1. Comprend les personnes qui sont décédées, les personnes institutionnalisées et celles qui ont déménagé à l'extérieur du pays.

Dans un premier temps, pour ce qui est de la taille de l'échantillon longitudinal, on peut observer une différence d'environ 3 600 personnes entre le 4^e panel et le 5^e panel pour le Canada. Une hausse a été observée dans toutes les provinces sauf en Nouvelle-Écosse où la taille a légèrement diminué au 5^e panel. On observe un scénario similaire pour la taille de l'échantillon transversal.

En ce qui concerne l'échantillon longitudinal des personnes non admissibles à l'échantillon transversal, on remarque une baisse de près de 50 % entre le panel 4 et le panel 5, et ce, pour toutes les provinces. Enfin, du côté du nombre de cohabitants au Canada, il est passé de 5 811 à 4 539 entre les deux panels, soit une différence d'environ 22 %.

Tableau 2.2 Composition de l'échantillon de l'EDTR par RMR en 2007

| Région métropolitaine de recensement | Taille de l'échantillon longitudinal | | Cohabitants | | Taille de l'échantillon transversal | |
|--------------------------------------|--------------------------------------|---------------|--------------|--------------|-------------------------------------|---------------|
| | Panel 4 | Panel 5 | Panel 4 | Panel 5 | Panel 4 | Panel 5 |
| Halifax | 427 | 586 | 90 | 91 | 517 | 677 |
| Québec | 401 | 447 | 141 | 95 | 542 | 542 |
| Montréal | 1 100 | 1 197 | 267 | 224 | 1 367 | 1 421 |
| Ottawa - Gatineau | 753 | 826 | 168 | 123 | 921 | 949 |
| Toronto | 1 322 | 1 650 | 313 | 241 | 1 635 | 1 891 |
| Hamilton | 365 | 432 | 70 | 39 | 435 | 471 |
| St. Catharines - Niagara | 389 | 364 | 68 | 56 | 457 | 420 |
| Kitchener | 397 | 427 | 71 | 65 | 468 | 492 |
| London | 360 | 485 | 85 | 94 | 445 | 579 |
| Windsor | 250 | 328 | 48 | 30 | 298 | 358 |
| Winnipeg | 857 | 1 093 | 207 | 151 | 1 064 | 1 244 |
| Calgary | 552 | 699 | 177 | 132 | 729 | 831 |
| Edmonton | 555 | 978 | 157 | 174 | 712 | 1 152 |
| Vancouver | 867 | 1 040 | 179 | 153 | 1 046 | 1 193 |
| Victoria | 232 | 281 | 39 | 49 | 271 | 330 |
| Autre RMR ou AR | 9 751 | 10 950 | 2 049 | 1 664 | 11 800 | 12 614 |
| N'habite pas dans une RMR | 8 130 | 9 477 | 1 682 | 1 148 | 9 812 | 10 625 |
| Pas disponible ¹ | 2 380 | 1 443 | 0 | 0 | 0 | 0 |
| Total | 29 088 | 32 703 | 5 811 | 4 529 | 32 519 | 35 789 |

0 Zéro absolu ou valeur arrondie à zéro

1. Ces renseignements ne sont disponibles que pour les personnes qui sont admissibles à l'échantillon transversal.

En ce qui a trait aux différences dans les tailles de l'échantillon longitudinal selon la région métropolitaine de recensement (RMR), à l'instar de l'augmentation canadienne, telle qu'illustrée précédemment, celle-ci a augmenté de façon similaire pour toutes les RMR, sauf pour St. Catharines – Niagara, où la taille a légèrement diminué. En ce qui a trait à la taille de l'échantillon transversal, les différences entre les 4^e et 5^e panel sont semblables à celles observées pour la taille d'échantillon longitudinal, avec aussi une

diminution pour St. Catharines – Niagara. Pour toutes les autres RMR, la taille a augmenté sauf pour la RMR de Québec où la taille est demeurée stable.

Pour ce qui est du nombre de cohabitants, il a baissé pour toutes les RMR dans le 5^e panel à l'exception de Halifax, où il est demeuré stable ainsi que London, Edmonton et Victoria, où il a augmenté.

Tableau 2.3 Statut des personnes de l'échantillon longitudinal en 2007

| Statut des personnes de l'échantillon longitudinal | Panel 4 | Panel 5 |
|---|----------------|----------------|
| Dans le champ de l'enquête (répondants) | 26 708 | 31 260 |
| Dans le champ de l'enquête (non-répondants) | 3 248 | 6 839 |
| Déménagé au Yukon, dans les T. N. O., au Nunavut | 9 | 7 |
| Déménagé à l'extérieur du Canada | 323 | 288 |
| Institutionnalisées | 657 | 398 |
| Décédées | 1 391 | 750 |
| Éliminées de l'échantillon ¹ | 9 859 | 2 773 |
| Doubles/erreurs ² | 37 | 15 |
| Total | 42 232 | 42 330 |

1. Les répondants peuvent être retirés de l'échantillon pour deux raisons. Les ménages dont tous les membres ont refusé de répondre pour 2 cycles consécutifs sont classés " refus catégoriques ", et il ne se fait pas d'autres efforts pour les dénombrier. De même, après deux ans, on ne tente plus de nouveaux essais de collecte des données auprès des ménages dont on a perdu la trace.
2. Répondants qui ont été inclus par erreur dans le ménage au début de la première année de l'existence d'un panel.

Si le nombre total de personnes dans les 4^e et 5^e panel était très semblable, on remarque des différences importantes en ce qui concerne le statut des personnes de l'échantillon longitudinal. Tout d'abord, le nombre de répondants et de non-répondants dans le champ de l'enquête est plus élevé dans le 5^e que dans le 4^e panel. Ces résultats étaient prévisibles du fait que le 5^e panel n'en était qu'à sa troisième vague comparativement au 4^e panel qui en était à sa 6^e et dernière vague. Pour les mêmes raisons, le nombre de personnes dans la catégorie « personnes éliminées de l'échantillon », est beaucoup plus élevée au 4^e panel, car après six vagues, il y a plusieurs cas pour lesquels ont a perdu la trace ou qui ont été abandonnés ou qui ont été considérés comme refus catégoriques.

3. Erreurs d'échantillonnage

Les erreurs d'échantillonnage viennent du fait que les inférences au sujet de la population observée sont fondées sur les données d'un échantillon de cette population plutôt que de l'ensemble de la population. Le plan d'échantillonnage, la variabilité des caractéristiques mesurées et la taille de l'échantillon sont autant de facteurs qui contribuent à l'ampleur de l'erreur d'échantillonnage.

L'erreur type est une mesure fréquente de l'erreur d'échantillonnage. L'erreur type mesure le degré de la variation introduite dans les estimations par le choix d'un

échantillon particulier plutôt que d'un autre de la même taille et du même plan. Une autre mesure fréquente de l'erreur d'échantillonnage est le coefficient de variation (CV), qui est l'erreur type estimée exprimée en pourcentage de l'estimation.

Dans l'EDTR, on utilise l'approche « bootstrap » pour le calcul des erreurs types. C'est une méthode de rééchantillonnage pour l'estimation de la variance, souvent appliquée aux estimations d'un plan d'échantillonnage complexe. Le tableau 3.1 fait voir les niveaux de CV provinciaux et nationaux pour un échantillon d'estimations de variables clés de l'EDTR.

