

# LEÇONS À TIRER DE L'UTILISATION DE DONNÉES ADMINISTRATIVES AU STADE DE LA PONDÉRATION

Michel Latouche<sup>1</sup>

## RÉSUMÉ

L'utilisation des données administratives dans différentes étapes d'une enquête est une pratique courante. L'étape de la pondération n'est pas une exception et beaucoup d'enquêtes ont recours au calage pour obtenir des gains en précision ou pour balancer un échantillon imparfait. Même si le calage contribue à améliorer la qualité de certaines estimations, il peut en même temps nuire à d'autres comme il le sera illustré dans cet article. L'élaboration d'une stratégie de calage peut devenir laborieuse et nécessiter beaucoup d'investissements en temps et ressources. Dans le cas où on cherche à équilibrer un échantillon, il est préférable de tout mettre en œuvre pour prévenir les problèmes au moment de la conception de l'enquête et de la collecte des données.

MOTS CLÉS : Calage; estimation par régression généralisée; pondération; qualité.

## ABSTRACT<sup>2</sup>

The use of administrative data in different survey steps is a common practice. The weighting step is no exception and many surveys use calibration to obtain gains in precision or to balance an imperfect sample. Even though calibration helps to improve the quality of certain estimates, it can at the same time worsen others as will be illustrated in this article. The development of a calibration strategy can become laborious and require a large investment in time and resources. Rather than trying to balance the sample, it is preferable to make every effort to prevent problems at the time of development and during data collection.

KEY WORDS: Calibration, Generalised Regression Estimator; Quality, Weighting.

## 1. INTRODUCTION

Les données administratives ont toujours été très utilisées dans les enquêtes par échantillonnage. Elles peuvent servir à construire des bases de sondage ou être employées pour fins de stratification, de validation ou d'estimation. Brackstone (1987) présente une revue de l'utilisation des données administratives ainsi que les inconvénients s'y rattachant. Le recours aux données administratives a des avantages certains. De telles utilisations permettent de réduire les coûts associés à la réalisation des enquêtes, d'améliorer la qualité des résultats ou d'obtenir des gains d'efficacité.

En particulier, les données administratives sont de plus en plus utilisées à l'étape de la pondération. Cette pratique, appelée calage, consiste à ajuster les poids des unités échantillonnées de manière à ce que certaines estimations faites à partir de cet échantillon respectent des jalons (totaux ou comptes) tirés d'autres sources fiables. Tandis que les méthodologistes ont recours au calage afin d'améliorer la qualité des estimations, les gestionnaires d'enquête y voient de plus en plus une façon de réduire les coûts de production et le fardeau des répondants. De leur côté, les analystes apprécient également la cohérence que le calage apporte facilitant d'autant plus les comparaisons entre les résultats de différentes enquêtes. Les avantages du calage sont si évidents qu'on est porté à oublier qu'il n'est qu'une solution palliative et ne contribue pas à régler les problèmes avant qu'ils ne se produisent.

---

<sup>1</sup> Michel Latouche, Immeuble R.H. Coats, 18<sup>e</sup> étage, Section O, 120 avenue Parkdale, Ottawa (Ontario) K1A 0T6, [michel.latouche@statcan.ca](mailto:michel.latouche@statcan.ca).

<sup>2</sup> This paper is also available in English. Please contact the author for a copy.

À la fin des années 90, Statistique Canada a initié un projet visant à remanier la stratégie de calage de l'ensemble des enquêtes portant sur les dépenses, le revenu et la richesse des ménages. Plusieurs enquêtes ont été affectées dont entre autres l'Enquête sur la dynamique du travail et du revenu (EDTR) et l'Enquête sur les dépenses des ménages (EDM). Ces enquêtes sont décrites par Lévesque et Franklin (2000) et par Arsenault et coll. (2001).

Bien que le calage permette de résoudre des problèmes importants, il n'en demeure pas moins qu'il représente une solution imparfaite et qu'il est préférable de l'utiliser avec précaution. Le but de cet article est de mettre en parallèle les avantages et les inconvénients de l'utilisation de données administratives à l'étape de la pondération. La section 2 décrit brièvement ce qu'est le calage au moyen de l'estimateur par régression généralisée. Ensuite, la section 3 discute des avantages et des inconvénients du calage à l'aide d'exemples tirés de l'EDTR et de l'EDM. L'article se termine par une brève discussion et une conclusion.

