

# PONDÉRATION DE L'ENQUÊTE SUR LA DYNAMIQUE DU TRAVAIL ET DU REVENU

Caroline Cauchon et Sylvie LaRoche<sup>1</sup>

## RÉSUMÉ

L'Enquête sur la dynamique du travail et du revenu (EDTR) est une enquête longitudinale par panel menée auprès de la population canadienne. Elle est la principale source d'information sur le revenu des particuliers, des familles et des ménages, incluant l'incidence des familles à faible revenu. L'échantillon de l'EDTR est composé de deux panels qui se chevauchent. À chaque année de référence, l'EDTR produit deux ensembles de poids : un ensemble de poids représentatif de la population initiale (l'aspect longitudinal) et un autre qui est représentatif de la population actuelle (l'aspect transversal). Ce document décrira la méthodologie employée à l'EDTR pour pondérer les échantillons longitudinaux et transversaux.

MOTS CLÉS : Calage aux marges; méthode de partage des poids; non-réponse; pondération longitudinale; pondération transversale.

## ABSTRACT

The Survey of Labour and Income Dynamics (SLID) is a longitudinal survey conducted on individuals. It is the main source for income data for Canadian families, households and individuals, including the family low income rates. The SLID sample is composed of two overlapping panels. Each reference year, SLID produces two sets of weights: one set of weights representing the initial population (longitudinal weights) and another set of weights representing the actual population (cross-sectional weights). This document will describe the methodology used in SLID to weight the longitudinal and the cross-sectional samples.

KEY WORDS: Calibration, Cross-sectional weighting strategy, Longitudinal weighting strategy; Non-response, Weight sharing method.

## 1. INTRODUCTION

L'Enquête sur la dynamique du travail et du revenu (EDTR) est une des premières enquêtes longitudinales développées à Statistique Canada. Le but de cette enquête est de mesurer les changements au niveau du bien-être économique des individus et d'étudier les facteurs qui influencent ces changements. L'EDTR mesure aussi l'impact sur les individus des changements dans la situation familiale. Outre son aspect longitudinal, elle est également la principale source de données transversales sur le revenu des Canadiens.

### 1.1 Description de l'échantillon

L'échantillon de l'EDTR est sélectionné à partir de l'échantillon de l'Enquête sur la population active (EPA). Les deux échantillons partagent donc le même plan de sondage. L'échantillon de l'EPA est prélevé à partir d'une base aréolaire et se fonde sur un plan stratifié à plusieurs degrés s'appuyant sur un échantillonnage probabiliste. La population cible est celle des dix provinces canadiennes, à l'exception des personnes vivant dans les réserves indiennes, les baraques militaires et les institutions. L'unité d'échantillonnage au dernier degré est le logement. Tous les membres des ménages habitant les logements choisis lors de la sélection de l'échantillon sont donc inclus dans l'échantillon de l'EPA. L'échantillon total

---

<sup>1</sup> Caroline Cauchon et Sylvie LaRoche, Division des méthodes d'enquêtes auprès des ménages, Statistique Canada, Édifice R.H.Coats, 18ième étage, Ottawa (Ontario), Canada, K1A 0T6, [caroline.cauchon@statcan.ca](mailto:caroline.cauchon@statcan.ca), [sylvie.laroche@statcan.ca](mailto:sylvie.laroche@statcan.ca)

(mensuel) est composé de six échantillons indépendants (groupes de rotation). Chaque groupe de rotation est représentatif de la population canadienne et un sixième de l'échantillon est remplacé tous les mois (Gambino, 1998).

L'EDTR pour sa part, est une enquête longitudinale par panel rotatif. Chaque panel est composé de deux groupes de rotation de l'EPA. Tous les ménages répondants lors du sixième et dernier contact de l'EPA (appelés ménages sortant de l'EPA) sont inclus dans l'échantillon de l'EDTR. Ainsi, un panel est composé d'environ 17 000 ménages, soit approximativement 35000 personnes. Chaque panel a une durée de vie de 6 ans, le panel le plus âgé étant remplacé à tous les 3 ans. Comme le deuxième panel a été introduit en 1996, il y a toujours deux panels qui se chevauchent depuis ce temps. Les panels peuvent être utilisés séparément pour fournir des estimations longitudinales ou encore combinés pour fournir des estimations transversales. Les panels peuvent aussi être combinés longitudinalement pour la période de chevauchement, soit un maximum de 3 ans (Naud, 2004).