Tableau 3.1 Coefficients de variation nationaux et provinciaux pour certaines variables de l'EDTR en 2007 (%)

| Variable (au niveau des familles sauf indication contraire) | T.-N. | Î.-P.-É. | N.-É. | N.-B. | Qc | Ont. | Man. | Sask. | Alb. | C.-B. | Canada |
|---|-------|----------|-------|-------|-----|------|------|-------|------|-------|--------|
| Revenu total médian | 3,0 | 2,2 | 2,0 | 2,2 | 1,3 | 1,2 | 1,7 | 1,4 | 1,5 | 1,7 | 0,6 |
| Revenu médian du marché | 3,5 | 3,3 | 2,4 | 2,7 | 1,6 | 1,2 | 2,5 | 3,1 | 1,9 | 2,3 | 0,8 |
| Médiane des salaires et traitements | 3,1 | 4,5 | 2,6 | 3,1 | 1,6 | 1,0 | 2,5 | 2,1 | 1,4 | 2,8 | 0,7 |
| Médiane des prestations d'assurance-emploi | 6,3 | 4,9 | 7,1 | 6,0 | 3,9 | 6,4 | 9,5 | 8,3 | 15,1 | 12,4 | 2,9 |
| Médiane de l'aide sociale | 15,0 | 14,2 | 5,8 | 7,3 | 6,7 | 5,0 | 19,1 | 12,8 | 5,8 | 12,6 | 4,0 |
| Médiane de la variable « Autre revenu » | 12,5 | 22,7 | 18,4 | 17,6 | 9,3 | 6,6 | 15,6 | 15,7 | 12,1 | 10,4 | 4,4 |
| Familles sous le SFR (après impôt) | 9,3 | 14,8 | 7,7 | 7,6 | 3,9 | 3,9 | 6,6 | 6,7 | 7,6 | 5,9 | 2,1 |
| Nombre de personnes employées | 1,8 | 2,0 | 1,5 | 1,4 | 1,1 | 1,0 | 1,9 | 1,6 | 1,5 | 1,7 | 0,5 |

Pour ce qui est du revenu total médian et du revenu du marché médian, on remarque que c'est Terre-Neuve qui enregistre le CV le plus élevé. Par contre, c'est à l'Île-du-Prince-Édouard où le CV est à son maximum pour la médiane des salaires et traitements, la médiane de la variable « autre revenu », pour les familles sous le SFR (après impôt) ainsi que pour le nombre de personnes employées. Le fait qu'on retrouve les tailles d'échantillon les plus petites dans ces deux provinces explique en partie ces CV plus élevés. Toutefois, pour la médiane des prestations d'assurance-emploi et la médiane pour l'aide sociale, les CV les plus élevés ont été observés pour l'Alberta et le Manitoba respectivement.

4. Erreurs de couverture

Pour qu'une enquête produise de bonnes estimations, il faut que son échantillon soit représentatif de la population observée. Pour ce faire, on utilise les projections démographiques du recensement comme totaux de contrôle pour corriger les poids de l'EDTR. Le taux de glissement est une mesure de la différence en pourcentage entre ces

projections du recensement et l'estimation de l'enquête par utilisation des poids avant l'application de ce rajustement pour le glissement. Plus spécifiquement, le glissement se calcule par la formule suivante :

$$glissement_c = \frac{\left(CP_c - \sum_{k \in S_c} w_{kc} \right)}{CP_c} * 100$$

où la classe C est la classe ou le groupe pour lequel nous voulons calculer les taux de glissement. Par exemple, au niveau détaillé, les groupes sont fondés sur la province, le sexe et le groupe d'âge.

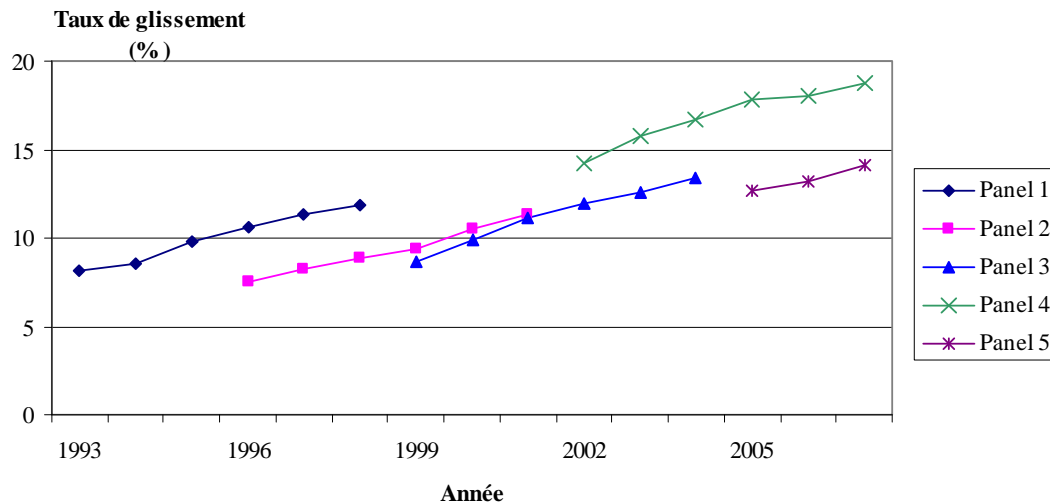
CP_C est la projection démographique du recensement pour la classe C

w_{kc} est le poids de l'enquête pour la k^e unité répondante de la classe C

S_C est l'ensemble de ménages-échantillon dans la classe C

Les taux de glissement pour les enquêtes-ménages sont généralement positifs en raison de la sous couverture de la base de sondage. Le graphique 4.1 montre l'historique des taux de glissement depuis 1993. Le tableau 4.1 montre les taux de glissement au niveau des personnes par province et par panel. Les taux de glissement ont aussi été calculés au niveau des ménages et sont résumés au tableau 4.2 ainsi qu'au graphique 4.2.

Graphique 4.1 Taux de glissement au niveau des personnes selon le panel et l'année de référence (%)



En regardant le graphique 4.1, on remarque que les courbes du taux de glissement sont semblables pour tous les panels alors que le taux a toujours augmenté entre la première et la dernière vague. Les taux de glissement plus élevés au niveau personne et ménage (voir le graphique 4.2) pour le Panel 4 sont dus (au moins en partie) à une comptabilité erronée des ménages faisant partie de l'échantillon de l'EDTR qui n'apparaissent pas sur le fichier de l'échantillon. Au début d'un panel, on accepte que l'effort requis pour obtenir une réponse de certains ménages serait trop élevé pour les envoyer à la collecte et,

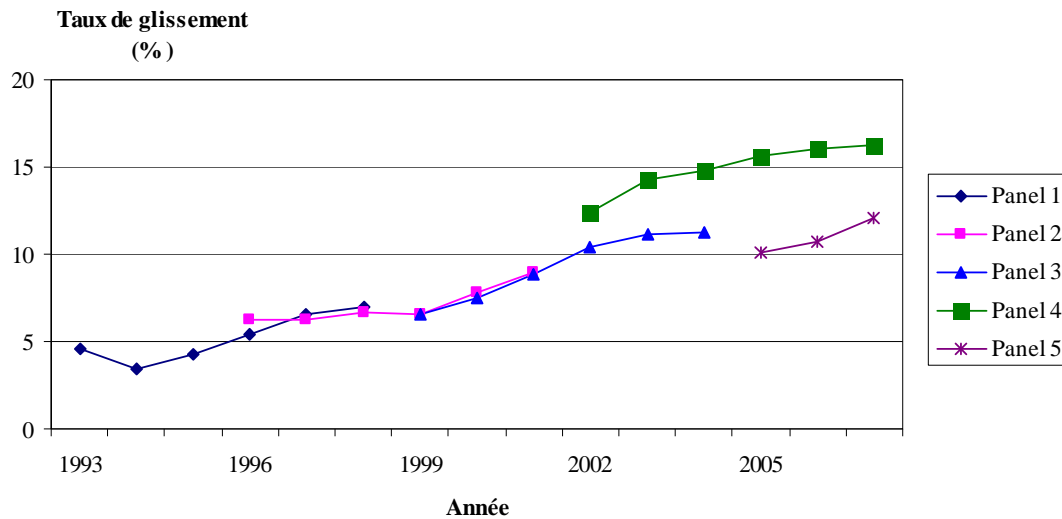
d'habitude, on les traite comme non-répondants pour la durée du panel. On estime que l'omission de ces ménages non-répondants a entraîné une augmentation du taux de glissement d'environ 2%. Cependant, l'impact sur les estimations de l'enquête devrait être négligeable, l'erreur étant en partie corrigée par le calage des poids finaux aux projections du recensement.

Tableau 4.1 Taux de glissement au niveau des personnes par province en 2007 (%)

| | T.-N.-L. | Î.-P.-É. | N.-É. | N.-B. | Qc | Ont. | Man. | Sask. | Alb. | C.-B. | Canada |
|---------|----------|----------|-------|-------|------|------|------|-------|------|-------|--------|
| Panel 4 | 8,7 | 7,7 | 10,0 | 12,4 | 12,9 | 21,5 | 15,7 | 11,9 | 22,3 | 25,3 | 18,8 |
| Panel 5 | 2,0 | 4,1 | 5,8 | 2,2 | 5,7 | 20,0 | 4,8 | 0,5 | 20,1 | 17,7 | 14,1 |

On remarque ici que le taux de glissement est plus élevé dans les provinces plus à l'Ouest (Alberta et Colombie-Britannique) ainsi qu'en Ontario, alors qu'il est plus bas dans les provinces plus à l'Est (Terre-Neuve et l'Île-du-Prince-Édouard) tant pour le 4^e que pour le 5^e panel. On a aussi enregistré un taux de glissement presque nul (0,5 %) en Saskatchewan pour le 5^e panel.

