

La qualité des données dans l'Enquête sur la dynamique du travail et du revenu (EDTR) en 2010

Wei Qian, Wisner Jocelyn

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

Table des matières

1. Introduction	3
2. Composition/érosion de l'échantillon	4
3. Erreurs d'échantillonnage	6
4. Erreurs de couverture	8
5. Taux de réponse	11
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. Taux d'entrevues par procuration	25
10. 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 a débuté en 1993. Elle permet 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 deux panels. Chaque panel demeure dans l'enquête pendant six années consécutives et un nouveau panel y entre tous les trois 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. Ils posent des questions sur le travail (l'activité sur le marché du travail, l'expérience de travail, les épisodes de chômage et l'emploi même) le niveau de scolarité et les sources de revenu.

Avant l'année de référence 2004, les répondants pouvaient être interviewés en janvier et en mai. L'entrevue du mois de mai avait pour but de collecter les données fiscales des répondants n'ayant pas donné la permission d'apparier avec leur dossier fiscal. Toutefois, depuis 2004, l'entrevue en mai a été abandonnée dans le but d'économiser sur les coûts de collecte. Ainsi, 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.

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 personnes 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/bsolc/olc-cel/olc-cel?lang=fra&catno=75F0011X>.

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 afin d'aider les 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 faits saillants des indicateurs de qualité pour l'année de référence 2010 pour le Canada.

Tableau 1.1 Principaux indicateurs de qualité de l'EDTR pour le Canada, 2010

Indicateur	Statistique
Taille d'échantillon longitudinal <ul style="list-style-type: none"> • Panel 5 • Panel 6 	42 330 40 912
Taille d'échantillon transversal (longitudinal admissible et cohabitants) <ul style="list-style-type: none"> • Panel 5 • Panel 6 	31 885 31 739
Coefficient de variation <ul style="list-style-type: none"> • Revenu total médian 	0,8%
Taux de glissement pour les personnes	13,5%
Taux de glissement pour les ménages <ul style="list-style-type: none"> • Ménages de taille 1 • Ménages de taille 2 • Tous 	8,4% 13,0% 11,8%
Taux de réponse <ul style="list-style-type: none"> • Transversal, personne • Transversal, ménage • Longitudinal, personne <ul style="list-style-type: none"> ○ Panel 5 ○ Panel 6 	66,0% 67,3% 67,2% 71,9%
Taux de permission <ul style="list-style-type: none"> • Panel 5 • Panel 6 	90,2% 86,6%
Taux de couplage avec le dossier fiscal (NAS trouvé)	95,1%
Taux d'imputation pour les personnes <ul style="list-style-type: none"> • Imputation totale • Imputation partielle 	2,6% 20,9%
Taux d'imputation pour les ménages <ul style="list-style-type: none"> • Imputation partielle 	39,6%

2. Composition/érosion de l'échantillon

La distribution de l'échantillon longitudinal par province et RMR en 2010 est montrée dans les tableaux 2.1 et 2.2 qui suivent. Seuls les répondants dans le champ de l'enquête ont une province ou une RMR.

Tableau 2.1 Nombre de personnes dans l'échantillon longitudinal, selon la province et le panel, 2010

Province	Panel 5	Panel 6
Terre-Neuve	1 231	1 193
Île-du-Prince-Édouard	743	794
Nouvelle-Écosse	1 646	1 533
Nouveau-Brunswick	1 582	1 555
Québec	4 867	5 153
Ontario	7 369	8 241
Manitoba	1 778	2 015
Saskatchewan	1 878	2 039
Alberta	2 685	2 829
Colombie-Britannique	2 321	2 967
Pas disponible ¹	16 230	12 593
Total	42 330	40 912

1. Comprend les personnes qui sont non-répondants dans le champ de l'enquête, qui ont déménagé au Yukon, dans les Territoires du Nord-Ouest ou au Nunavut, qui ont déménagé à l'extérieur du pays, qui sont des personnes institutionnalisées, qui sont décédées, qui ont été éliminées de l'échantillon ou qui ont été inclus par erreur.

Tableau 2.2 Nombre de personnes dans l'échantillon longitudinal, selon la région métropolitaine de recensement et le panel, 2010

Région métropolitaine de recensement	Panel 5	Panel 6
Halifax	502	651
Québec	409	418
Montréal	1 032	1 052
Ottawa - Gatineau	729	940
Toronto	1 337	1 490
Hamilton	340	442
St. Catharines - Niagara	311	407
Kitchener	383	361
London	413	453
Windsor	259	351
Winnipeg	952	1 142
Calgary	564	620
Edmonton	844	716
Vancouver	844	1 078
Victoria	218	426
Autre RMR ou AR	9 343	10 070
N'habite pas dans une RMR	7 620	7 702
Pas disponible ¹	16 230	12 593
Total	42 330	40 912

1. Comprend les personnes qui sont non-répondants dans le champ de l'enquête, qui ont déménagé au Yukon, dans les Territoires du Nord-Ouest ou au Nunavut, qui ont déménagé à l'extérieur du pays, qui sont des personnes institutionnalisées, qui sont décédées, qui ont été éliminées de l'échantillon ou qui ont été inclus par erreur.

L'échantillon transversal est formé des répondants longitudinaux faisant partie du champ de l'enquête et des personnes ne faisant pas partie de l'échantillon longitudinal mais qui

vivent avec un répondant de l'échantillon longitudinal (« cohabitants »). La répartition de l'échantillon transversal par province en 2010 se trouve dans le tableau suivant.

Tableau 2.3 Nombre de personnes dans l'échantillon transversal, selon la province et le panel, 2010

Province	Répondants dans le champ de l'enquête		Cohabitants		Taille de l'échantillon transversal	
	Panel 5	Panel 6	Panel 5	Panel 6	Panel 5	Panel 6
Terre-Neuve	1 231	1 193	243	127	1 474	1 320
Île-du-Prince-Édouard	743	794	159	70	902	864
Nouvelle-Écosse	1 646	1 533	343	188	1 989	1 721
Nouveau-Brunswick	1 582	1 555	360	187	1 942	1 742
Québec	4 867	5 153	1 238	625	6 105	5 778
Ontario	7 369	8 241	1 478	962	8 847	9 203
Manitoba	1 778	2 015	365	253	2 143	2 268
Saskatchewan	1 878	2 039	441	261	2 319	2 300
Alberta	2 685	2 829	700	444	3 385	3 273
Colombie-Britannique	2 321	2 967	458	303	2 779	3 270
Total	26 100	28 319	5 785	3 420	31 885	31 739

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 à chaque année pendant la vie du panel.

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.

À l'EDTR, on utilise l'approche « bootstrap » pour le calcul des erreurs types. Étant donné que l'EDTR utilise un plan d'échantillonnage à plusieurs degrés combiné à une stratégie de calage complexe, il n'y a pas par conséquent de formule simple d'estimation

de variance qu'on puisse appliquer. Il faut donc procéder par approximation. On utilise à cet égard la méthode bootstrap de Rao-Wu, décrite dans leur article de 1987 : Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83, 231-241.