## 2. CALAGE ET ESTIMATION PAR RÉGRESSION GÉNÉRALISÉE

La stratification a posteriori est la forme la plus simple de calage (Cochran, 1977). Dans ce cas on veut s'assurer que pour un échantillon  $s$  l'estimateur du total estime de façon exacte le nombre total d'unités (individus, entreprises, etc) de la population par catégorie d'une variable auxiliaire  $X$  particulière. Le nombre d'individus par croisement de genre, groupe d'âge et province de résidence est un exemple de variable auxiliaire fréquemment employée dans les enquêtes sociales. Ainsi, pour une unité  $k$  dans la catégorie  $X_j$ , le poids d'échantillon  $a_k$  est ajusté par un facteur  $g_k$  donné par  $g_k = \frac{N_j}{\hat{N}_j}$  où  $N_j$  est le nombre d'unités de la population qui se trouve dans la catégorie  $j$ , et  $\hat{N}_j = \sum_{k \in s_j} w_k$  est le nombre d'unités dans la catégorie  $j$  tel qu'estimé par l'échantillon avant d'ajuster les poids. Ainsi, les poids finaux,  $w_k = g_k a_k$ , sont dit calés selon la caractéristique  $x$  et l'estimation du total pour une variable d'intérêt  $y$  est donnée par  $\hat{Y}_{REG} = \sum_{k \in s} w_k y_k$ .

De façon générale, il est possible de caler selon plusieurs variables simultanément. Par exemple, caler sur le nombre d'individus par croisement de groupe d'âge, genre et province ainsi que sur le nombre de ménages selon leur taille (personne seule, 2 personnes, 3 personnes, etc) par province. Dans ce cas, on utilise un estimateur par régression généralisée (Särndal et coll. 1992). Il s'agit d'une méthode d'estimation assistée par un modèle. En utilisant une notation matricielle similaire à celle utilisée par ces auteurs, on a pour chaque unité  $k$  le vecteur de variables auxiliaires  $\mathbf{x}_k = (x_{1k}, \dots, x_{jk}, \dots, x_{Jk})'$  où  $J$  est le nombre de totaux (ou de comptes) à respecter et  $x_j$  vaut 1 si l'individu  $k$  appartient à ce groupe ou 0 sinon. Le modèle utilisé est donné par :

$$\begin{aligned} \text{Modèle } \xi : \quad & y_k = \mathbf{x}'_k \boldsymbol{\beta} + \varepsilon_k \\ \text{avec} \quad & E_\xi(\varepsilon_k) = 0, \text{Var}_\xi(\varepsilon_k) = \sigma^2 \text{ et } \text{Cov}_\xi(\varepsilon_k, \varepsilon_l) = 0 \end{aligned}$$

Ainsi, selon ce modèle, l'estimation du total pour une variable d'intérêt  $y$  est donnée par :

$$\hat{Y}_{REG} = \sum_{k \in s} a_k g_k y_k \quad (2.1)$$

où le facteur d'ajustement des poids est donné par :

$$g_k = 1 + (\mathbf{X} - \hat{\mathbf{X}})' \hat{\mathbf{T}}^{-1} \mathbf{x}_k \quad (2.2)$$

avec  $\mathbf{X} = \sum_{k \in U} \mathbf{x}_k$ ,  $\hat{\mathbf{X}} = \sum_{k \in s} a_k \mathbf{x}_k$  et  $\hat{\mathbf{T}} = \sum_{k \in s} a_k \mathbf{x}_k \mathbf{x}'_k$ .

Un grand nombre d'enquêtes à Statistique Canada utilise ce genre d'estimateur. L'Enquête sur la dynamique du travail et du revenu (LaRoche, 2005) de même que l'Enquête sur les dépenses des ménages (Arsenault et coll. 2001) ajustent les poids d'échantillon selon les comptes démographiques provinciaux d'individu par croisement de groupe d'âge et genre, comptes provinciaux de ménage par taille ou type, taille de famille et également selon le salaire des individus. Bien que les enquêtes sociales calent essentiellement sur des comptes, il est également possible de caler sur des variables quantitatives. Par exemple la série d'enquêtes sur les cultures menées par Statistique Canada calent les poids selon la superficie cultivable totale au niveau des provinces (Yung, 2000).