Graphique 4.2 Taux de glissement au niveau des ménages selon le panel et l'année de référence (%)



Le taux de glissement au niveau des ménages a toujours augmenté d'une vague à l'autre pour tous les panels à l'exception du 1^{er} entre la 1^{re} vague (en 1993) et la 2^e (1994) où il a quelque peu baissé.

Tableau 4.2 Taux de glissement au niveau des ménages selon la province et la taille du ménage en 2007 (%)

| Province | Panel 4 Taille des ménages | | | | Panel 5 Taille des ménages | | | |
|-----------------------|-------------------------------|-------------|-------------|-------------|-------------------------------|-------------|-------------|-------------|
| | 1 | 2 | 3 ou plus | Tous | 1 | 2 | 3 ou plus | Tous |
| Terre-Neuve | 12,1 | 3,7 | 4,9 | 6,0 | 3,0 | -3,0 | -1,0 | -0,9 |
| Île-du-Prince-Édouard | -3,5 | 5,7 | 4,1 | 2,9 | 15,6 | 2,3 | -2,2 | 3,8 |
| Nouvelle-Écosse | -0,9 | 6,2 | 13,2 | 6,6 | 4,0 | 10,7 | -1,4 | 4,7 |
| Nouveau-Brunswick | -3,3 | 17,2 | 11,5 | 9,9 | 3,6 | 0,2 | -1,9 | 0,3 |
| Québec | 15,9 | 13,9 | 10,4 | 13,4 | 11,1 | 5,5 | 4,4 | 7,0 |
| Ontario | 13,1 | 15,1 | 23,0 | 17,9 | 10,3 | 19,0 | 19,6 | 17,1 |
| Manitoba | 17,2 | 13,3 | 14,5 | 14,9 | -0,2 | -0,5 | 5,3 | 1,7 |
| Saskatchewan | -10,4 | 18,6 | 12,9 | 8,1 | -0,5 | -11,9 | 5,3 | -2,6 |
| Alberta | -2,0 | 13,5 | 28,8 | 15,9 | 6,2 | 20,3 | 19,8 | 16,6 |
| Colombie-Britannique | 27,7 | 22,8 | 23,5 | 24,5 | 14,6 | 17,0 | 16,4 | 16,1 |
| Canada | 13,3 | 15,2 | 19,5 | 16,3 | 9,6 | 12,5 | 13,6 | 12,1 |

Enfin, si on compare les taux de glissement selon la taille des ménages, on remarque pour le Canada que plus la taille des ménages est élevée, plus le taux de glissement augmente.

Pour les deux panels à l'étude, comme pour les taux de glissement des personnes, c'est en Ontario, en Alberta et en Colombie-Britannique où on enregistre les taux de glissement les plus élevés pour les ménages de taille 3 ou plus ainsi que pour l'ensemble des ménages. Ceux de la Colombie-Britannique au 4^e panel se démarquent plus particulièrement alors qu'ils sont tous au-dessus de la barre des 20 % peu importe la taille des ménages. Toujours pour le 4^e panel, pour les ménages composés de deux personnes, les taux de glissement sont plus bas dans les provinces Maritimes à l'exception du Nouveau-Brunswick qui montre un taux supérieur au taux canadien. Par ailleurs, on constate des taux négatifs dans la moitié des provinces canadiennes pour les ménages composés d'une seule personne soit à l'Île-du-Prince-Édouard, en Nouvelle-Écosse, au Nouveau-Brunswick, en Saskatchewan et en Alberta.

Du côté du 5^e panel, de façon générale, les taux de glissement sont plus bas que ceux du 4^e panel. Si les provinces avec les taux les plus élevés sont les mêmes, cette fois-ci, c'est en Saskatchewan que les taux sont les plus bas à l'exception des ménages de taille 3 ou plus. C'est au Manitoba qu'on observe la plus forte chute du taux de glissement alors que celui-ci est passé de 14,9 % pour le 4^e panel à 1,7 % pour le 5^e.

5. Taux de réponse

Puisque l'EDTR vise à fournir des estimations longitudinales et transversales, les taux de réponse doivent être calculés pour ces deux aspects. Les taux de réponse transversaux sont calculés à la fois au niveau des personnes et au niveau des ménages. Comme les personnes faisant partie de l'échantillon ont l'option d'accorder la permission qu'on

consulte leur dossier fiscal et d'éviter ainsi les questions sur le revenu, il est possible d'avoir des données complètes sur le revenu sans communiquer avec le répondant pendant l'année de référence. La définition d'un non-répondant n'est par conséquent pas une définition simple.

Si toutes les personnes d'un ménage constituent des non-répondants aux questions sur le travail et le revenu, ces personnes (et les ménages) seront classées en tant que non-répondants.

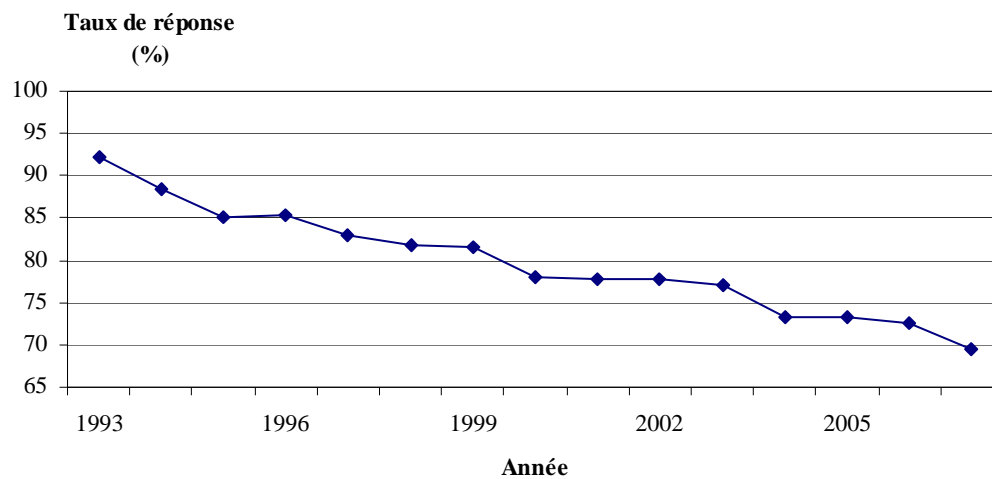
En ce qui concerne les personnes des ménages non-répondants aux questions sur le travail dans le cas desquelles nous disposons des données fiscales, on détermine si la personne fait partie du même ménage que l'année précédente (au 31 décembre). Si le ménage n'est pas le même, cela signifie que le répondant s'est séparé du ménage initial. Puisque nous n'avons pas le moindre renseignement sur la composition du nouveau ménage, ces personnes sont définies comme non-répondantes.

Les personnes des ménages non-répondants aux questions sur le travail dans le cas desquelles nous disposons toutefois des données sur le revenu et dont le ménage est le même que l'année précédente sont considérées comme des non-répondants si le ménage était un ménage non-répondant aux questions sur le travail en janvier précédent. Comme on effectue la collecte des mises à jour sur la composition des ménages en même temps que sont posées les questions sur le travail, on n'aura pas mis à jour la composition des ménages pendant deux années consécutives. Les personnes des ménages qui n'ont pas répondu aux questions sur le travail au cours de deux collectes de janvier consécutives sont en conséquence considérées comme des non-répondants de l'EDTR.

Le graphique 5.1 illustre les taux de réponse à l'EDTR des personnes de l'échantillon transversal au fil des années de l'enquête. On calcule les taux de réponse au niveau des personnes en divisant le nombre de répondants admissibles à l'échantillon transversal aux questions sur le travail et/ou le revenu par le nombre total de personnes admissibles à l'échantillon transversal. On suppose que les non-répondants sont encore dans la population cible, à moins d'indication contraire. Par conséquent, cela peut donner une certaine sous-estimation des taux de réponse. Le graphique 5.2 illustre les taux de réponse des ménages de l'échantillon transversal selon la région géographique.