Tous les membres des ménages sélectionnés sont suivis pour la durée du panel et sont contactés 2 fois par année pour recueillir de l'information sur la composition et le revenu du ménage, ainsi que sur l'activité des individus sur le travail. On recueille aussi ces renseignements pour les individus qui se sont ajoutés au ménage depuis la sélection du panel afin de maintenir une représentativité transversale. Ces nouveaux membres, appelés cohabitants, ne seront pas pris en compte dans la pondération longitudinale mais feront partie du processus de pondération transversale.

Les prochaines sections décrivent la méthodologie utilisée lors de la pondération longitudinale et transversale de l'EDTR. Le tableau 1 situé à la fin du document présente une liste des étapes effectuées pour les deux types de pondération. Afin d'illustrer le processus de pondération longitudinal et transversal, l'année 2003 sera utilisée comme période de référence dans la suite du document.

## **2. PONDÉRATION LONGITUDINALE**

Les poids longitudinaux d'un panel donné doivent représenter la population initiale, c'est-à-dire la population cible au moment de la sélection du panel. Par exemple, le panel 3 représente la population cible en date du 31 décembre 1998 alors que le panel 4 représente la population cible en date du 31 décembre 2001. À la sélection d'un nouveau panel, appelé vague 0, l'information disponible pour tous les ménages répondants et non-répondants est celle provenant de l'EPA. Celle-ci sera utilisée pour deux des étapes de la pondération soit l'ajustement de la non-réponse et le calage aux marges.

Chaque personne longitudinale se voit d'abord attribuer un poids initial. Ce poids est équivalent au sous-poids de l'EPA ajusté pour la non-réponse et multiplié par 3 car le panel n'est constitué que de 2 des 6 groupes de rotations de l'EPA et que chacun des groupes de rotation représente un sixième de la population canadienne. Une personne longitudinale est considérée comme répondante si elle a répondu à l'entrevue sur le travail ou le revenu pour la vague courante. Au niveau du ménage, celui-ci est considéré répondant si au moins un de ses membres est répondant. Les données des personnes non-répondantes à l'intérieur d'un ménage répondant seront imputées.

La pondération longitudinale de chaque panel est faite séparément et compte 3 étapes : l'ajustement de non-réponse, l'ajustement pour les valeurs influentes et finalement le calage servant à améliorer l'efficacité de l'échantillon et sa représentativité pour des caractéristiques-clés de l'enquête. La première étape de la pondération longitudinale est l'ajustement pour la non-réponse. La méthode utilisée à l'EDTR pour construire les groupes de réponse homogènes (GRH) est l'analyse par segmentation (Kass, 1980; Angoss Software, 1995). Les GRH sont créés à partir des variables explicatives disponibles à la vague 0. Par exemple, le niveau d'éducation, les groupes d'âge, le sexe et le fait d'être propriétaire ou locataire sont des variables issues de l'EPA qui sont souvent significatives dans la composition des GRH. La méthode par segmentation construit un arbre de décision pour la composition des GRH, c'est-à-dire le choix des variables explicatives et chaque point de branchement est étudié de façon indépendante ce qui ne force pas l'arbre à être symétrique. Une fois les GRH construits, on calcule l'ajustement basé sur le taux de réponse à l'intérieur de ces groupes. L'ajustement de non-réponse sera appliqué au poids de tous les répondants. Pour les enfants, l'ajustement appliqué sera en fait la moyenne des ajustements appliqués aux adultes du ménage. L'arbre de décision ainsi que l'ajustement de non réponse sont refaits à chaque vague car l'échantillon des répondants change d'une vague à l'autre. Le modèle est donc recréé à chaque année de référence sans toutefois avoir d'impact sur le poids longitudinal des années précédentes.