La méthode peut-être décrite de la façon suivante : On prélève indépendamment dans chaque strate un échantillon aléatoire simple avec remise de taille $(n-1)$ parmi les n unités de l'échantillon. À noter que, comme la sélection se fait avec remise, une unité peut être choisie plusieurs fois. On reprend l'opération R fois pour ainsi obtenir R échantillons bootstrap. Pour chacun des R échantillons bootstrap, des poids bootstrap sont calculés pour chaque unité de l'échantillon bootstrap (les unités non sélectionnées dans un échantillon bootstrap donné reçoivent un poids de zéro). Ces poids bootstrap sont formés à partir du poids d'échantillonnage initial, le nombre de fois qu'une unité donnée a été sélectionnée, la taille de l'échantillon initial, ainsi que la taille de l'échantillon bootstrap. Ces poids sont par la suite corrigés selon la même procédure de pondération que les poids courants : ajustement pour tenir compte de la non-réponse, calage, etc. L'ensemble du processus (sélection d'échantillons aléatoires simples, nouveau calcul des poids pour chaque strate) est repris plusieurs fois, ce qui produit R poids bootstrap différents pour chaque unité de l'échantillon original. L'EDTR utilise $R=1,000$ pour produire 1 000 échantillons bootstrap comportant 1 000 poids possibles différents pour chaque unité.

Le résultat final est 1 000 poids bootstrap pour chaque unité de l'échantillon. La variation à l'intérieur des 1 000 estimations possibles fondées sur les 1 000 poids bootstrap est liée à la variance de l'estimateur fondé sur les poids courants et peut être utilisée pour l'estimer.

Le tableau 3.1 illustre les CV provinciaux et nationaux pour certaines variables clés de l'EDTR.

Tableau 3.1 Coefficients de variation nationaux et provinciaux pour certaines variables de l'EDTR, 2010

Variable (pour les familles sauf indication contraire)	T.-N.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Canada
Revenu total médian	3,3	3,6	1,8	2,7	1,3	1,7	2,7	2,0	1,9	2,4	0,8
Revenu médian du marché	4,5	4,2	2,7	3,2	1,9	1,5	2,9	2,2	2,0	2,8	1,0
Médiane des salaires et traitements	5,8	4,0	3,1	3,1	1,8	2,0	3,6	2,9	3,1	3,0	1,0
Médiane des prestations d'assurance-emploi	5,4	6,3	6,5	7,5	6,0	7,2	7,9	14,8	9,5	8,8	3,8
Médiane de l'aide sociale	13,1	15,1	10,8	9,2	6,2	6,7	7,6	33,0	6,9	7,5	4,8
Médiane de la variable « Autre revenu »	15,8	20,9	8,7	16,3	11,4	9,7	17,6	10,5	15,4	7,7	4,9
Nombre de personnes sous le SFR (après impôt)	15,8	19,3	8,8	10,0	6,0	6,3	9,9	10,6	11,2	7,2	3,2
Nombre de personnes avec un emploi ¹	2,1	2,6	1,7	1,6	1,2	1,5	1,8	1,7	2,4	2,0	0,7

1. Comprend les personnes qui ont :
- occupé un emploi toute l'année,
 - occupé un emploi une partie de l'année et été en chômage une partie de l'année,
 - occupé un emploi une partie de l'année et été inactives une partie de l'année, ou
 - été occupé, en chômage et inactives à différents moments de l'année.

Étant donné qu'on retrouve les tailles d'échantillon les plus petites à l'Île-du-Prince-Édouard, les CV les plus élevés ont été observés dans cette province pour toutes les variables à l'étude à l'exception de la médiane des prestations d'assurance-emploi où le Manitoba a enregistré le CV le plus haut.

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 en utilisant des poids avant l'application de ce rajustement pour le glissement. Plus précisément, le glissement se calcule de la façon suivante :

$$\text{glissement}_c = \frac{(CP_c - \sum_{k \in S_c} w_{kc})}{CP_c} * 100$$

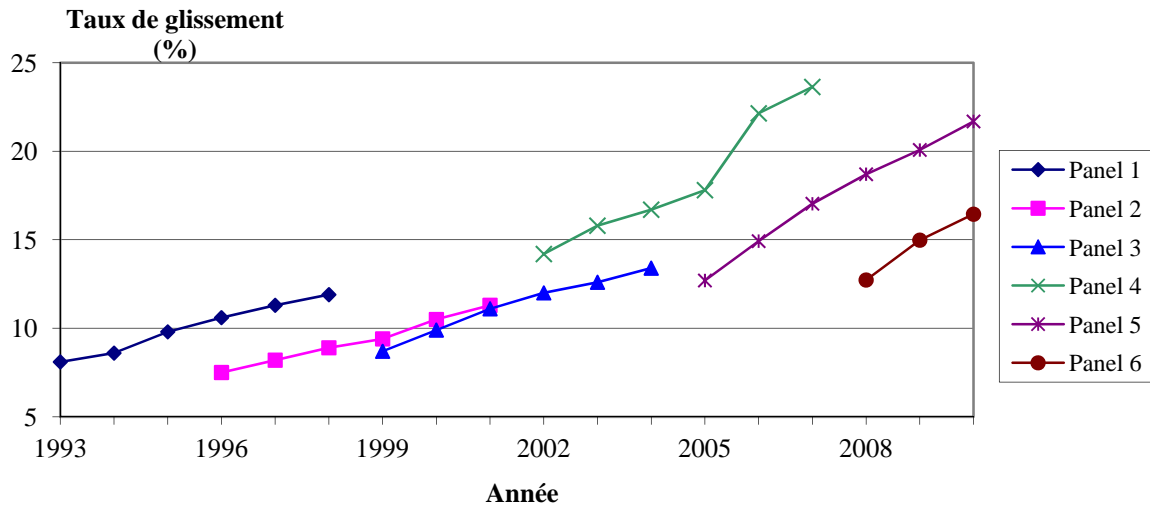
- où la classe C est la classe ou le groupe pour lequel on veut calculer les taux de glissement. Par exemple, les groupes pourraient être basés sur la province, le sexe et/ou le groupe d'âge.
 CP_c est la projection démographique du recensement pour la classe C
 w_{kc} est le poids de sondage avant calage 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.

Les taux de glissement pour les personnes sont montrés par panel et selon l'année de référence dans le graphique 4.1 et par province pour l'année de référence 2010 dans le tableau 4.1.

Graphique 4.1 Taux de glissement pour les personnes, selon l'année de référence et le panel



Le graphique 4.1 montre que les courbes du taux de glissement pour tous les panels suivent sensiblement la même tendance. Ainsi, le taux augmente de manière continue de la première vague jusqu'à la dernière. Cependant, quand on compare les trois premiers panels aux suivants, on note une augmentation plus rapide du taux de glissement pour les panels 4, 5 et 6. Il faut également noter qu'à partir de 2006, on utilise les projections du recensement de 2006 pour calculer le taux de glissement ce qui pourrait expliquer la forte augmentation du taux pour le panel 4 entre 2005 et 2006. En effet, les projections de 2001 ont été utilisées pour calculer le taux à la 4^{ième} vague du panel 4 alors que celles de 2006 ont été utilisées pour le calcul du taux à la 5^{ième} vague.