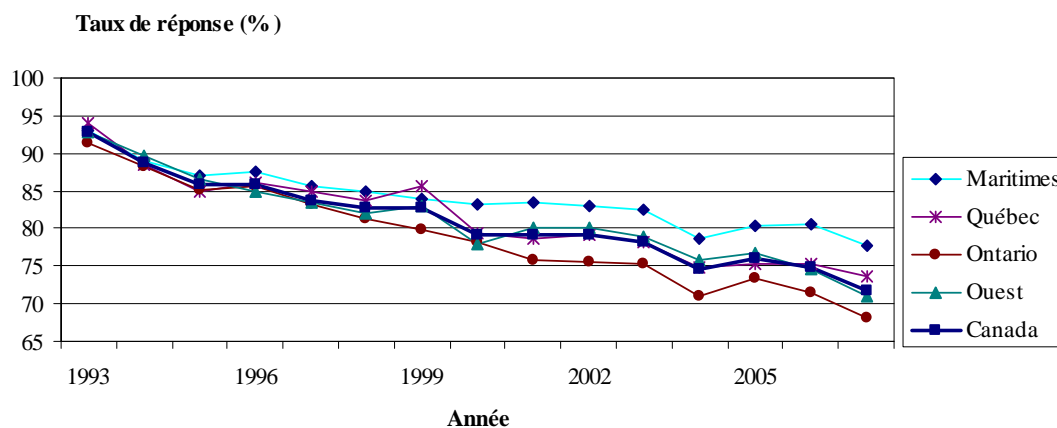
Le ménage est considéré comme ménage répondant si au moins une personne du ménage est considérée comme répondant. On calcule les taux de réponse des ménages en divisant le nombre de ménages répondants admissibles à l'échantillon transversal par le nombre total de ménages admissibles à l'échantillon transversal. Encore une fois, on fait une hypothèse : on suppose que les ménages non répondants sont encore dans la population cible. Cela peut donc amener une certaine sous-estimation des taux de réponse. La non-réponse risque d'introduire un biais dans les données. Il y a un biais lorsque les caractéristiques des répondants diffèrent de celles des non-répondants et que cette différence a un effet sur la variable observée. Il est difficile d'établir si la non-réponse introduit un biais étant donné que la quantité de renseignements disponibles pour les non-répondants est limitée.

Graphique 5.1 Taux de réponse des personnes de l'échantillon transversal selon l'année de référence (%)



Le graphique 5.1 nous montre clairement que le taux de réponse des personnes est en baisse constante depuis les débuts de l'enquête. Celui-ci a débuté à 92,1 % en 1993 pour atteindre un creux à 69,6 % en 2007.

Graphique 5.2 Taux de réponse des ménages de l'échantillon transversal selon la région géographique et l'année de référence (%)



Le graphique ci-dessus illustre encore une fois la tendance à la baisse du taux de réponse au fil des ans. On peut observer une baisse plus importante pour l'année 2004. Le taux est ensuite remonté en 2005 pour ensuite poursuivre sa baisse jusqu'à son niveau plancher de 71,8 % en 2007, pour le Canada. La courbe des taux pour les Maritimes montre les taux les plus élevés tandis que celle de l'Ontario montre les taux les plus faibles.

Le tableau 5.1 illustre les taux de réponse des personnes par phase. Les pourcentages de « répondants aux questions sur le travail » et de « répondants aux questions sur le revenu » correspondent aux pourcentages de personnes qui ont répondu seulement aux

séries de questions sur le travail ou sur le revenu, respectivement, alors que les « répondants aux deux séries » représentent le pourcentage de personnes qui ont répondu en totalité ou en partie aux deux séries de questions.

Tableau 5.1 Taux de réponse des personnes de l'échantillon transversal par phase selon l'année de référence¹ (%)

| Année | Répondants aux deux séries | Répondants aux questions sur le travail | Répondants aux questions sur le revenu | Non-réponse |
|-------|----------------------------|---|--|-------------|
| 1993 | 75,6 | 10,3 | 6,2 | 7,9 |
| 1994 | 75,1 | 10,5 | 2,8 | 11,6 |
| 1995 | 71,7 | 10,0 | 3,3 | 14,9 |
| 1996 | 71,6 | 10,8 | 2,9 | 14,6 |
| 1997 | 68,9 | 12,2 | 2,2 | 16,7 |
| 1998 | 68,8 | 10,4 | 2,6 | 18,2 |
| 1999 | 65,5 | 13,6 | 2,5 | 18,5 |
| 2000 | 56,1 | 17,3 | 4,6 | 22,0 |
| 2001 | 63,3 | 10,4 | 4,1 | 22,2 |
| 2002 | 61,6 | 10,8 | 5,4 | 22,2 |
| 2003 | 63,9 | 7,9 | 5,4 | 22,9 |
| 2004 | 62,3 | 5,8 | 5,1 | 26,8 |
| 2005 | 62,1 | 8,3 | 2,9 | 26,7 |
| 2006 | 59,3 | 7,2 | 6,0 | 27,5 |
| 2007 | 56,9 | 7,0 | 5,8 | 30,4 |

1. Depuis 2004, les entrevues sur le travail et sur le revenu sont combinées en une seule qui a lieu en janvier.

En lien avec le graphique 5.1, on remarque une baisse générale du taux pour les personnes qui ont répondu aux deux séries de question, le taux maximum ayant été enregistré lors de la première année de l'enquête (75,6 %) et le taux minimum lors de la dernière année (56,9 %). En lien avec ces taux, ce n'est pas surprenant de constater que la proportion de non-répondants a atteint son apogée en 2007 alors qu'elle était à son plus bas à ses débuts en 1993.

Par contre, si on analyse les taux des répondants qui n'ont répondu qu'à une seule série de questions, la tendance est différente. Pour la série de questions sur le travail, à part en 1997, 1999 et 2000, le taux a oscillé autour de la barre des 10 % entre 1993 et 2002. Par la suite, il est tombé beaucoup plus bas, variant entre 5,8 % et 8,3 %. Quant à la série de questions sur le revenu, après avoir été relativement stable entre 1994 et 1999, le taux a doublé pour ensuite se stabiliser entre 4,6 % et 6,0 % exception faite de l'année 2005 où le taux a descendu à 2,9 %.

Étant donné qu'il est difficile de définir conceptuellement le ménage de l'échantillon longitudinal, seuls les taux de réponse longitudinaux au niveau des personnes sont calculés. Le tableau 5.2 montre ces derniers pour chacun des quatre panels. On calcule ces taux en divisant le nombre de répondants de l'échantillon longitudinal par le nombre initial de personnes de l'échantillon longitudinal choisies dans ce panel.

Tableau 5.2 Taux de réponse des personnes de l'échantillon longitudinal selon le panel et la vague (%)

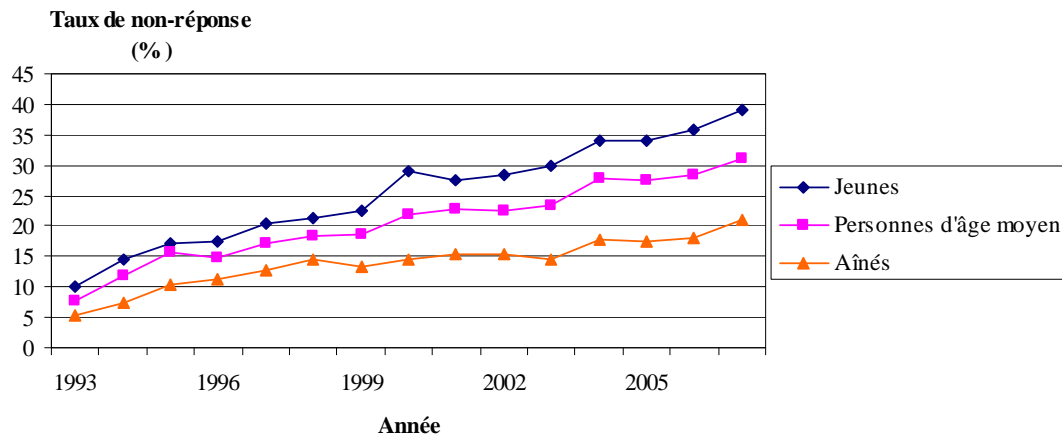
| Panel (année de début) | Vague du panel | | | | | |
|------------------------|----------------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Panel 1 (1993) | 93,3 | 89,6 | 86,5 | 83,9 | 82,6 | 81,5 |
| Panel 2 (1996) | 89,5 | 86,8 | 85,2 | 82,7 | 78,5 | 77,4 |
| Panel 3 (1999) | 83,9 | 83,0 | 83,0 | 79,6 | 76,4 | 73,7 |
| Panel 4 (2002) | 81,2 | 83,2 | 78,3 | 75,0 | 71,6 | 68,9 |
| Panel 5 (2005) | 78,8 | 80,6 | 77,3 | ... | ... | ... |

... N'ayant pas lieu de figurer

Le tableau 5.2 fait ressortir la diminution progressive du taux de réponse longitudinal par panel d'une vague à l'autre. La différence étant que les taux de réponse à la première vague débutent avec un taux plus bas d'un panel à l'autre. Par exemple, le taux est passé de 93,3 % à la vague 1 à 81,5 % à la vague 6 pour le 1^{er} panel, alors qu'il a chuté de 81,2 % à la vague 1 à 68,9 % à la vague 6 pour le 4^e panel.