### 3. AVANTAGES ET INCONVÉNIENTS DU CALAGE

Le calage sert avant tout à améliorer la qualité des estimations. L'utilisation de variables auxiliaires contribue à réduire la variance des estimations si la relation linéaire entre la variable d'intérêt et les variables auxiliaires est bonne (Cochran, 1977, Särndal et coll. 1992). Le calage peut également servir à améliorer l'exactitude des estimations. Combiné avec un traitement approprié des données (vérification et imputation) le calage réduit les problèmes de sous couverture et les erreurs de non réponse. Par exemple, on a recours au calage pour équilibrer un échantillon ayant un important problème de couverture causée par une base de sondage défailante ou par la non réponse (Tremblay, 2005). Ce point sera discuté dans la section 3.1.

Il y a cependant d'autres avantages à caler les poids. Lorsque le plan de sondage de l'enquête a été établi et que les données ont été recueillies, le calage permet a posteriori d'améliorer les estimations. Le calage permet également d'améliorer les estimations diffusées antérieurement et de maintenir la cohérence dans les séries chronologiques lorsque des changements majeurs sont apportés à la méthodologie d'enquêtes répétées. Par exemple en 1992, l'Enquête sur l'emploi, la rémunération et les heures a changé de base de sondage ce qui a entraîné une augmentation importante de la couverture et des différentes estimations produites par l'enquête. Le calage a permis de réviser la série chronologique en calant sur de nouveaux totaux basés sur les données de la nouvelle base de sondage (Gossen et Latouche, 1992).

#### 3.1 Cohérence et exactitude

L'utilisation de jalons provenant de sources administratives assure une plus grande cohérence entre les estimations de l'enquête et la source administrative utilisée lors du calage. Par extension, le calage permet également de mieux réconcilier les estimations produites par différentes enquêtes utilisant les mêmes jalons. Les analystes se sentiront plus à l'aise de comparer les caractéristiques des ménages ou des individus provenant d'enquêtes différentes si ces enquêtes produisent les mêmes estimations de base. Les utilisateurs des données produites par une organisation statistique seront plus en confiance si les estimations de bases (e.g. tailles de population) sont les mêmes d'une enquête à l'autre. Cependant, il faut s'assurer que les jalons utilisés sont fiables, disponibles en permanence et que les concepts s'y rattachant soient compatibles avec ceux de l'enquête. Finalement, comme il n'existe pas de source parfaite, l'évaluation des sources administratives doit se faire sur une base régulière.

Le calage peut améliorer l'exactitude pour des variables non contrôlées mais reliées aux variables auxiliaires utilisées lors du calage. L'Enquête sur la dynamique du travail et du revenu cale, entre autres variables, sur le nombre d'individus ayant un revenu de salaires et traitements situé dans une classe déterminée. Chaque province est divisée en sept classes de salaires et traitements (LaRoche, 2005). Le tableau 1 illustre l'impact de cette variable sur les estimations. Dans ce tableau les estimations de l'EDTR pour le revenu total selon certaines sources sont comparées à celles produites par le Système des comptes nationaux (SCN). Cette source est la plus exacte possible car la comptabilité nationale compile les estimations en réconciliant plusieurs sources indépendantes. Avant l'utilisation de cette variable auxiliaire, l'EDTR avait des gains totaux (salaires, traitements, commissions et revenu de travailleur autonome) 6% supérieur à ceux du SCN. Par contre, avec cette variable auxiliaire dans le calage, l'EDTR est maintenant plus proche bien que 2% inférieure au SCN. Notez que la correction n'est pas parfaite car le calage est fait sur le nombre de salariés et non pas sur les montants reçus en salaires et traitements. De même, les transferts gouvernementaux (assurance emploi, aide sociale, etc) qui étaient 4,3% supérieurs sont maintenant à 1% du SCN. Par contre, les investissements ont été peu affectés et on constate une légère augmentation de l'écart au niveau du revenu total.