Le deuxième ajustement consiste à ajuster les poids pour les valeurs influentes. Pour savoir si on doit ajuster le poids d'une personne donnée, on évalue d'abord la contribution de cette personne à l'estimation transversale pondérée du revenu total d'une province. Les valeurs influentes sont donc déterminées de façon transversale et peuvent changer d'une vague à l'autre. Comme cette étape est réalisée dans le cadre de la pondération transversale, elle sera discutée plus longuement dans la section 3. Il existe cependant une différence entre l'aspect longitudinal et transversal en ce qui a trait au choix du facteur d'ajustement. En effet, afin de garantir une certaine cohérence dans les données longitudinales, le facteur d'ajustement pour l'aspect longitudinal est égal au minimum des facteurs calculés pour les estimations transversales, à partir de la première vague du panel jusqu'à la vague courante. Pour les personnes jugées non influentes, ce facteur est égal à un. Les poids ajustés pour les valeurs influentes s'obtient donc en multipliant les poids longitudinaux ajustés pour la non-réponse par ce facteur d'ajustement pour valeurs influentes.

L'étape finale de la pondération longitudinale est le calage aux marges. Cette étape consiste à ajuster les poids intermédiaires obtenus après les deux ajustements précédents de façon à ce que les estimations tirées de l'échantillon soient égales aux totaux connus de la population pour certaines caractéristiques-clés de l'enquête, tout en respectant certaines conditions sur la valeur des poids à la fin du processus. Les classes de calage sont donc définies en fonction des sous-groupes de l'échantillon que l'on veut ajuster lors du calage. La seule restriction imposée sur les poids finaux de l'EDTR est que ceux-ci doivent être égaux ou supérieurs à 1 à la fin du calage. Le calage se fait au moyen d'une régression généralisée (LaRoche, 2005) et est effectué indépendamment pour les deux panels.

Il y a deux sortes de totaux de contrôle utilisés à l'EDTR; les totaux démographiques au niveau provincial soient le nombre de personnes par groupe d'âge et de sexe, le nombre de ménages<sup>2</sup> et d'entités économiques<sup>3</sup> selon la taille et les totaux basés sur les revenus, soit le nombre de salariés par classe de salaires et traitements. Les projections démographiques utilisées sont celles dérivées à partir du recensement effectué en 2001 alors que la distribution du nombre de salariés par classe de salaires et traitements est obtenue à partir du fichier de l'état de la rémunération payée par l'employeur tel que recueilli par l'Agence du revenu (fichier T4). Il y a 7 classes de salaire utilisées. Les bornes de ces classes varient d'une province à l'autre car elles sont choisies selon des centiles spécifiques (10<sup>e</sup>, 25<sup>e</sup>, 50<sup>e</sup>, 65<sup>e</sup>, 75<sup>e</sup> et 95<sup>e</sup>). Dans le cas des grandes provinces, la borne de la dernière classe est remplacée par le 98<sup>e</sup> ou 99<sup>e</sup> centile si le nombre de personnes dans la dernière classe est suffisant. Le nombre de salariés dans chaque classe de salaires et traitements est calculé à partir du T4 pour chaque province et on obtient donc, de cette façon, les totaux de contrôle souhaités.

Comme l'échantillon longitudinal doit représenter la population cible au moment de la sélection du panel, le classement des individus dans les différentes classes de calage se fait en utilisant les caractéristiques des individus longitudinaux à la vague 0. L'information disponible pour tous les répondants provient ainsi de l'enquête sur la population active (EPA). Or, il n'est pas possible de classer les individus longitudinaux dans les différentes classes de salaires et traitements à la vague 0 puisque l'information n'est pas disponible dans l'EPA. Pour contourner ce problème, différentes options ont été proposées et évaluées. La solution retenue a été de faire un calage en deux étapes en utilisant l'information disponible provenant de la vague 1. La première étape du calage consiste à ajuster les estimations d'enquête correspondantes à la vague 1 aux totaux de contrôle pour cette même période. Pour cette première étape, tous les totaux de contrôle sont utilisés. Par la suite, la deuxième étape consiste à réajuster les estimations d'enquête de façon à ce qu'elles représentent la population cible, soit celle de la population à la vague 0. Seuls les totaux de contrôle démographiques sont utilisés puisqu'ils sont disponibles dans l'EPA. Étant donné que les totaux démographiques ne varient pas beaucoup pour deux années consécutives, compléter le second calage sans les classes de salaires a peu d'impact sur la distribution salariale résultante. Le deuxième calage permet donc d'ajuster les poids de façon à représenter la population initiale tout en gardant une distribution salariale acceptable.