Les taux de glissement plus élevés pour les personnes et pour les ménages (voir le graphique 4.2) pour le panel 4 sont dus, en partie, à un dénombrement erroné des ménages faisant partie de l'échantillon de l'EDTR qui n'apparaissent pas dans le fichier composé de l'échantillon. Au début d'un panel, on accepte que l'effort requis pour obtenir une réponse de certains ménages soit trop élevé afin de pouvoir envoyer ces ménages à la collecte. Par conséquent, ces ménages sont traités habituellement 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.

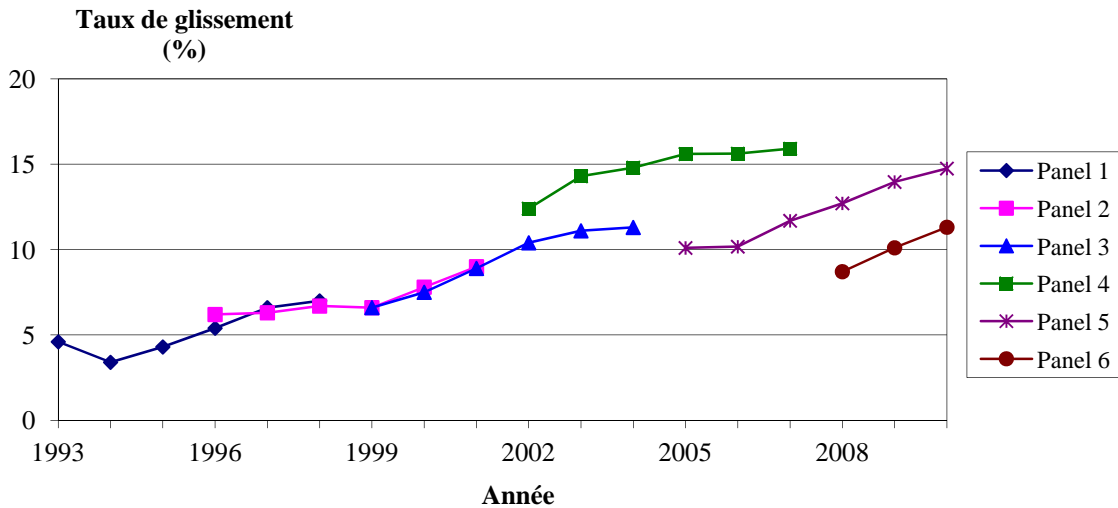
Le tableau 4.1 ci-dessous montre que les taux de glissement pour les personnes les plus élevés ont été enregistrés en Alberta et en Colombie-Britannique.

Tableau 4.1 Taux de glissement pour les personnes, selon la province, 2010

T.-N.-L.	Î.-P.-É.	N.-É.	N.-B.	Qc	Ont.	Man.	Sask.	Alb.	C.-B.	Canada
-1,8	6,0	0,1	7,7	4,9	15,6	4,5	6,0	22,9	24,5	13,5

Le graphique 4.2 montre les taux de glissement pour les ménages selon l'année de référence et le panel tandis que le tableau 4.2 indique les taux selon la province et la taille des ménages.

Graphique 4.2 Taux de glissement pour les ménages, selon l'année de référence et le panel



On constate à partir du graphique 4.2 qu'en général, le taux de glissement augmente d'une vague à l'autre à l'exception d'une vague dans les panels 1 et 2.

Tableau 4.2 Taux de glissement pour les ménages, selon la province et la taille du ménage, 2010

Province	Taille des ménages		
	1	2	Tous
Terre-Neuve	-6,6	-5,8	-2,2
Île-du-Prince-Édouard	13,3	3,6	6,3
Nouvelle-Écosse	-4,5	1,1	-1,4
Nouveau-Brunswick	-2,3	3,5	4,8
Québec	-1,7	1,5	2,0
Ontario	13,5	18,5	14,9
Manitoba	6,9	5,3	4,5
Saskatchewan	7,4	6,5	6,8
Alberta	13,5	24,2	21,5
Colombie-Britannique	19,3	21,7	22,5
Canada	8,4	13,0	11,8

Comme pour les taux de glissement des personnes, les taux de glissement les plus élevés pour les ménages ont été enregistrés en Alberta et en Colombie-Britannique.

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 pour les personnes et pour les 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 ne répondent pas 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 pour lesquelles on dispose 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 la personne ne se trouve plus dans le ménage initial. Puisqu'on n'a pas le moindre renseignement sur la composition du nouveau ménage, ces personnes sont donc considérées comme étant non-répondantes.

Les personnes des ménages non-répondants aux questions sur le travail pour lesquelles on dispose 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 consécutives de janvier sont en conséquence considérées comme des non-répondants de l'EDTR.

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. Le tableau 5.1 montre le statut en 2010 pour les personnes choisies au départ dans l'échantillon longitudinal pour les panels 5 et 6. 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 5.1 Nombre de personnes dans l'échantillon longitudinal, selon le statut et le panel, 2010

Statut des personnes de l'échantillon longitudinal	Panel 5	Panel 6
Dans le champ de l'enquête (répondants)	26 100	28 319
Dans le champ de l'enquête (non-répondants)	4 528	6 326
Déménagé au Yukon, dans les T. N. O., au Nunavut	9	9
Déménagé à l'extérieur du Canada	275	134
Institutionnalisées	649	323
Décédées	1 409	634
Éliminées de l'échantillon ¹	9 344	5 166
Doubles/erreurs ²	16	1
Total	42 330	40 912

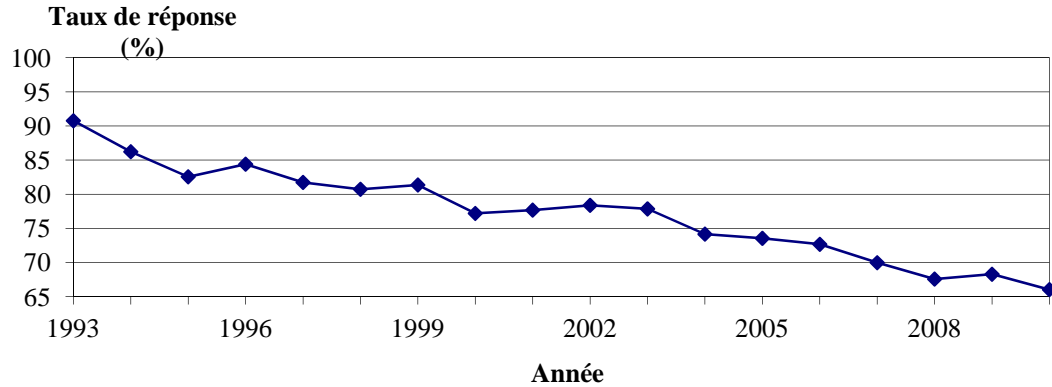
1. Les répondants peuvent être retirés de l'échantillon pour deux raisons. Les ménages dont tous les membres ont refusé de répondre pour 2 cycles consécutifs sont classés « refus catégoriques », et aucun autre effort n'est déployé pour les dénombrer. De même, après deux ans, on ne tente plus de nouveaux 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 5^e et 6^e panels était semblable, on remarque des différences importantes en ce qui concerne le statut des personnes de l'échantillon longitudinal. La proportion de non-répondants dans le champ de l'enquête est nettement plus élevée dans le 6^{ième} panel (près de 50%). Le nombre de personnes dans la catégorie « personnes éliminées de l'échantillon » est beaucoup plus élevé au 5^e panel car on a perdu la trace de plusieurs cas, certains ont été abandonnés et d'autres ont été considérés comme refus catégoriques. Ce résultat était prévisible du fait que le 5^e panel en était à sa sixième et dernière vague en 2010.