Même si beaucoup de variables auxiliaires sont disponibles, il n'est pas recommandé de les utiliser toutes car cela risque de nuire à d'autres estimations. Pour illustrer ce point, considérons les estimations faite par l'EDTR lors de l'année de référence 1998. Le tableau 2 montre l'impact de l'ajout de variables auxiliaires sur les estimations produites par l'enquête. Lorsqu'aucun calage n'est fait, l'EDTR sous-estime la population totale de 13%. Les entités économiques de taille 1 sont également sous-représentées de 8% mais les ménages de taille 1 sont surreprésentés de 6%. En calant sur les projections démographiques au niveau des groupes d'âge-genre par province, les comptes de population sont maintenant exacts mais les ménages de taille 1 sont surreprésentés de 20% au lieu de 6 et les entités économiques sont surreprésentées de 7% alors

qu'elles étaient sous-représentées auparavant. En ajoutant la variable taille de ménage au calage, les entités économiques de taille 1 sont sous-représentées de 15% et il faut donc ajouter cette variable au calage. En faisant cela, on surestime le nombre de salariés et sous-estime le nombre de familles monoparentales de 6%. Finalement, l'ajout de la variable nombre de salariés au calage ne semble pas influencer sur l'estimation du nombre de familles monoparentales.

Tableau 1  
 Comparaison entre l'EDTR et le système des comptes nationaux  
 par sources de revenu (année 1997)

Sources de revenu des individus	proportion du revenu total (%)	Sans calage selon le salaire (%)	Avec calage selon le salaire (%)
Gains	65	6,0	-2,0
Investissements	15	-46,0	-45,0
Transferts gouvernementaux	13	-4,3	-1,0
REVENU TOTAL	100	0,8	-1,8

Tableau 2  
 Écarts (%) entre les comptes de l'EDTR et les sources externes  
 (année 1998, basée sur les comptes du recensement 1996)

variables auxiliaires utilisées lors du calage	Population totale	Nombre de ménages de taille 1	Nombre d'entités économiques de taille 1	Nombre de salariés	Nombre de familles monoparentales
aucune	-13	6	-8	3	-
province-âge-genre (1)	0	20	7	-	-5
(1)+ ménages (2)	0	0	-15	-	-12
(1)+(2)+entités économiques (3)	0	0	0	6	-6
(1)+(2)+(3)+salaires	0	0	0	0	-6

Cet exemple illustre que le calage peut avoir des effets indésirables sur les variables exclues du calage et ce, même si la taille de l'échantillon peut supporter un calage important. Le calage règle des problèmes au niveau des agrégations mais ne garantit pas que la multitude de relations entre les différentes variables soit préservée. La meilleure façon de maintenir ces relations est de minimiser les erreurs à la source en appliquant les principes de base de la méthodologie d'enquête au moment de la conception.

### 3.2 Gain en précision

Cochran (1976) a démontré que, dans de grands échantillons aléatoires simples, la variance de l'estimateur par régression sera plus petite que la variance de l'estimateur avec facteur d'extension (Horvitz-Thompson) si la corrélation entre la variable d'intérêt et les variables auxiliaires est élevée. Dans les enquêtes à plan complexe, la démonstration n'est pas aussi simple. La plupart des enquêtes sociales de Statistique Canada utilisent un plan à plusieurs degrés et calcule les variances à l'aide de la méthode du Bootstrap (Efron, 1979, Särndal et coll. 1992).

Figure 1  
 Enquête sur les dépenses des ménages  
 Effets du calage sur les CV des estimations de revenu

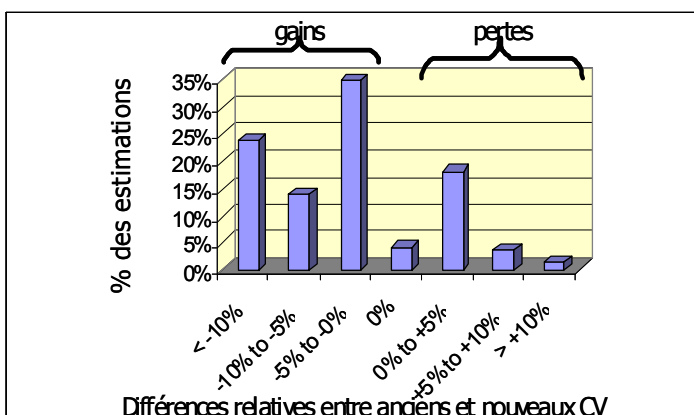
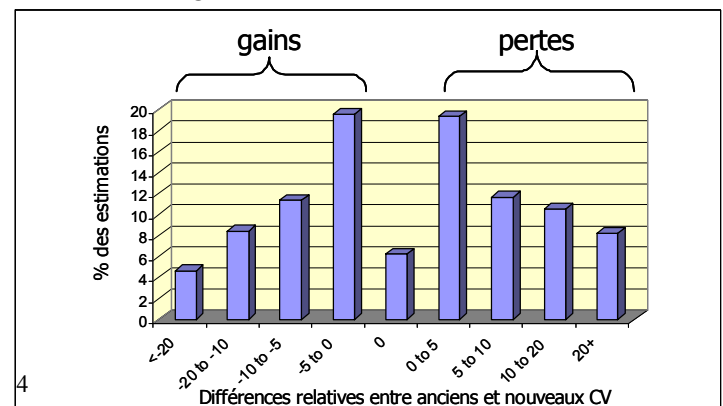