### 3. PONDÉRATION TRANSVERSALE

---

<sup>2</sup> Un ménage est constitué d'une personne ou un groupe de personnes occupant un même logement.

<sup>3</sup> Une entité économique est constituée d'un groupe de deux personnes ou plus qui vivent dans le même logement et qui sont apparentées par le sang, par alliance, par union libre ou par adoption.

L'aspect transversal de l'enquête sur la dynamique du travail et du revenu est un aspect important de l'enquête car la pondération transversale est utilisée pour la plupart des tableaux de la publication de celle-ci. En effet, les tableaux d'estimations provenant de l'EDTR sont la principale source de données sur le revenu de la population canadienne. La pondération transversale se compose de deux séries de poids : une série de poids travail et une série de poids intégrés au niveau du ménage. Dans ce document, il sera question de la série de poids intégrés au niveau du ménage car c'est la pondération la plus couramment utilisée à l'EDTR. La pondération transversale est semblable à la pondération longitudinale sauf qu'il y a certaines étapes supplémentaires dû au fait que cette pondération doit représenter la population à la vague courante. Par exemple, l'échantillon transversal pour l'année de référence 2003 doit être représentatif de la population des dix provinces canadiennes au 31 décembre 2003. On doit donc déterminer qui sont les personnes répondantes éligibles transversalement. Il y a tout d'abord les répondants longitudinaux faisant partie de la population cible à la période de référence ainsi que leurs cohabitants. Un ménage est considéré comme étant répondant si au moins un de ses membres a répondu à l'une des deux entrevues (l'entrevue sur le travail ou l'entrevue sur le revenu) pour l'année de référence.

Du point de vue transversal, on détient trop peu d'information sur les cohabitants non-répondants pour permettre un ajustement pour la non-réponse à ce niveau. En effet, pour plusieurs cohabitants, les seules variables disponibles sont la province, l'âge et le sexe, variables qui seront de toute façon utilisées à l'étape du calage. Il a donc été décidé de faire un ajustement pour la non-réponse pour les individus longitudinaux seulement, et d'utiliser la méthode du partage de poids afin d'attribuer un poids aux cohabitants. En procédant de cette façon, les poids des cohabitants sont indirectement ajustés pour la non-réponse. La détermination de la valeurs des poids initiaux ainsi que l'ajustement de non-réponse demeurent donc les mêmes pour la pondération transversale.

En théorie, les deux panels peuvent être utilisés indépendamment l'un de l'autre pour produire des estimations transversales, mais en pratique, il est plus efficace de combiner les échantillons transversaux des deux panels afin de produire celles-ci. La taille d'échantillon est alors doublée, ce qui réduit la variabilité des estimations et offre la possibilité d'utiliser un plus grand nombre de totaux de contrôle lors de l'étape du calage aux marges.

Afin de développer une méthode qui permet de combiner l'information des échantillons transversaux des deux panels, il suffit de considérer les deux panels au moment de l'introduction du second panel. La méthode développée par Merkouris (1999) est celle que l'EDTR utilise pour jumeler ses échantillons qui se chevauchent. Merkouris (1999) et Lavallée (1994) suggèrent de combiner les deux panels avant d'intégrer les cohabitants par la méthode du partage de poids. Pour estimer  $Y$ , il s'agit d'utiliser l'estimateur combiné suivant (qui donne une importance relative à chacun des panels selon un certain facteur  $p$ ):  $\hat{Y} = p \hat{Y}_1 + (1 - p) \hat{Y}_2$  où  $\hat{Y}$  est l'estimation transversale pour une variable d'intérêt,  $\hat{Y}_i$  est l'estimation produite pour cette même variable d'intérêt pour le panel  $i$ , et  $p$  est le facteur d'allocation des panels. Le facteur d'allocation des panels est une fonction de l'effet de plan ( $d_i$ ) et du nombre de personnes longitudinales répondantes et