Le graphique 5.3 illustre les taux de non-réponse longitudinale enregistrés chaque année, par groupes d'âge tel que défini au début du panel. Les « jeunes » sont des personnes ayant au moins 16 ans mais pas plus de 30 ans. Les « personnes d'âge moyen » correspondent aux personnes de 30 ans ou plus mais de moins de 60 ans, et les « aînés » sont les personnes âgées d'au moins 60 ans. À noter que les groupes d'âge sont définis au début du panel.

Graphique 5.3 Taux de non-réponse longitudinal selon le groupe d'âge (%)



Enfin, si on veut comparer les taux de non-réponse longitudinaux selon différents groupes d'âge, il ressort clairement du graphique 5.3, que chacune des courbes montre une tendance à la hausse de son taux de réponse. Si on compare les taux entre 1993 et 2007, on peut voir que ces derniers ont quadruplé au cours des 15 ans de l'EDTR. Les jeunes, soit ceux âgés entre 16 et 30 ans montrent des taux de réponse environ deux fois plus élevés que les aînés qui sont âgés de plus de 60 ans. En effet, en 2007, 39,2 % des jeunes n'avaient pas répondu à l'enquête comparativement à 20,9 % des aînés. Ces résultats sont

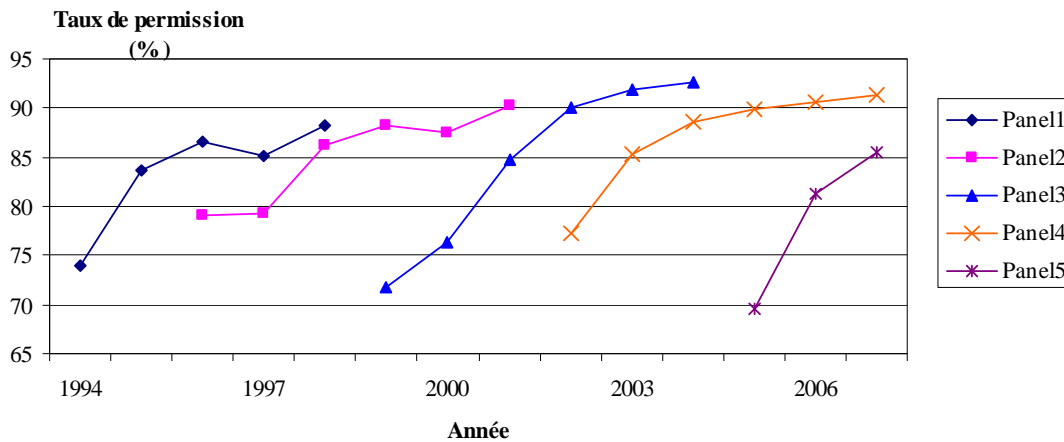
peu surprenants étant donné que les jeunes sont plus difficiles à joindre en général que les aînés qui sont plus souvent chez eux.

6. Taux de permission d'accès au dossier fiscal

Avant l'année de référence 2004, on tenait deux entrevues par année : l'entrevue de janvier concernait des activités comme le travail, la fréquentation scolaire, la recherche de travail ou la retraite. La seconde entrevue, en mai, touchait le revenu, mais elle n'était pas nécessaire si le répondant accordait à Statistique Canada la permission d'obtenir les données requises de son dossier fiscal. La source fiscale devrait livrer des données cohérentes de grande qualité, de sorte qu'un taux élevé de permission devrait assurer des estimations de revenu de grande qualité. Statistique Canada demandait cette permission au répondant à la fin de l'entrevue de janvier. Si le répondant refusait, on communiquait encore une fois avec lui en mai. À ce moment-là, on lui demandait de nouveau s'il préférerait donner la permission qu'on accède à son dossier fiscal. S'il refusait, l'entrevue avait lieu. Depuis l'année de référence 2004, on demande cette permission seulement une fois, en janvier. Si le répondant n'accorde pas sa permission, l'entrevue se poursuit aussitôt avec les questions sur le revenu.

Le graphique 6.1 montre les taux annuels de permission pour l'enquête. L'option de donner la permission de consultation du dossier fiscal a été donnée pour la première fois lors de la collecte de mai pour l'année de référence 1994. Avant cela, toutes les données sur le revenu étaient recueillies par voie d'entrevue. Les pourcentages du graphique sont fondés sur le nombre de répondants de plus de 15 ans qui sont admissibles à l'échantillon transversal. La permission d'accès au dossier fiscal est valide jusqu'à la fin de la durée de vie du panel sauf avis contraire du répondant. Par conséquent, pour une année donnée, l'effet cumulatif de la permission peut masquer l'effort fourni à la collecte pour obtenir la permission de nouveaux répondants.

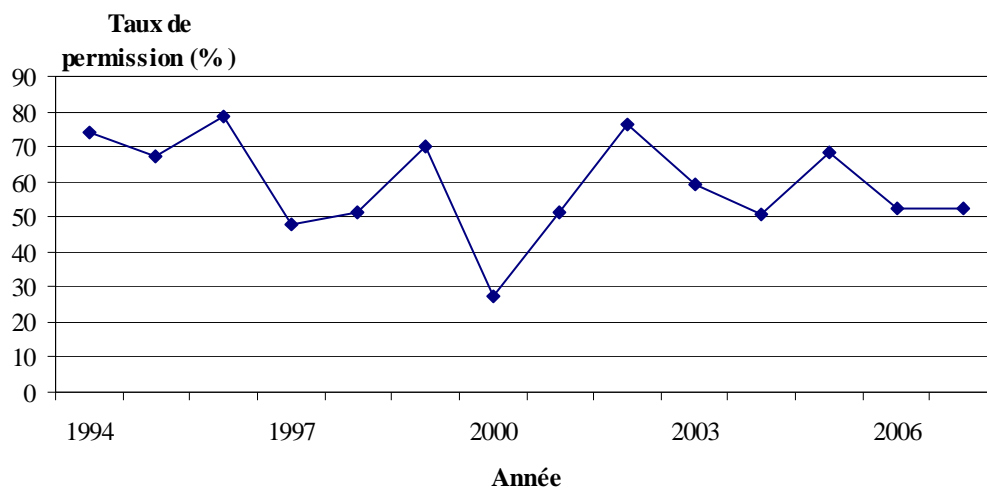
Graphique 6.1 Taux de permission par panel selon l'année de référence (%)



On remarque des tendances similaires dans le taux de permission annuel pour chacun des panels. Le taux a subi une forte augmentation dans les trois premières vagues, à l'exception de la 2^e vague du 2^e panel où le taux est demeuré stable. Ensuite, il a poursuivi sa hausse, mais de façon moins importante lors des trois dernières. On a même enregistré une baisse du taux de permission entre les 4^e et 5^e vagues pour les deux premiers panels, mais ce dernier a terminé à la hausse lors de la dernière vague.

Le graphique 6.2 présente le taux de permission annuel des nouveaux répondants qui ont donné la permission d'accès au dossier fiscal selon l'année de référence.

Graphique 6.2 Taux de permission chez les nouveaux répondants selon l'année de référence (%)



Le taux de permission des nouveaux répondants a subi de fortes variations au cours de la période à l'étude. Il a fluctué entre 27,2 % en 2000 et 78,4 % en 1996. On remarque aussi que l'année d'introduction d'un nouveau panel (1996, 1999, 2002 et 2005) a toujours enregistré les plus hauts taux de permission chez les nouveaux répondants. Également, le taux était très bas en 2000 car il s'agissait de la première année de référence pour laquelle l'entrevue du mois de mai n'a pas eu lieu.

7. Taux de couplage avec le dossier fiscal

Les répondants peuvent accorder à Statistique Canada la permission d'utiliser leurs données fiscales, mais on ne leur demande pas leur numéro d'assurance sociale (NAS). Sans NAS pour repérer les répondants à l'EDTR dans le fichier de données fiscales, il faut effectuer une opération de couplage pour trouver le NAS du répondant. Le Système généralisé de couplage d'enregistrements (SGCE) mis au point à Statistique Canada sert à effectuer ce couplage.