Figure 2  
 Enquête sur la dynamique du travail et du revenu  
 Effets du calage sur les CV des estimations de revenu



Le modèle  $\xi$  introduit à la section 2 a été développé pour une seule variable  $Y$ . Les variables auxiliaires utilisées ont été sélectionnées en fonction de cette seule variable. On peut donc s'attendre à des gains en précision pour les estimations de  $Y$  ainsi que pour d'autres variables corrélées avec le vecteur  $\mathbf{x}$ . Ainsi pour l'EDTR, l'utilisation du nombre de salariés dans le calage donne des estimations provinciales du nombre de salariés ayant une variance nulle et une baisse de la variance pour les estimations provinciales du total des montants de salaires et traitements. Cependant, qu'en est-il des autres variables?

Le modèle  $\xi$  est défini à un niveau d'agrégation donné (e.g. comptes par province) alors que des estimations sont produites à des niveaux plus fins (e.g. agglomérations urbaines). Autrement dit, le nombre de variables auxiliaires (ou la dimension de  $\mathbf{x}$ ) est limité alors qu'on voudrait utiliser encore plus de variables auxiliaires. De plus, il peut y avoir une certaine dépendance entre les variables auxiliaires. Ces facteurs ont pour effet de limiter les gains en précision. La figure 1 montre la distribution des différences relatives entre les coefficients de variation (CV) avec et sans calage selon les salaires et traitements. Cette figure a été produite à l'aide d'une centaine d'estimations agrégées tirées de l'EDM. Une différence négative signifie un gain de précision tandis qu'une différence positive indique une perte de précision. Lorsque les estimations sont produites à des niveaux similaires à ceux des totaux de contrôle, les gains en précision sont appréciables. Par contre, la figure 2 montre la même distribution tirée à partir de l'EDTR sur plusieurs milliers d'estimations pour des niveaux désagrégés. Dans ce cas, les résultats sont plus mitigés; il y a autant de pertes que de gains. Les effets du calage sont moins probants à des niveaux plus désagrégés car le calage ne se fait qu'au niveau provincial.

La dimension du vecteur  $\mathbf{x}$  utilisé dans le modèle  $\xi$  est non seulement limitée par les sources externes disponibles mais surtout par la taille de l'échantillon. En effet, il est hautement désirable que les poids ajustés soient le plus près possible des poids de sondage afin de préserver les avantages de l'approche probabiliste. On veut donc que les facteurs  $g$  de l'équation (2.2) soient le plus proche de 1. Cela est possible si la dimension de  $\mathbf{x}$  est petite comparativement à la taille de l'échantillon et si les données ont un minimum de congruence avec les contraintes imposées. À ce titre, la distribution des facteurs  $g$  est révélatrice. La figure 3 par exemple montre l'effet de l'ajout de la variable nombre de salariés au calage de l'EDM.

Un autre élément à considérer est que les variables auxiliaires interagissent entre elles. En effet, les individus  $k$  se retrouvent dans plusieurs catégories simultanément (e.g. groupe des hommes de 25-34 ans, dans un ménage de taille 5 et une classe de très hauts salariés). Même si chaque élément du vecteur  $\hat{\mathbf{X}} = \sum_{k \in S} a_k \mathbf{x}_k$  est grand, il faut se rappeler que ces éléments se croisent comme si le modèle  $\xi$  comportait des termes d'interaction et que le nombre d'observations dans les différents croisements peut être beaucoup plus petit. Pour l'EDTR, la distribution des éléments du vecteur  $\hat{\mathbf{X}}$  selon leur taille est présentée à la figure 4. On peut constater que 89 éléments ont moins de 300 observations indiquant que plusieurs croisements sont potentiellement petits, contribuant ainsi à aplanir la distribution des facteurs  $g$ .