Le graphique 5.1 illustre les taux de réponse à l'EDTR des personnes de l'échantillon transversal selon l'année de référence. On calcule les taux de réponse pour les 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 dans 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.

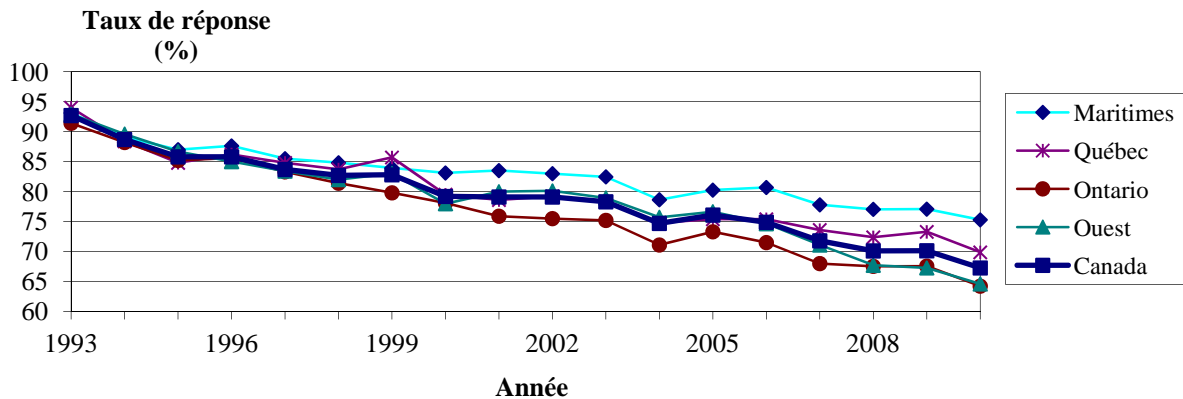
Graphique 5.1 Taux de réponse transversal pour les personnes, selon l'année de référence



Le graphique 5.1 fait voir clairement que le taux de réponse des personnes de l'échantillon transversal a baissé depuis le début de l'enquête, à l'exception d'une faible hausse en 1996, 1999, 2002 et 2009. Le taux a débuté à 90,8 % en 1993 pour atteindre un creux à 66,0 % en 2010.

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.

Graphique 5.2 Taux de réponse transversal pour les ménages, selon l'année de référence et la région



Le graphique ci-dessus illustre la tendance à la baisse du taux de réponse au fil des ans. On peut observer une baisse plus importante pour l'année 2004. Par la suite, le taux remonte en 2005 pour ensuite redescendre jusqu'à son niveau plancher de 67,3 % en 2010, pour le Canada. De façon générale, la courbe pour les Maritimes montre les taux les plus élevés tandis que celle de l'Ontario montre les taux les plus faibles.

Le tableau 5.2 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 5.2 Taux de réponse transversal pour les personnes, selon l'année de référence¹ et la phase

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
2008	56,4	7,2	3,9	32,5
2009	55,9	6,3	6,0	31,9
2010	53,7	5,7	6,2	34,4

1. Depuis 2004, les questions sur le travail et sur le revenu sont posées lors d'une seule entrevue qui a lieu en janvier.

En lien avec le graphique 5.1, on remarque une baisse générale du taux au tableau 5.1 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 à l'étude (53,7 %). Également, à deux exceptions près (2005 et 2009), la proportion de non-répondants était à son plus bas à ses débuts en 1993 (7,9 %) pour ensuite augmenter graduellement et atteindre son apogée en 2010 (34,4 %). 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 3,9 % et 6,0 % exception faite de l'année 2005 où le taux est descendu à 2,9 %.

Étant donné la difficulté de définir conceptuellement un ménage « longitudinal », seuls les taux de réponse longitudinaux pour les personnes sont calculés. Le tableau 5.2 montre ces derniers pour chacun des six 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 longitudinal pour les personnes, selon le panel et la vague

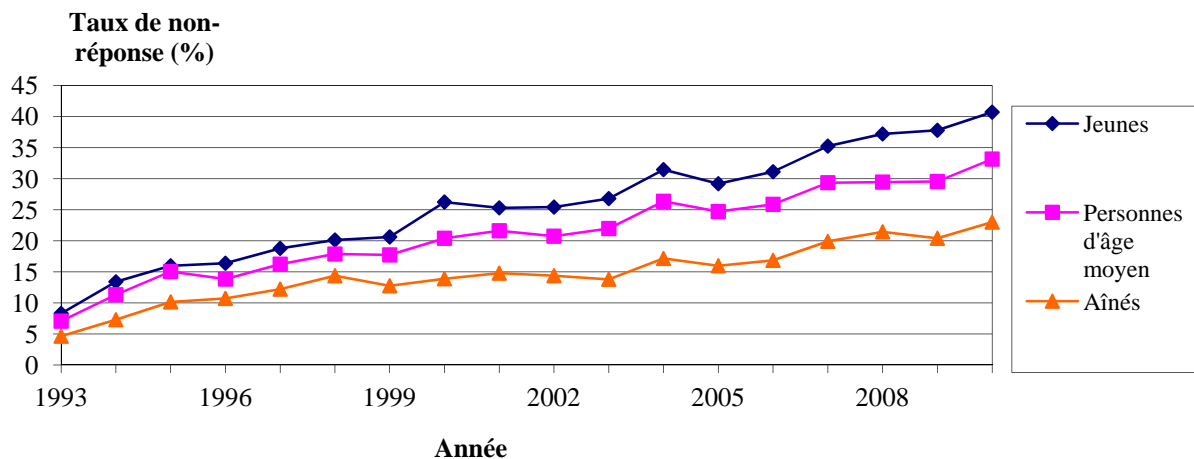
Panel (année de début)	Vague du panel					
	1	2	3	4	5	6
Panel 1 (1993)	93,3	89,6	86,5	83,9	82,6	81,5
Panel 2 (1996)	89,5	86,8	85,2	82,7	78,5	77,4
Panel 3 (1999)	83,9	83,0	83,0	79,6	76,4	73,7
Panel 4 (2002)	81,2	83,2	78,3	75,0	71,6	68,9
Panel 5 (2005)	78,8	80,6	77,3	72,8	69,3	67,2
Panel 6 (2008)	71,0	75,6	71,9

... Ne s'applique pas.