Après un traitement préliminaire du fichier de données fiscales et du fichier de l'EDTR pour vérifier la compatibilité du formatage de toutes les variables d'appariement, on effectue un appariement direct à l'aide de 7 principales variables d'appariement. Ces variables d'appariement sont : le sexe, la province, le code Soundex¹ pour le nom de famille, le nom de famille, la date de naissance, le code postal et la première initiale. L'enregistrement de l'EDTR ne peut pas avoir de données manquantes pour les variables d'appariement clés. Les résultats de l'appariement direct sont soumis à une opération manuelle de détection des erreurs lorsqu'un enregistrement de l'EDTR correspond à plus d'un enregistrement fiscal, lorsque plus d'un enregistrement fiscal correspond à un enregistrement de l'EDTR, et lorsque le premier prénom n'est pas le même dans les 2 sources (seule la première initiale est utilisée dans l'appariement). Le taux d'appariement dans l'appariement direct est d'environ 55 %.

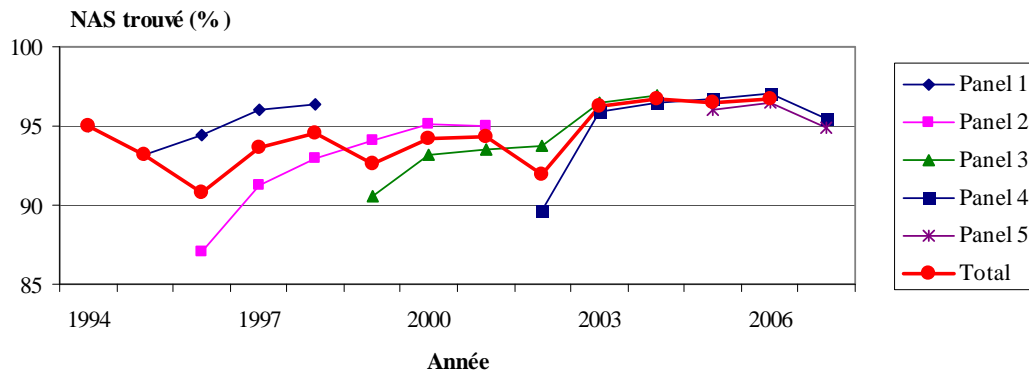
Les dossiers non appariés sont alors soumis à un appariement statistique. On définit des cases² pour l'appariement. Les fichiers sont segmentés en cases, où les variables déterminantes sont le sexe, la province et le nom de famille en code Soundex. Chaque enregistrement d'une case dans le fichier de l'EDTR est comparé à chaque enregistrement de la même case dans le fichier fiscal. On attribue des facteurs d'importance pour la concordance parfaite, la concordance partielle et la non-concordance. Ces facteurs sont des valeurs numériques, qui servent à évaluer la probabilité qu'une paire d'enregistrements (un de l'EDTR et un du dossier fiscal) représente la même personne. On définit des facteurs pour chacune des variables d'appariement. On définit des seuils où l'on détermine qu'il y a concordance certaine entre les enregistrements si leur facteur total dépasse le seuil supérieur ou non-concordance certaine si leur facteur total est inférieur au seuil inférieur. On effectue une vérification manuelle pour s'assurer de la qualité des appariements.

Le graphique 7.1 donne le pourcentage des répondants de l'échantillon de l'EDTR qui donnent leur permission et pour lesquels un NAS a été trouvé. Puisque certains répondants qui donnent leur permission n'ont pas produit de déclaration de revenus, il n'est pas possible d'obtenir un couplage positif dans tous les cas où un NAS est trouvé.

1 .Soundex est le nom d'une routine de codage du nom qui sert à éliminer les erreurs d'orthographe fréquentes dans les noms de famille des répondants. Ce codage s'effectue en fonction de la prononciation du nom de famille.

2. Les cases sont des groupes de particuliers dans le fichier de données fiscales et le fichier de l'EDTR, pour lesquels le sexe, la province et le code Soundex sont les mêmes.

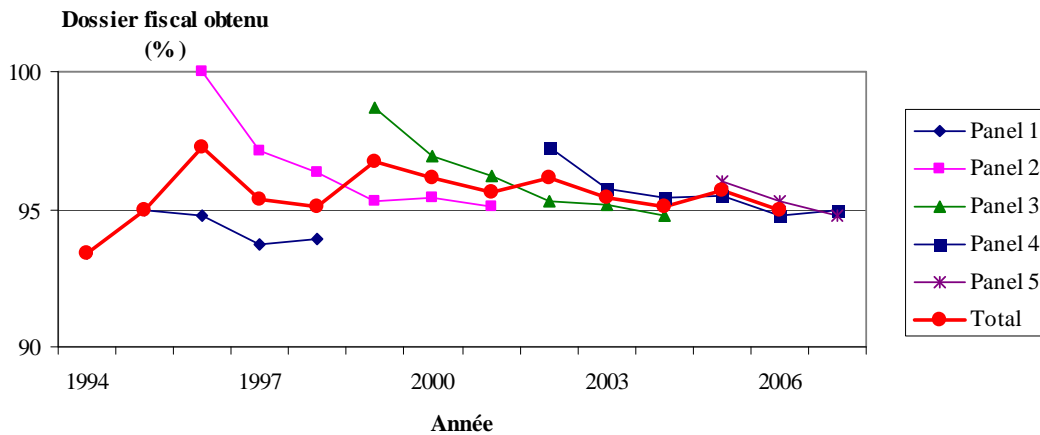
Graphique 7.1 Personnes ayant donné leur permission et pour lesquels le NAS a été trouvé selon l'année de référence (%)



En général, le pourcentage de répondants ayant donné leur permission et pour lesquels le NAS a été trouvé a montré une tendance à la hausse lors des 6 vagues de chacun des panels. On a observé une hausse plus importante entre la 1^{re} et la 2^e vague tandis que la pente de la courbe devient moins importante lors des vagues subséquentes. Entre la 5^e et 6^e vague, le taux s'est stabilisé et a même diminué quelque peu pour certains panels.

Le graphique 7.2 donne les taux de couplage pour les personnes de l'échantillon de l'EDTR pour lesquelles nous avons réussi à trouver un NAS.

Graphique 7.2 Taux de couplage lorsque le NAS a été trouvé selon l'année de référence (%)



En ce qui a trait au taux de couplage, celui-ci montre la même tendance pour tous les panels. Lors de la première vague, le taux est à son maximum et puis il descend graduellement pour se stabiliser autour des 95 % lors de la dernière vague. La courbe globale illustre bien ce scénario alors que le taux est à son maximum lors de l'année d'introduction d'un nouveau panel (1996, 1999, 2002 et 2005) tandis qu'il a baissé lors des autres années.

Toutefois, on remarque que les taux de couplage débutent à un seuil de plus en plus bas au fil des panels 2 à 5. En effet, lors de la première vague du panel 2, le taux de couplage était à 100 % alors qu'il était à 98,7 %, 97,2 % et 96,0 % lors de la première vague des panels 3, 4 et 5 respectivement.

Enfin, le tableau 7.1 compare la provenance des données en pourcentage selon qu'elles ont été obtenues à partir des dossiers fiscaux ou via une entrevue téléphonique.

Tableau 7.1 Proportion des répondants venant du dossier fiscal ou de l'entrevue selon l'année de référence³ (%)

| Année | Dossier fiscal | Entrevue | Autres ¹ |
|-------|----------------|----------|---------------------|
| 1999 | 71,9 | 12,0 | 16,2 |
| 2000 | 74,0 | 0,0 | 26,0 |
| 2001 | 78,9 | 5,0 | 16,1 |
| 2002 | 74,2 | 8,8 | 17,0 |
| 2003 | 81,4 | 5,2 | 13,4 |
| 2004 | 83,4 | 5,0 | 11,7 |
| 2005 | 73,6 | 9,8 | 16,6 |
| 2006 | 78,8 | 5,9 | 15,3 |
| 2007 | 79,8 | 4,7 | 15,5 |

0 Zéro absolu ou valeur arrondie à zéro.

1. Comprend les enregistrements qui ne sont pas couplés avec le dossier fiscal et sans réponse aux questions sur le revenu.