Figure 3  
Enquête sur les dépenses des ménages  
Distribution des facteurs  $g$

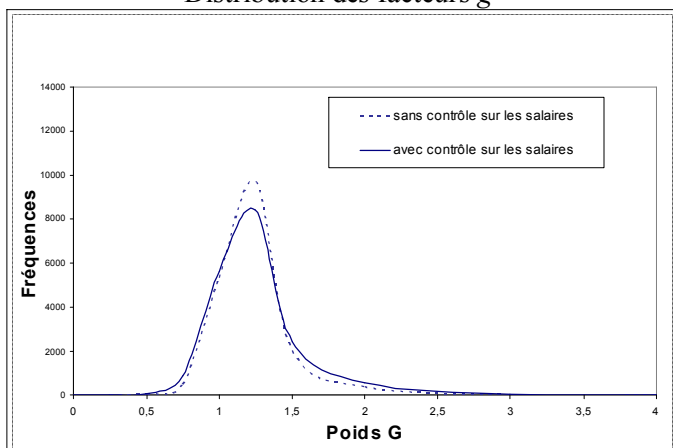
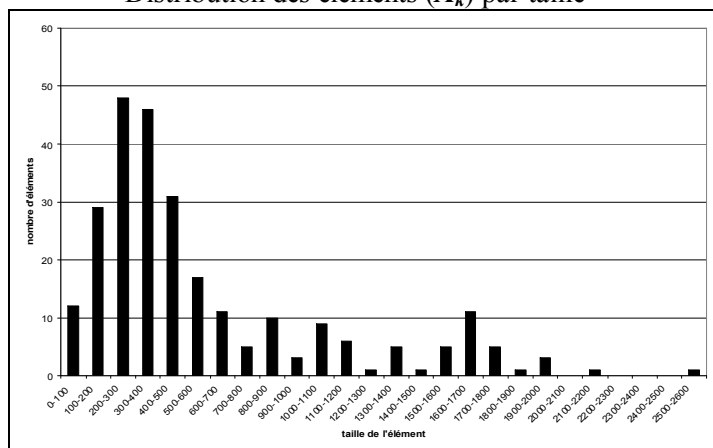


Figure 4  
Enquête sur la dynamique du travail et du revenu  
Distribution des éléments ( $\mathbf{X}_k$ ) par taille



Un dernier élément qui contribue à étirer la distribution des facteurs d'ajustement est la contrainte de la pondération intégrée. Souvent les analystes demandent à ce que les poids des individus d'un même ménage soient le même (Lemaître et Dufour 1987). L'intégration des poids au niveau du ménage impose des restrictions supplémentaires importantes qui sont presque exclusivement absorbées par certains types d'unités. Dans le cas de l'EDTR il s'agit des entités économiques de taille 1 vivant dans des ménages de taille 2 ou supérieure car les entités économique de taille 1 sont beaucoup plus sur-représentées que les ménages de tailles 1 (Cf. table 2). Un phénomène similaire explique la sous-représentation des familles monoparentales dans l'EDM.

De façon générale, l'augmentation de la dimension de  $x$  contribue à aplanir la distribution des facteurs et favorise la création de valeurs très grandes contribuant à augmenter la variance dans des petits domaines. Cela crée également des facteurs plus petits que l'unité et même négatifs ce qui crée un certain malaise chez les utilisateurs et analystes minant la crédibilité de l'enquête. Bien qu'il existe des méthodes pour éliminer les petits facteurs d'ajustement (Lévesque et Franklin, 1998), le calcul des variances par des méthodes numériques est plus laborieux, voire problématique car la fréquence des petits facteurs est plus grande.

#### **4. DISCUSSION ET CONCLUSION**

L'élaboration d'une stratégie de calage demande beaucoup de précaution et d'analyses. Après avoir trouvé des sources d'informations externes il faut en vérifier régulièrement la qualité. L'impact du calage peut se voir au niveau de la précision et de l'exactitude et par conséquent la stratégie doit être évaluée en termes de variance et de biais calculés à différents niveaux d'agrégation. Comme le calage peut avoir un impact différent d'une variable à l'autre, cette tâche est d'autant plus laborieuse que le contenu de l'enquête est grand.