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 78,8 % à la vague 1 à 67,2 % à la vague 6 pour le 5^e panel. Les taux sont encore plus bas pour les premières vagues du panel 6.

Le graphique 5.3 illustre les taux de non-réponse longitudinaux 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 moins 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 pour les personnes, selon l'année de référence et 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 non-réponse. Si on compare les taux entre 1993 et

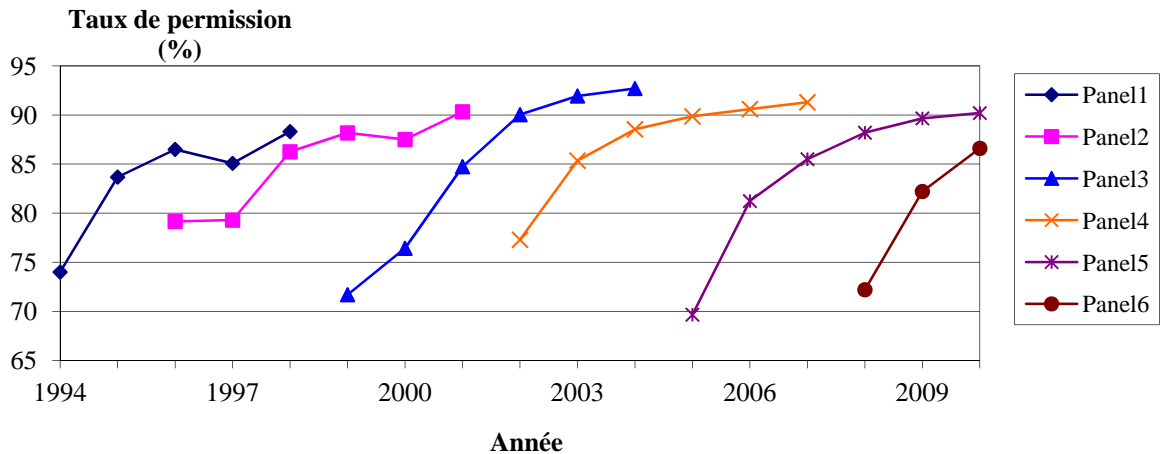
2009, on peut voir que ces derniers ont quadruplé au cours des 18 ans de l'EDTR. Les jeunes, soit ceux âgés entre 16 et 29 ans montrent des taux de réponse environ deux fois plus élevés que les aînés qui sont âgés de plus de 60 ans. En effet, en 2010, 40,7% des jeunes n'ont pas répondu à l'enquête comparativement à 23,0% 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 à la maison.

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 portait sur 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 selon l'année de référence et le panel. 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.

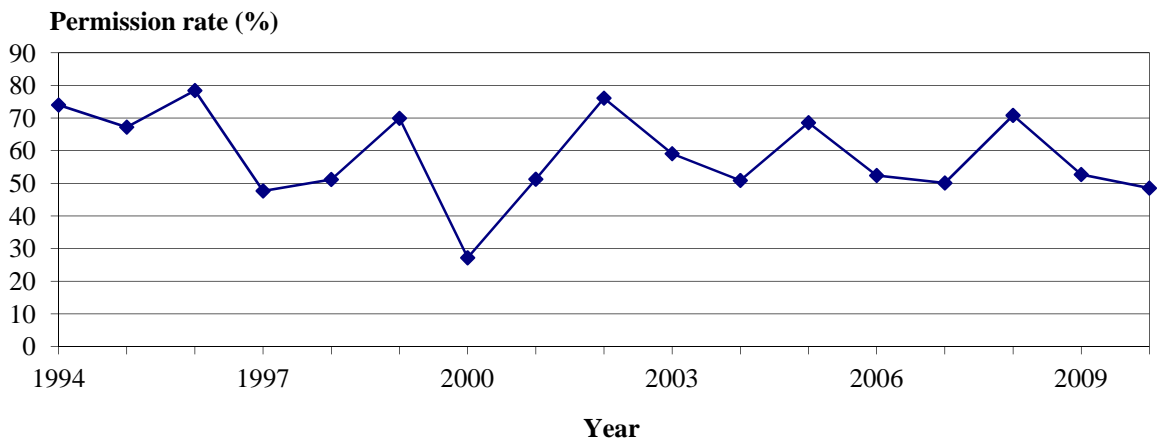
Graphique 6.1 Taux de permission, selon l'année de référence et le panel



On remarque des tendances similaires dans les 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 vagues. 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 les taux de permission annuels des nouveaux répondants qui ont autorisé qu'on consulte leur 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, 2005 et 2008) a toujours enregistré les plus haut taux de permission chez les nouveaux répondants.

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

La première étape du processus consiste à standardiser les variables d'appariement sur les fichiers de données fiscales et de l'EDTR afin d'assurer la compatibilité du formatage de toutes les variables d'appariement. Ce processus de standardisation implique, entre autres, qu'on enlève les caractères manquants du champ d'adresse et qu'on utilise le codage phonétique tel qu'utilisé dans les systèmes NYSIIS et SNDX¹. Les variables standardisées disponibles pour effectuer l'appariement sont donc : l'adresse, la ville, la date de naissance, le prénom, le nom de famille, le sexe, la province, les codes NYSIIS et SNDX pour le nom de famille, le code postal, l'état matrimonial, le numéro de téléphone et la première initiale.

Un programme SAS mis au point à Statistique Canada sert à comparer les données provenant des deux sources (fiscales et EDTR). Afin de gérer efficacement le processus d'appariement, les enregistrements provenant des deux sources sont partitionnés selon la date de naissance, le code postal et le code SNDX pour le nom de famille. À l'intérieur d'une partition donnée, tous les enregistrements de l'EDTR sont comparés à ceux provenant des fichiers de données fiscales. On attribue des facteurs d'importance pour la concordance basés sur la probabilité qu'une paire d'enregistrements (un de l'EDTR et un du dossier fiscal) représente la même personne. 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.