âgées de 16 ans et plus ( $n_i$ ):  $p = \frac{\left(\frac{n_1}{d_1}\right) \left(\frac{n_1 + n_2}{d_1 + d_2}\right)^{-1}}{n_1 + n_2 \left(\frac{d_1}{d_2}\right)}$  Par exemple, pour les panels 3 et 4, on sait que la

différence entre les effets de plan des deux panels est négligeable étant donné qu'ils ont été sélectionnés à l'aide du même plan de sondage. Le facteur d'allocation  $p$  devient donc le rapport entre le nombre de personnes longitudinales de plus de 15 ans dans le panel le plus ancien et la somme des personnes longitudinales répondantes de plus de 15 ans des deux panels. Omettre les cohabitants dans le calcul des facteurs  $p$  contribue à réduire l'importance du panel le plus ancien, et ainsi minimiser les impacts potentiels dus au biais d'érosion.

Les facteurs  $p$  sont appliqués à toutes les personnes longitudinales du panel le plus ancien (troisième panel) éligibles transversalement et faisant partie d'un ménage répondant. Les facteurs  $1-p$  devraient être appliqués à toutes les personnes longitudinales du panel le plus récent (quatrième panel) dans un ménage répondant et éligibles transversalement, à l'exception des personnes hors de la population cible au moment de l'introduction du panel le plus ancien (au 31 décembre 1998). Pour des raisons pratiques, seuls les nouveaux-nés et les immigrants absents lors de l'introduction du troisième panel et présents lors de l'introduction du quatrième panel sont exclus lors de l'application des facteurs  $1-p$ . Les personnes exclues conserveront donc leur poids ajusté pour la non-réponse longitudinale. À partir de ce moment, on ne considère plus qu'un seul échantillon (transversal) dont la période de référence est le moment de l'introduction du panel le plus récent.

L'étape suivante de la pondération transversale est le partage des poids. Cette étape est unique à la pondération transversale et vient du fait que l'EDTR est une enquête longitudinale et que la constitution des ménages change dans le temps. La méthode de partage des poids a été présentée par Lavallée (1995) et a été utilisée pour la première fois à l'EDTR. Elle est utilisée afin d'assigner un poids aux cohabitants puisque ces derniers ne font pas partie de l'échantillon initial et n'ont donc pas de poids initiaux (les probabilités de sélection de ces individus sont inconnues). L'idée sous-jacente à cette méthode est de partager les poids des personnes longitudinales entre elles-mêmes et leurs cohabitants. La première chose à faire lors du partage des poids est de déterminer à quel niveau on souhaite faire le partage des poids. L'EDTR a choisi d'effectuer le partage au niveau de l'unité finale d'échantillonnage, soit le ménage. Une fois cela décidé, il s'agit de déterminer le statut de chaque personne dans l'échantillon transversal, selon que la personne est un individu longitudinal, un cohabitant initialement présent ou initialement absent. Le statut de chaque personne est attribué selon la situation de celle-ci au moment de l'introduction du panel le plus récent (panel 4, au 31 décembre 2001 pour cet exemple). Un cohabitant initialement présent est défini comme étant une personne qui faisait partie de la population cible lors de la sélection du panel le plus récent alors qu'au contraire, un cohabitant initialement absent n'en faisait pas partie. Cette distinction doit être faite pour les cohabitants de façon à déterminer si ces derniers avaient ou non une chance d'être sélectionnés au moment de l'introduction du panel le plus récent. On obtient donc le poids partagé en calculant le rapport entre la somme des poids des longitudinaux ajustés pour la combinaison des panels et le nombre total de personnes longitudinales et de cohabitants initialement présents dans le ménage. Tous les membres du ménage se verront assigner ce poids, incluant les cohabitants initialement absents. Habituellement, le partage des poids est fait seulement pour les ménages contenant au moins un cohabitant. Cependant, pour le poids intégré au niveau du ménage, les poids doivent être égaux pour tous les membres d'un même ménage (Lemaître et Dufour, 1987) alors le partage des poids est effectué pour tous les ménages de l'échantillon.