Il faut veiller à ne pas abuser du calage. Bien qu'il soit tentant de régler le plus de problèmes possibles, l'addition de jalons ne garantit pas une réduction de variance pour l'ensemble des estimations et des domaines d'intérêt. Le calage étant défini au niveau de certaines marginales, il n'est pas garanti que les relations à des niveaux désagrégés soient préservées ou que l'échantillon soit capable d'absorber tous ces jalons. Comme nous l'avons vu à la section 3.1, l'ajout de jalons peut entraîner une augmentation du biais pour d'autres variables non contrôlées.

Le recours au calage pour équilibrer un échantillon défaillant est un outil essentiel. Il permet de corriger des problèmes mais pas de les éviter. Il en résulte un produit amélioré mais malgré tout imparfait. Autant que possible il est préférable de tout mettre en œuvre pour prévenir les problèmes au moment de la conception de l'enquête et de la collecte des données. Le calage est un outil qui contribue à régler des problèmes évidents, mais peut également avoir des impacts négatifs.

Après une série de travaux échelonnés sur cinq ans, la Division de la statistique du revenu de Statistique Canada a finalisé la mise en œuvre de la nouvelle stratégie de calage. La stratégie implantée utilise toutes les variables prévues originalement mais le nombre de contrôles par variable a été réduit afin de minimiser les effets indésirables. Les utilisateurs ont très bien accueilli les nouvelles estimations. Cependant, les méthodologistes évaluent régulièrement les résultats ainsi que la qualité des variables auxiliaires. Un effort important est également consacré au travail sur le terrain pour améliorer la couverture et réduire autant que possible le recours au calage.

#### **REMERCIEMENTS**

L'auteur tient à remercier Didier Garriguet, Normand Laniel, Johanne Tremblay, Sylvie LaRoche et Caroline Cauchon pour leurs précieux commentaires et suggestions.

## RÉFÉRENCES

- Arsenault, S. et Tremblay, J. (2001). *La méthodologie de l'Enquête sur les dépenses des ménages*. Catalogue no. 62F0026MIF-01003, Statistique Canada.
- Arsenault, S., Gaudet, J., Nadeau, C. et Tremblay, J. (2001). *Introduction of a New Calibration Strategy for the Survey of Household Spending*. Proceedings of the Joint Statistical Meetings, American Statistical Association.
- Brackstone, G.J. (1987). *Utilisation des dossiers administratifs à des fins statistiques*. Techniques d'enquête, 13, 35-51.
- Cochran, W.G. (1977). *Sampling Techniques* (3e éd.). New York: John Wiley & Sons.
- Efron, B. (1979). *Bootstrap Methods: Another Look at the Jackknife*. Annals of Statistics 7, pp. 1-21.
- Gossen, M. and Latouche, M. (1992). *Post-Stratification to Reduce Sample Bias in an Establishment Survey*. Internal document of Statistics Canada.
- LaRoche, S. (2005). *Stratégie de calage de l'Enquête sur la dynamique du travail et du revenu*. Recueil de la Conférence annuelle de la Société Statistique du Canada, Section de la méthodologie d'enquête.
- Lavigne, M. and Michaud, S. (1998). *Aspects généraux de l'Enquête sur la dynamique du travail et du revenu*. Article tiré de la Série de documents de recherche sur le revenu de Statistique Canada No. 75F0002MIF1998005.
- Lemaître, G. et Dufour, J. (1987). *Une méthode intégrée de pondération des personnes et des familles*. Techniques d'enquête, vol. 13, no 2, pp. 211-220.
- Lévesque, I. et Franklin, S. (2000). *Pondération longitudinale et transversale de l'Enquête sur la dynamique du travail et du revenu, année de référence 2000*. Série de document de recherche – Revenu, Statistique Canada. No au catalogue 75F0002MIF-00004, juin 2000.
- Särndal, C.-E., Swensson, B. et Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Tremblay, J. (2005). *Aperçu de la stratégie de calage harmonisée pour les statistiques du revenu*. Recueil de la Conférence annuelle de la Société Statistique du Canada, Section de la méthodologie d'enquête.
- Yung, W. (2000). *Addressing Frame Attrition in the Crops Surveys Through Poststratification*. Document interne de la Division des méthodes d'enquête auprès des entreprises, Statistique Canada.