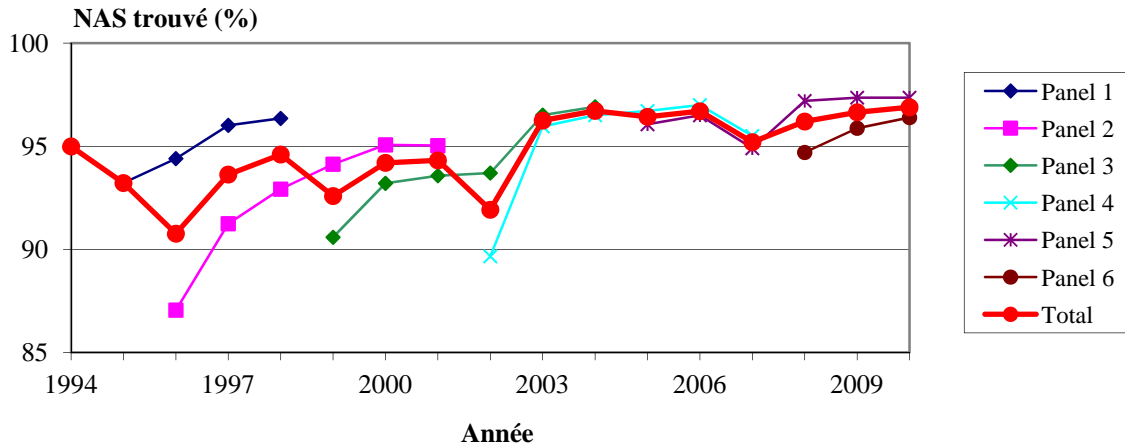
Pour chaque cycle de traitement, on entreprend deux fois plutôt qu'une le processus d'appariement de l'EDTR : une première fois en utilisant le fichier final de données fiscales de l'année de référence précédente et une autre fois en utilisant le fichier préliminaire de données fiscales de l'année de référence en cours. Pour les appariements potentiels, le résultat de l'appariement peut donner lieu à neuf possibilités de résultats différents. Ces possibilités dépendent du fait qu'un appariement puisse être considéré comme étant définitif ou questionnable. À la fin du processus, les appariements potentiels sont soit acceptés ou soit révisés manuellement. Les doublons liés au fait que deux unités de l'EDTR puissent être attachées au même NAS sont résolus à la toute fin du processus d'appariement.

À partir des NAS qu'on vient d'obtenir, on extrait les informations fiscales pour les répondants qui ont donné la permission d'accéder à leurs dossiers fiscaux. Le graphique 7.1 illustre le pourcentage des répondants de l'échantillon de l'EDTR qui

1 .NYSIIS et SNDX sont des routines de codage du nom qui servent à éliminer les fautes d'orthographe fréquentes dans les noms de famille des répondants. Ce codage s'effectue en fonction de la prononciation du nom de famille.

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.

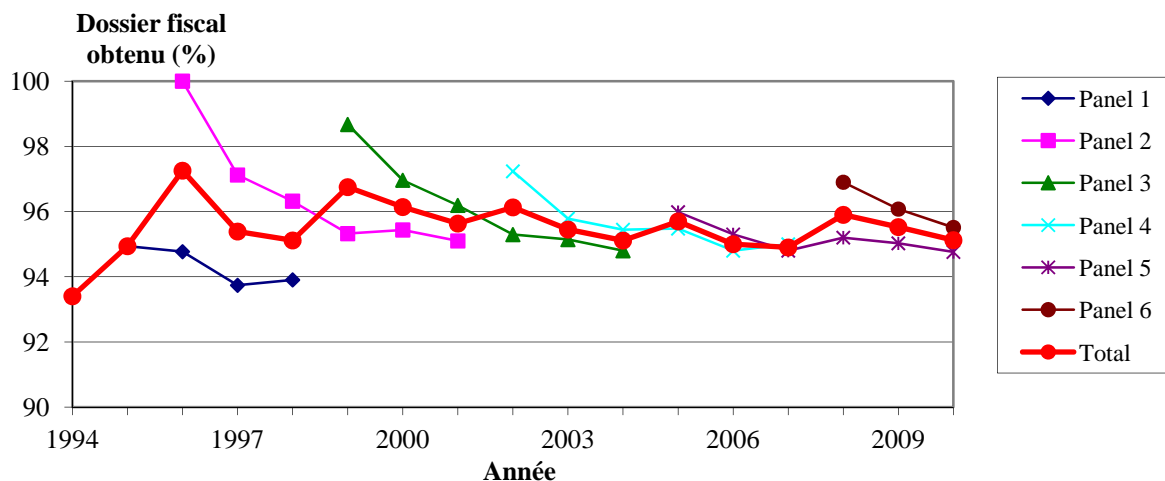
Graphique 7.1 Pourcentage de personnes ayant donné leur permission et pour lesquels le NAS a été trouvé, selon l'année de référence et le panel



En général, la proportion de répondants ayant donné leur permission et pour lesquels le NAS a été trouvé a montré une tendance à la hausse lors des six vagues de chacun des panels. On a observé une hausse plus importante entre la 1^{re} et la 2^e vague de chaque panel 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 montre les taux de couplage pour les personnes de l'échantillon de l'EDTR qui ont donné la permission et pour lesquelles un NAS a été trouvé.

Graphique 7.2 Taux de couplage lorsque le NAS a été trouvé, selon l'année de référence et le panel



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, 2005 et 2008).

Toutefois, on remarque que les taux de couplage débutent à un seuil de plus en plus bas au fil des panels 2 à 5 (il a remonté quelque peu pour le panel 6). 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. Il est remonté quelque peu à 96,9 % pour le panel 6.

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 Pourcentage de répondants venant du dossier fiscal ou de l'entrevue, selon l'année de référence¹

Année	Dossier fiscal	Entrevue	Autres ²
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
2008	74,4	8,9	16,7
2009	79,3	5,8	14,9
2010	81,5	4,4	14,1

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

2. 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 proviennent des dossiers fiscaux. Les proportions varient entre 71,9 % en 1999 et 83,4 % en 2004. En fait, ce pourcentage fluctue autour de 80 % sauf pour les années où débutait un nouveau panel soit en 1999, 2002, 2005 et en 2008 où il était plus bas. Par le fait même, c'est aussi lors de ces mêmes années qu'on observe une proportion plus élevée de réponses provenant des entrevues (entre 9 % et 12 %) alors que le pourcentage était plus autour de 5 % pour les autres années.

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 les 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 à certaines d'entre elles. On peut considérer que les taux d'imputation dans l'EDTR sont une mesure de la réponse partielle à l'enquête.

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 en cours. La méthode transversale utilise l'approche du plus proche voisin. Quelques variables sont aussi imputées de façon déterministe.

Pour l'imputation par le donneur, on définit un ensemble de règles de cohérence de base et un ensemble de donneurs pour un enregistrement donné nécessitant l'imputation. 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.

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, jusqu'à dix-huit variables du revenu sont imputées pendant l'imputation du revenu pour 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 la personne.

Tableau 8.1 Taux d'imputation des variables de revenu pour les ménages, selon la province et le type d'imputation, 2010

Province	Imputation totale ¹	Imputation partielle ²	Pas d'imputation
Terre-Neuve	1,3	18,7	80,0
Île-du-Prince-Édouard	2,3	20,1	77,7
Nouvelle-Écosse	1,8	18,7	79,5
Nouveau-Brunswick	2,5	19,1	78,4
Québec	2,1	17,2	80,7
Ontario	3,1	23,9	73,0
Manitoba	2,6	19,8	77,6
Saskatchewan	1,8	17,8	80,4
Alberta	3,3	23,3	73,3
Colombie-Britannique	2,8	23,3	73,9
Canada	2,6	20,9	76,5

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 montre que les 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 80,7 % des enregistrements de la province n'ont pas été imputés. Par contre, c'est en Ontario où on a dû avoir le plus recours à l'imputation. Le taux d'imputation partielle a franchi la barre des 24 % pour la province.