L'étape de l'ajustement pour la migration inter-provinciale est une deuxième étape directement engendrée par la nature longitudinale de l'EDTR. À l'occasion, il arrive que certaines personnes échantillonnées déménagent d'une province à une autre. Étant donné que la probabilité de sélection peut varier beaucoup selon la géographie, un grand écart est parfois observé entre les poids les plus grands et les poids les plus petits. Par conséquent, si une personne avec une probabilité de sélection assez faible (donc un poids assez grand) déménage dans un endroit où la probabilité de sélection des gens est beaucoup plus forte (donc des poids beaucoup plus petits), ceci peut potentiellement créer un problème. En effet, du point de vue analytique, les résultats d'analyse pourraient se trouver faussés dû à certains cas de migration inter-provinciale. Par exemple, un haut salarié qui quitterait la province de l'Ontario pour aller s'établir à l'Île-du-Prince-Édouard aurait une contribution beaucoup trop importante aux estimations provinciales en raison de son poids trop élevé par rapport à la moyenne des poids à l'Île-du-Prince-Édouard. Il viendrait ainsi gonfler artificiellement le revenu total de sa nouvelle province. La solution choisie pour calculer l'ajustement inter-provincial est de ramener le poids de la personne influente au 95<sup>ième</sup> percentile de la distribution des poids des personnes originaires de la province. Cela permet d'ajuster les personnes dont le poids était supérieur aux poids des personnes sélectionnées dans la province d'origine et de minimiser le changement pour ne pas trop s'éloigner des probabilités de sélection. Pour le poids intégré, l'ajustement est fait au niveau ménage afin de préserver l'égalité des poids à l'intérieur d'un ménage. L'ajustement pour la migration sera donc appliqué à tous les membres du ménage et il est refait à chaque année.

Suite à l'ajustement inter-provincial vient l'ajustement pour les valeurs influentes. On utilise d'abord la source externe T1, qui est le fichier sur l'impôt des revenus des particuliers, pour obtenir la distribution des personnes à hauts revenus dans la population. On considère comme un haut revenu, tout revenu supérieur à 100 000 dollars canadiens. On compare ensuite la distribution provenant de cette source à celle pondérée provenant de l'échantillon, obtenue à la suite d'un calage temporaire (l'étape du calage est expliquée dans les prochains paragraphes). Dans les provinces où l'estimation du nombre de hauts salariés tiré de l'enquête est supérieure à celui provenant du T1, on identifiera alors les individus ayant les plus importantes contributions à la distribution. La méthodologie employée pour identifier les plus importants contributeurs et calculer le facteur d'ajustement provient d'une analyse produite en 1998 sur les données de l'Enquête sur les finances des consommateurs (Tremblay, 1998). Une fois le facteur d'ajustement calculé pour la vague courante, on doit l'appliquer au poids ajusté pour la migration inter-provinciale. Comme le poids transversal est un poids intégré au niveau du ménage, tous les membres du ménage se verront attribuer le même ajustement.

La dernière étape de la pondération transversale est le calage aux marges. Cette étape a pour but d'améliorer la représentativité de l'échantillon en ajustant les poids de sorte que certaines estimations-clés tirées de l'échantillon soient égales à des totaux connus de la population. La stratégie de calage pour l'aspect transversal est similaire à la stratégie

employée pour la première étape du calage dans la pondération longitudinale, c'est-à-dire que tous les totaux de contrôle sont utilisés. Ainsi, on utilise les totaux démographiques issus de recensement 2001 ainsi que le nombre de salariés par classe de salaires et traitements, à la différence près que ce sont les totaux de l'année de référence en cours et non pas les totaux correspondants à la vague 1. Aussi, le classement des individus dans les différentes classes de calage se fait en utilisant les caractéristiques de l'année de référence en cours pour les individus éligibles transversalement. Une autre particularité pour la pondération transversale est que le poids doit être intégré au niveau du ménage (Lemaître et Dufour, 1987). ajoutant ainsi une nouvelle contrainte à satisfaire lors de la régression généralisée. Un enregistrement par ménage sera gardé en additionnant les classes de chacun des membres du ménage de façon à garder la composition du ménage ainsi que ses caractéristiques lors du calage. Il est à noter que pour que la méthodologie soit valide et obtenir un poids intégré après le calage on doit d'abord utiliser un poids avant le calage qui soit un poids intégré après l'ajustement des valeurs influentes. Cette condition est valide pour tous les ménages à l'EDTR car la méthodologie de la pondération transversale a été développée en tenant compte de cette restriction. Les poids transversaux intégrés sont de première importance pour l'EDTR car ils sont utilisés pour produire la plupart des estimations publiées par celle-ci.