On remarque dans le tableau ci-dessus que la majorité des enregistrements provient des dossiers fiscaux. Les proportions varient entre 71,9 % en 1999 et 83,4 % en 2004. Depuis 2003, ce pourcentage fluctue autour de 80 % sauf en 2005 où il était plus bas (73,6 %). Par le fait même, c'est aussi en 2005 qu'on observe une proportion d'enregistrements d'entrevue plus élevés soit près de 10 % alors que le pourcentage était plus autour de 5 % pour les autres années entre 2003 et 2007.

8. Taux d'imputation

Pour compenser pour les ménages non répondants de l'échantillon de l'EDTR, on applique une correction de la non-réponse aux poids de l'EDTR. Cependant, on conserve dans l'échantillon les ménages qui ont donné une réponse partielle et on impute les données sur le revenu qui sont manquantes pour des personnes des ménages qui ont répondu. Il peut être nécessaire de procéder à une imputation complète de toutes les variables du revenu de ces personnes, ou seulement de certaines zones. On peut considérer que les taux d'imputation dans l'EDTR sont une mesure de la réponse partielle à l'enquête.

3. En excluant les enregistrements qui ne sont pas admissibles à l'imputation sur le revenu.

L'imputation des variables du revenu dans l'EDTR se fait soit de manière transversale ou longitudinale. L'imputation longitudinale consiste à utiliser les données de la vague précédente pour imputer les données manquantes de la vague courante. La méthode transversale utilise l'approche du plus proche voisin. On définit un ensemble de règles de cohérence de base et, pour un enregistrement donné nécessitant l'imputation, un ensemble de donneurs. On définit également un ensemble de variables d'appariement, dont chacune est corrélée avec les variables à imputer. Par l'utilisation combinée d'une fonction de score (pour les variables d'appariement catégoriques) et une fonction de distance (pour les variables d'appariement numériques), on trouve l'enregistrement donneur le plus semblable au receveur et on utilise les données de l'enregistrement pour imputer. Il faudrait enfin souligner que certaines variables de l'enquête sont également imputées de manière déterministe.

Le tableau 8.1 indique le pourcentage des personnes des ménages répondants de l'EDTR qui ont été soumises à une imputation totale ou partielle. Il faut se rappeler qu'un ménage répondant de l'EDTR est un ménage qui compte au moins un membre ayant répondu en partie ou en totalité aux questions sur le travail ou sur le revenu de l'enquête. Au total, dix-huit variables du revenu sont imputées pendant l'imputation du revenu dans l'EDTR. Pour un grand nombre de personnes, une imputation partielle suffit. L'imputation est dite partielle lorsqu'un ou plusieurs postes de revenu sont imputés à l'aide de renseignements fournis par le particulier.

Tableau 8.1 Répondants pour lesquels des variables du revenu ont dû être imputées selon la province en 2007 (%)

| Province | Imputation totale ¹ | Imputation partielle ² | Pas d'imputation |
|-----------------------|--------------------------------|-----------------------------------|------------------|
| Terre-Neuve | 1,7 | 20,1 | 78,2 |
| Île du Prince Édouard | 2,1 | 19,3 | 78,6 |
| Nouvelle-Écosse | 2,5 | 20,7 | 76,8 |
| Nouveau-Brunswick | 1,8 | 19,9 | 78,3 |
| Québec | 1,9 | 18,4 | 79,7 |
| Ontario | 3,0 | 24,3 | 72,7 |
| Manitoba | 2,3 | 24,1 | 73,6 |
| Saskatchewan | 2,1 | 21,8 | 76,1 |
| Alberta | 3,2 | 25,1 | 71,7 |
| Colombie-Britannique | 2,5 | 25,7 | 71,8 |
| Canada | 2,5 | 22,4 | 75,1 |

1. Pas de renseignements fournis par le répondant. Toutes les données ont été imputées.

2. Une ou plusieurs données ont été imputées à l'aide de certains renseignements fournis par le répondant.

Le tableau ci-dessus nous montre que le trois quarts des enregistrements n'ont nécessité aucune imputation pour le Canada. C'est au Québec où le taux d'imputation fut le plus bas alors que près de 80 % des enregistrements de la province n'ont pas été imputés. Par contre, c'est dans l'Ouest canadien (Alberta et Colombie-Britannique) où on a dû avoir le plus recours à l'imputation. Les taux d'imputation partielle ont franchi la barre des 25 % pour ces deux provinces.

Peu d'enregistrements ont nécessité de l'imputation totale, alors que les taux varient entre 1,7 % et 3,2 % dans les différentes provinces canadiennes.

Au tableau 8.2, nous comparons le pourcentage des enregistrements de données fiscales nécessitant une imputation avec le pourcentage des enregistrements pour lesquels les données sont recueillies par l'entrevue téléphonique. On détermine la nécessité d'imputer partiellement la section du revenu après avoir combiné les réponses aux questions sur le travail et sur le revenu. On corrige les cas incohérents lors du processus d'imputation.

Le tableau montre également le pourcentage des personnes soumises à l'imputation partielle pour lesquelles il faut imputer entre une et dix-sept variables.

Tableau 8.2 Enregistrements fiscaux et enregistrements d'entrevue nécessitant une imputation totale ou partielle en 2007 (%)

| Imputation | Source des données | | | Tous |
|-------------------------------|--------------------|--------------|---------------------|--------------|
| | Dossiers fiscaux | Entrevue | Autres ¹ | |
| Partielle (1 variable) | 8,4 | 13,7 | 0,0 | 7,3 |
| Partielle (2 à 9 variables) | 0,3 | 37,1 | 0,0 | 2,0 |
| Partielle (10 à 17 variables) | 0,0 | 0,1 | ... | 13,1 |
| Totale | ... | ... | 100 | 2,5 |
| Pas d'imputation | 91,3 | 49,1 | ... | 75,1 |
| Total | 100,0 | 100,0 | 100,0 | 100,0 |

... N'ayant pas lieu de figurer

0 Zéro absolu ou valeur arrondie à zéro.

1. Enregistrements qui n'ont pu être appariés aux données fiscales et qui n'ont pas d'information sur le revenu. Certains de ces cas ont partiellement été imputés à l'aide d'information tirée de l'entrevue sur le travail.

Le tableau précédent montre que les enregistrements pour lesquels on a pu obtenir les données fiscales ont nécessité très peu d'imputation. En effet, 91,3 % des enregistrements n'ont pas demandé d'imputation et 8,4 % ont exigé de l'imputation partielle pour une seule variable, ce qui nous amène tout près des 100 %.

Pour ce qui est des enregistrements obtenus via l'entrevue, environ la moitié n'a requis aucune imputation et plus du tiers ont exigé une imputation partielle de 2 à 9 variables, ce qui est en contraste avec les enregistrements fiscaux.

En 2002, en collaboration avec la Société canadienne d'hypothèques et de logement (SCHL), on a ajouté à l'EDTR de nouvelles questions sur le logement au profit de la recherche et de l'élaboration des politiques de logement. L'enquête réunit désormais les renseignements suivants depuis l'année de référence 2002 : la nécessité de réparations (selon l'occupant du logement); le combustible principal utilisé pour le chauffage du logement; et la présence d'une ferme ou d'une entreprise dans la propriété. Également, dans le cas des propriétaires-occupants : le montant des paiements hypothécaires périodiques; le montant des impôts fonciers annuels; et si le logement est un logement en condominium. Dans le cas des locataires, on recueille les variables suivantes : le montant

du loyer mensuel, les commodités incluses dans le loyer (*p. ex.*, chauffage, eau, électricité); et si le loyer est subventionné par le gouvernement ou un employeur.

Les renseignements qui précèdent s'ajoutent aux renseignements sur la propriété et le type de logement (depuis 1994) et aux renseignements sur la présence d'un prêt hypothécaire et le nombre de chambres à coucher dans le logement (depuis 1999).

À cause de la non-réponse à certaines questions, il a fallu commencer à imputer des variables du logement dans l'EDTR en 2002. Deux méthodes d'imputation ont été utilisées : l'imputation longitudinale et l'imputation transversale par donneur. L'imputation transversale par enregistrement donneur utilise une méthode semblable à celle utilisée pour l'imputation du revenu, faisant appel à la fonction de pointage décrite plus haut.

Le tableau 8.3 montre le pourcentage des ménages répondants de l'EDTR qui ont été soumis à une imputation totale ou partielle.