Peu d'enregistrements ont nécessité de l'imputation totale, alors que les taux ont varié entre 1,8 % et 3,3 % dans les différentes provinces canadiennes.

Au tableau 8.2, on compare 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 incohérences lors du processus d'imputation. Ce tableau fournit également un aperçu de l'importance de l'imputation partielle (1 variable, 2 à 9 variables et 10 à 17 variables).

Tableau 8.2 Taux d'imputation des variables de revenu pour les ménages, selon le type d'imputation et la source des données, 2010

Imputation	Source des données			Tous
	Dossiers fiscaux	Entrevue	Autres ¹	
Partielle (1 variable)	8,4	12,0	0,0	7,4
Partielle (2 à 9 variables)	0,4	37,8	0,0	2,0
Partielle (10 à 17 variables)	0,0	0,2	...	11,5
Totale	100,0	2,6
Pas d'imputation	91,2	50,1	...	76,5
Total	100,0	100,0	100,0	100,0

... Ne s'applique pas

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, 91,2 % des enregistrements n'ont pas eu besoin d'imputation tandis que 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 raison de la non-réponse à certaines questions, il a fallu commencer à imputer des variables reliées au logement. Deux méthodes d'imputation ont été utilisées : l'imputation longitudinale et l'imputation transversale par donneur. L'imputation transversale par 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 pour les variables relatives à la condition du logement.

Tableau 8.3 Taux d'imputation des variables de logement pour les ménages, selon la province et le type d'imputation, 2010¹

Province	Imputation totale ²	Imputation partielle ³	Pas d'imputation
Terre-Neuve	...	38,4	61,6
Île du Prince Édouard	...	40,8	59,2
Nouvelle-Écosse	...	33,6	66,4
Nouveau-Brunswick	...	35,0	65,0
Québec	...	33,8	66,2
Ontario	...	43,3	56,7
Manitoba	...	38,7	61,3
Saskatchewan	...	39,6	60,4
Alberta	...	42,6	57,4
Colombie-Britannique	...	43,7	56,3
Canada	...	39,6	60,4

1. Pour l'année de référence 2010, on n'a pas tenu compte de la variable indiquant si le logement est un condominium étant donné le changement dans la façon dont cette variable est définie. Si cette variable avait été conservée, le taux d'imputation partielle aurait été augmenté de façon très importante.
2. Pas de renseignements fournis par le répondant. Toutes les données ont été imputées.
3. Un ou plusieurs éléments sont imputés à l'aide de certains renseignements fournis par le répondant.

Pour le Canada, 39,6 % des ménages ont eu besoin d'imputation partielle pour les variables portant sur le logement. Le taux d'imputation le plus élevé a été observé en Colombie-Britannique. Le taux s'approche même de la barre des 44 % pour cette province. Le taux le plus bas, quant à lui, a été enregistré en Nouvelle-Écosse à 33,6 %.

Au total, jusqu'à dix-huit variables reliées au ménage ont été imputées dans le cadre de l'imputation des variables relatives au logement. Une grande proportion de ménages ne nécessite qu'une imputation partielle. Le tableau 8.4 montre la répartition de ceux qui nécessitent une imputation partielle.

Tableau 8.4 Taux d'imputation des variables de logement pour les ménages, selon l'année de référence et le nombre de variables demandant l'imputation

Année	Nombre de variables ménages ayant besoin d'imputation			
	1	2 à 5	6 à 17 ¹	Un ou plus
2004	10,5	10,2	10,6	31,3
2005	10,2	10,6	15,2	36,0
2006	10,0	7,4	22,3	39,7
2007	9,8	6,9	22,3	39,0
2008	8,8	5,9	28,4	43,1
2009	8,8	6,1	27,4	42,2
2010	8,6	5,9	25,1	39,6

1. À partir de l'année de référence 2006, les valeurs des deux variables indiquant le type de combustible à chauffage utilisé dans le logement ont été mises à S/O pour tous les ménages. Résultat, seulement 18 variables au maximum sont imputées

Le nombre de variables à imputer a augmenté annuellement entre 2004 et 2008 avant de se mettre à diminuer systématiquement à partir de 2008.

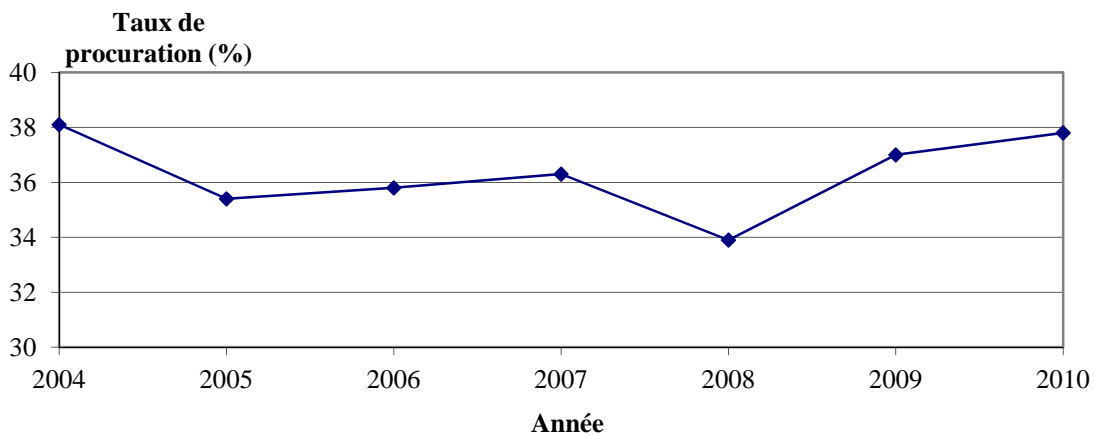
9. Taux d'entrevues par procuration

Une entrevue réalisée par procuration signifie que l'intervieweur a obtenu l'information sur une certaine personne du ménage via un autre membre du ménage prêt à répondre pour cette dernière. L'information sur le nombre d'entrevues s'étant déroulées par procuration est disponible depuis l'année de référence 2000. Une variable spécifique est utilisée pour indiquer si l'entrevue a été réalisée par procuration. Avant l'année de référence 2004, les répondants étaient questionnés deux fois par année, une fois en janvier et une fois en mai, créant ainsi deux variables de procuration. Depuis l'année de référence 2004, l'entrevue du mois de mai a été éliminée; entraînant ainsi la création d'une seule variable de procuration pour l'entrevue du mois de janvier. Afin de pouvoir comparer les résultats d'une année à l'autre, seuls les taux de procuration depuis l'année de référence 2004 seront présentés dans ce chapitre.

Les taux de procuration sont calculés en fonction des répondants de l'EDTR âgés de 16 ans et plus qui ont fait parvenir leurs réponses lors de l'entrevue, eux-mêmes ou via une personne interposée pour l'année de référence (nombre total de répondants). Les taux de procuration sont obtenus en divisant le nombre de répondants par procuration par le nombre total de répondants.

Le graphique 9.1 ci-dessous illustre les taux de procuration depuis l'année 2004.