**Tableau 1 – Résumé des étapes effectuées pour les pondérations longitudinale et transversale**

<b>Étapes de pondération</b>	<b>Longitudinale</b>	<b>Transversale</b>
<b>Sous-poids de l'EPA</b>	✓	✓
<b>Ajustement de non-réponse</b>	✓	✓
<b>Combinaison de panels</b>		✓
<b>Partage des poids</b>		✓
<b>Migration inter-provinciale</b>		✓
<b>Valeurs influentes</b>	✓	✓
<b>Calage</b>	✓	✓

#### **4. PROJETS FUTURS**

L'enquête sur la dynamique de travail et du revenu est une enquête d'une grande complexité dû en grande partie au fait qu'elle est de nature longitudinale et également utilisée pour produire des estimations transversales. C'est aussi une enquête en constante évolution ; les étapes de pondération de celle-ci sont évaluées fréquemment et sont constamment améliorées. Les procédures de la collecte ont été revues et modifiées pour la prochaine période de collecte qui aura lieu en janvier 2006. De nombreuses modifications à la base de données ainsi qu'à la méthodologie employée pour la pondération ont été faites récemment lors de la révision historique effectuée en mai 2005 (LaRoche, 2005). D'autres activités sont prévues dans un futur proche afin d'augmenter la quantités d'estimation produites pour l'EDTR. Il est question de faire l'imputation des variables sur l'emploi et de repenser la méthodologie utilisée pour la création du poids travail.

#### **REMERCIEMENTS**

Les auteurs tiennent à remercier Julie Bernier, Marie-Noëlle Parent et Michelle Simard pour leurs précieux commentaires.

#### **RÉFÉRENCES**

- Angoss Software (1995), "Knowledge Seeker IV for Windows – User's Guide." Angoss Software International Limited.
- Gambino, J.G., Singh, M.P., Dufour, J., Kennedy, B. et Lindeyer, J. (1998). Méthodologie de l'enquête sur la population active du Canada, Statistique Canada, Catalogue 71-526-XPB
- Kass, G. V. (1980), "An explanatory technique for investigating large quantities of categorical data." *Applied Statistics*, Vol. 29, no 2, 119-127.

LaRoche, S. (2005). «Stratégie de calage de l'Enquête sur la dynamique du travail et du revenu ». *Recueil 2005 de la section des méthodes d'enquête, Société statistique du Canada.*

Lavallée, P. (1995). Pondération transversale des enquêtes longitudinales menées auprès des individus et des ménages à l'aide de la méthode du partage de poids, *Techniques d'enquête*, **vol. 21, no 1**, 27-35.

Lemaître, G. et Dufour, J. (1987). « Une méthode intégrée de pondération des personnes et des familles ». *Techniques d'enquête*, **vol. 13, no 2**, 211-220.

Merkouris, T. (1999). Cross-sectional Estimation in Multiple-Panel Household Surveys, document de travail de la direction de méthodologie, Statistique Canada, HSMD-99-004E.

Naud, J.-F. «Pondération longitudinale avec panels combinés, Enquête sur la dynamique du travail et du revenu.» *Série de documents de recherche - Revenu*, Statistique Canada, numéro au catalogue 75F0002MIF2004008, juin 2004.

Tremblay, J. (1998). Détection des observations influentes pour l'enquête sur les finances des consommateurs (EFC) et l'enquête sur la dynamique du travail et du revenu (EDTR), document interne, Statistique Canada.