Tableau 8.3 Ménages nécessitant une imputation selon la province en 2007 (%)

| Province | Imputation totale ¹ | Imputation partielle ² | Pas d'imputation |
|-----------------------|--------------------------------|-----------------------------------|------------------|
| Terre-Neuve | ... | 38,3 | 61,7 |
| Île du Prince Édouard | ... | 38,6 | 61,4 |
| Nouvelle-Écosse | ... | 35,8 | 64,2 |
| Nouveau-Brunswick | ... | 36,0 | 64,0 |
| Québec | ... | 29,7 | 70,3 |
| Ontario | ... | 40,8 | 59,2 |
| Manitoba | ... | 44,0 | 56,0 |
| Saskatchewan | ... | 40,8 | 59,2 |
| Alberta | ... | 43,6 | 56,4 |
| Colombie-Britannique | ... | 46,0 | 54,0 |
| Canada | ... | 38,9 | 61,1 |

1. Pas de renseignements fournis par le répondant. Toutes les données sont imputées.

2. Un ou plusieurs éléments sont imputés à l'aide de certains renseignements fournis par le répondant.

Pour l'imputation au niveau des ménages, près de 40 % des ménages ont eu besoin d'imputation partielle. Les taux d'imputation les plus élevés ont été observés dans l'Ouest canadien (au Manitoba, en Alberta et en Colombie-Britannique), tandis que le taux le plus bas a été enregistré au Québec. Il s'agit de la seule province ayant montré un taux sous la barre des 30 %.

Au total, vingt variables reliées au ménage sont imputées dans le cadre de l'imputation des variables du ménage dans l'EDTR. Un grand nombre de ménages ne nécessitent qu'une imputation partielle. Le tableau 8.4 montre la répartition de ceux qui nécessitent une imputation partielle.

Tableau 8.4 Ménages nécessitant une imputation selon le nombre de variables demandant l'imputation et l'année de référence (%)

| Année | Nombre de variables ménages ayant besoin d'imputation | | | |
|-------|---|-------|--------|------------|
| | 1 | 2 à 5 | 6 à 19 | Un ou plus |
| 2004 | 10,5 | 9,9 | 10,9 | 31,3 |
| 2005 | 10,2 | 10,1 | 15,7 | 36,0 |
| 2006 | 10,0 | 7,1 | 22,6 | 39,7 |
| 2007 | 9,8 | 6,6 | 22,5 | 38,9 |

Enfin, si on regarde les taux d'imputation partielle au niveau des ménages selon le nombre de variables requérant une imputation, on peut remarquer qu'en 2006 et 2007, les taux sont plus élevés que ceux des deux années précédentes. Plus particulièrement, dans la catégorie 6 à 19 variables ayant besoin d'imputation, la proportion de ménages a plus que doublé en 2 ans, passant de 10,9 % en 2004 à 22,6 % en 2006, pour ensuite se stabiliser en 2007.

9. Arrondissement des données sur le revenu

Un faible pourcentage de données sur le revenu dans l'EDTR provient des données recueillies dans le cadre d'une entrevue téléphonique. On peut considérer que la plupart des données obtenues du fichier de données fiscales sont cohérentes mais on dénote que la qualité des données obtenues par voie de collecte comporte certaines lacunes. Un certain nombre de répondants peuvent se reporter à leurs documents d'impôt et donner des montants précis, mais il est possible qu'un grand nombre des réponses soient des estimations ou des approximations, donc exprimées en centaines ou milliers de dollars plutôt qu'en dollars et en cents.

Pour vérifier la présence possible d'un arrondissement, on a produit des distributions de chacune des tranches des 4 derniers chiffres des variables déclarées. On s'attendrait normalement que la distribution soit à peu près uniforme, et que les chiffres 0 à 9 forment chacun environ 10 % de la distribution. On pourrait conclure que les montants ont été arrondis à la dizaine près si les 0 sont fréquents dans le dernier chiffre, à la centaine près s'ils sont fréquents dans l'avant-dernier chiffre également, etc. Le tableau 9.1 indique la distribution de chacun de ces chiffres pour toutes les valeurs déclarées supérieures à 10 000 de la variable « salaires et traitements », à la fois pour les données recueillies (c.-à-d. recueillies par entrevue) et pour les données fiscales.

Tableau 9.1 Distribution des quatre derniers chiffres des salaires et traitements supérieur à 9 999 \$ en 2007 (%)

| Chiffre | Quatrième chiffre de la fin | | Troisième chiffre de la fin | | Avant-dernier chiffre | | Dernier chiffre | |
|---------|-----------------------------|----------------|-----------------------------|----------------|-----------------------|----------------|-----------------|----------------|
| | Recueilli | Dossier fiscal | Recueilli | Dossier fiscal | Recueilli | Dossier fiscal | Recueilli | Dossier fiscal |
| 0 | 33,2 | 11,2 | 90,8 | 12,1 | 95,8 | 12,9 | 96,9 | 14,2 |
| 1 | 3,6 | 10,9 | 0,5 | 10,0 | 0,3 | 9,7 | 0,4 | 9,3 |
| 2 | 9,4 | 10,7 | 0,9 | 9,9 | 0,2 | 9,8 | 0,4 | 9,9 |
| 3 | 5,6 | 10,2 | 0,8 | 9,5 | 0,5 | 9,4 | 0,3 | 9,4 |
| 4 | 6,2 | 9,9 | 0,8 | 9,6 | 0,7 | 9,6 | 0,3 | 9,5 |
| 5 | 19,6 | 10,0 | 2,8 | 10,2 | 0,5 | 9,8 | 0,2 | 9,8 |
| 6 | 5,9 | 9,8 | 1,1 | 9,8 | 0,4 | 9,7 | 0,4 | 9,6 |
| 7 | 4,8 | 9,3 | 0,5 | 9,6 | 0,3 | 9,6 | 0,4 | 9,5 |
| 8 | 8,6 | 9,5 | 1,1 | 9,7 | 0,9 | 9,8 | 0,3 | 9,5 |
| 9 | 3,3 | 8,4 | 0,8 | 9,6 | 0,4 | 9,8 | 0,4 | 9,2 |

Le tableau 9.1 illustre clairement que les salaires et traitements de 10 000 \$ ou plus ont été arrondis. Le troisième, deuxième et dernier chiffre de la fin a été un zéro dans 90,8 %, 95,8 % et 96,9 % des cas respectivement pour les données qui ont été recueillies, alors que la distribution est plus uniforme pour chacun des nombres de 0 à 9 pour les données provenant des dossiers fiscaux.

En ce qui a trait aux données recueillies, pour le 4^e chiffre de la fin, environ le tiers des enregistrements ont montré un zéro et près de 20 % ont montré un cinq. Même si ces résultats ne sont pas aussi frappants que pour les trois derniers chiffres, ils montrent quand même qu'il y a eu arrondissement.

Le tableau 9.2 indique la fréquence de zéros dans chacun des quatre derniers chiffres pour toutes les valeurs autres que zéro qui ont été déclarées pour certaines variables de l'EDTR.

Tableau 9.2 Proportion de zéros dans les quatre derniers chiffres des données déclarées pour certaines variables en 2007 (%)

| Variable | Chiffre | | | |
|---|---------------------|---------------------|---------------|---------|
| | Quatrième de la fin | Troisième de la fin | Avant-dernier | Dernier |
| Salaires et traitements | 26,7 | 82,2 | 93,7 | 96,1 |
| Revenu de placements | 11,5 | 29,6 | 59,6 | 71,9 |
| Aide sociale | 12,8 | 25,6 | 69,8 | 86,0 |
| Prestations d'assurance-emploi | 6,3 | 42,2 | 85,7 | 95,2 |
| Revenu de travail autonome non agricole | 37,3 | 82,8 | 98,2 | 97,6 |

Ces derniers résultats illustrent sans l'ombre d'un doute l'augmentation graduelle du nombre de 0 quand on observe l'emplacement du chiffre de la fin en passant du 4^e chiffre

jusqu'au dernier. Pour les salaires et traitements et le revenu de travail autonome non agricole, on observe un plus grand nombre de zéros que les autres variables dès le 4^e chiffre avant la fin et de façon plus marquée pour le 3^e.

Pour les revenus de placements, l'aide sociale et les prestations d'assurance-emploi, on remarque une forte hausse du nombre de zéros en comparant le 3^e chiffre avant la fin avec le 2^e chiffre. Ces augmentations varient de 30,0 % à 44,2 %.

Enfin, toutes les variables ont la valeur zéro lors du dernier chiffre dans plus de 95 % des cas, sauf pour les revenus de placement (71,9 %) et l'aide sociale (86,0 %) où la proportion de zéros a quand même été élevée.