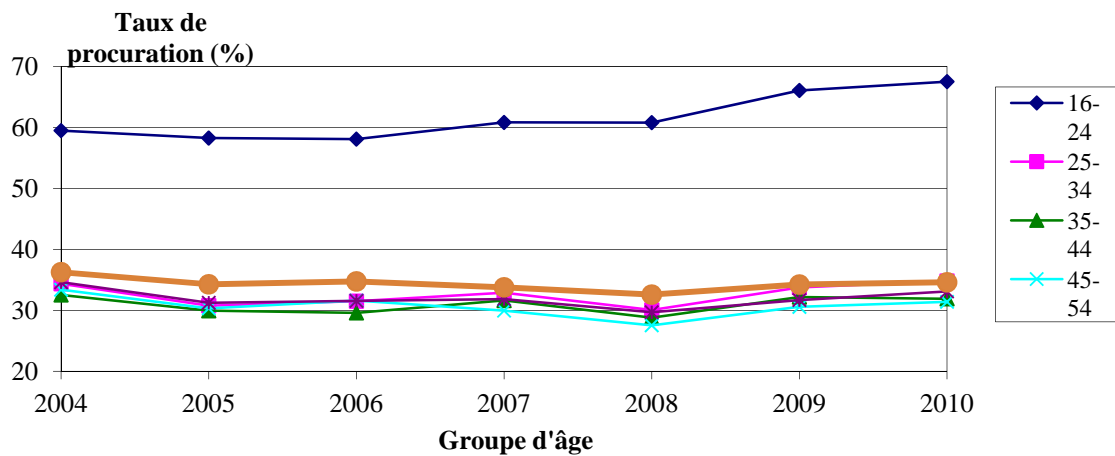
Graphique 9.1 Taux de procuration selon l'année de référence



Le graphique ci-dessus illustre que le taux de procuration a baissé de 2004 à 2005, passant de 38,1 % à 35,4 %. Par la suite, la courbe des taux a montré une tendance à la hausse jusqu'en 2009, à l'exception de l'année 2008 où le taux a atteint un creux pour la période à l'étude soit 33,9 %.

Le graphique 9.2 ci-dessous montre le taux de procuration selon le groupe d'âge pour les années de référence 2004 à 2010. En général, des tendances semblables sont observées selon les différents groupes d'âge pour chacune des années à l'étude. Peu importe l'année, le taux de procuration est très élevé chez les 16 à 24 ans, environ 60 % avant 2008 à 68% pour l'année 2010, alors que les autres groupes d'âge montrent un taux moyen oscillant entre 31 % et 34 % pour la période à l'étude. Cette différence peut être expliquée par le fait que les jeunes sont plus difficiles à rejoindre que le reste de la population et que la plupart d'entre eux vivent toujours avec leurs parents, et donc, les parents auront tendance à répondre pour eux. En effet, en 2010, l'échantillon contenait 11 % de jeunes de 16 à 24 ans, avec un taux de réponse longitudinal de 63 % comparativement à 69 % pour le reste de la population âgée de 25 ans et plus.

Graphique 9.2 Taux de procuration, selon l'année de référence et le groupe d'âge



10. Arrondissement des données sur le revenu

Un faible pourcentage de données sur le revenu de 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 quatre 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 10.1 indique la distribution de chacun de ces chiffres pour toutes les valeurs déclarées supérieures ou

égales à 10 000 \$ pour la variable « salaires et traitements », à la fois pour les données recueillies (lors de l'entrevue) et pour les données fiscales.

Tableau 10.1 Proportion des enregistrements, selon la valeur des quatre derniers chiffres de la variable salaires et traitements¹ et la provenance des données, 2010

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	36,7	11,7	89,3	11,9	94,5	13,1	96,6	14,5
1	3,7	10,5	0,4	9,6	0,4	9,3	0,3	9,2
2	8,8	10,6	1,0	9,9	0,6	9,8	0,5	9,5
3	5,7	9,7	1,3	9,5	0,5	10,1	0,3	9,6
4	4,9	10,2	1,3	9,9	0,4	9,8	0,3	10,1
5	20,1	9,9	3,0	9,6	0,7	9,9	0,8	10,0
6	6,2	9,5	1,5	10,2	0,7	9,7	0,3	9,0
7	4,1	9,4	0,5	9,4	0,4	9,5	0,2	9,2
8	6,5	9,5	1,0	10,0	0,9	9,3	0,5	9,3
9	3,3	9,0	0,6	9,9	0,7	9,5	0,3	9,7

¹ Seulement pour les cas où la valeur de la variable était supérieure à 9 999 \$.

Le tableau 10.1 illustre clairement que les salaires et traitements de 10 000 \$ ou plus ont été arrondis. Les troisième, deuxième et dernier chiffres de la fin ont été un zéro dans 89,3 %, 94,5 % et 96,6 % 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, le tiers des enregistrements ont montré un zéro et 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.

Les données recueillies lors de la collecte ont été étudiées plus en détails afin de vérifier s'il y avait une différence significative entre les données obtenues directement par le répondant et celles acquises par procuration. Un premier résultat indique que, pour la variable salaires et traitements, le 3^e chiffre de la fin était un zéro dans 89,1 % des cas par procuration et 89,4 % lorsque la personne répondait elle-même. Dans le cas du dernier chiffre, c'était un zéro dans 97,3 % des cas par procuration et 96,1 % lorsque la personne répondait elle-même. Des résultats semblables ont été observés pour les autres chiffres considérés. Par conséquent, on conclue que les réponses obtenues directement par le répondant ou par procuration n'indiquent aucune différence significative dans l'arrondissement pour la variable salaires et traitements.

Le tableau 10.2 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 10.2 Proportion des enregistrements ayant un zéro, pour les quatre derniers chiffres des données déclarées pour certaines variables, 2010

Variable	Chiffre			
	Quatrième de la fin	Troisième de la fin	Avant-dernier	Dernier
Salaires et traitements	36,7	89,3	94,5	96,6
Revenu de dividendes	17,6	35,1	76,7	86,7
Prestations d'assurance-emploi	7,5	53,3	83,9	93,5
Revenu de travail autonome non agricole	25,0	80,4	95,7	95,8

Ces derniers résultats illustrent, en général, l'augmentation graduelle de la proportion de zéros quand on observe l'emplacement du chiffre de la fin en passant du 4^e chiffre de la fin jusqu'au dernier. Pour les salaires et traitements et le revenu de travail autonome non agricole, on observe une plus grande proportion de zéros que les autres variables pour les 4^e, 3^e et 2^e chiffres avant la fin.

Pour les revenus de dividendes et les prestations d'assurance-emploi, on remarque une forte hausse de la proportion de zéros en comparant le 3^e chiffre avant la fin avec le 2^e chiffre. Ces augmentations varient de 41,6 % à 30,6 %.

Enfin, toutes les variables ont la valeur zéro au dernier chiffre dans plus de 90 % des cas, sauf pour les revenus de dividendes (86,7 %) où la proportion de zéros est quand même élevée.