Techniques d'enquête, décembre 1990
Vol. 16, n° 2, pp. 177–190
Statistique Canada

Analyse d'enquêtes à passages répétés

WAYNE A. FULLER[1]

RÉSUMÉ

Dans cet article, nous nous intéressons principalement aux enquêtes à passages répétés où une partie des unités de l'échantillon est observée sur plusieurs périodes et une partie n'est pas observée à certaines périodes. Nous voyons en quoi consiste l'estimation par les moindres carrés pour de telles enquêtes. Nous nous arrêtons aussi à des méthodes d'estimation en vertu desquelles les estimations existantes n'ont pas à être révisées lorsque de nouvelles données sont connues. Par ailleurs, nous considérons des méthodes pour estimer des paramètres longitudinaux; mentionnons à cet égard les tableaux de variation brute. Nous décrivons aussi la méthode d'estimation utilisée dans une enquête à passages répétés sur l'utilisation du sol, réalisée par le U.S. Soil Conservation Service. Enfin, nous illustrons l'effet de l'erreur de mesure sur les estimations de la variation brute et montrons que les plans de sondage qui permettent d'estimer les paramètres du processus d'erreur de mesure peuvent être très efficients.

MOTS CLÉS: Échantillon d'enquête; moindres carrés; erreur de mesure; variation brute.

1. INTRODUCTION

L'analyse d'enquêtes à passages répétés suscite beaucoup d'intérêt. Soulignons à cet égard la publication récente des actes d'un symposium sur les enquêtes par panel, colligés par Kasprzyk, Duncan, Kalton et Singh (1989), la tenue de séances sur la question lors des assemblées de l'Institut international de Statistique de 1987 et de 1989, et le Symposium sur l'analyse des données dans le temps, organisé par Statistique Canada en octobre 1989. Dans l'article qu'ils ont présenté à la session de l'IIS de 1989 à Paris, Smith et Holt (1989) parlent d'un intérêt renouvelé pour l'élaboration et l'analyse d'études longitudinales. Ils soulignent que des spécialistes de domaines comme la sociologie et la santé réalisent depuis longtemps des enquêtes par panel et des études de cohorte. Ils citent Lazarsfeld et Fiske (1938). Dans le domaine de la santé, mentionnons l'article de Garcia, Battese et Brewer (1975).

Les organismes officiels réalisent de nombreuses enquêtes périodiques, comme l'enquête sur la population active. Ces enquêtes produisent habituellement une suite de rapports comme ceux portant sur l'emploi et le chômage pour la période courante. En règle générale, les enquêtes réalisées par les organismes officiels fournissent très peu de données sur le comportement des unités de l'échantillon dans le temps. La U.S. Survey of Income and Program Participation est un exemple d'enquêtes qui servent à produire des estimations longitudinales. Voir à ce sujet Kasprzyk et McMillen (1987). Bien que nous en sachions moins sur les enquêtes réalisées par le secteur privé que sur celles réalisées par les administrations publiques, il semble que les premières servent surtout, comme les secondes, à produire une suite de rapports pour des périodes données. Toutefois, le secteur public comme le secteur privé doivent répondre à une demande accrue d'analyses longitudinales.

L'élaboration d'une taxinomie pour les enquêtes à passages répétés a pour effet de mettre en relief les questions complexes qui accompagnent ce genre d'enquêtes. Duncan et Kalton (1987) énumèrent sept objectifs des enquêtes à passages répétés, soit:

A. produire des estimations de paramètres de la population pour des périodes déterminées;
B. produire des estimations de paramètres de la population pour des périodes combinées;
C. mesurer la variation nette;

[1] Wayne A. Fuller, Département de statistique, Iowa State University, Ames, Iowa, 50011, É.-U.

D. mesurer des éléments de la variation, dont
   i) la variation brute
   ii) la variation pour une unité
   iii) la variabilité pour une unité
E. produire des données agrégées sur les unités prises individuellement
F. déterminer la fréquence, le moment et la durée d'événements
G. accumuler des données sur des populations peu courantes.

Bien que cela ne soit pas explicite, plusieurs de ces objectifs supposent l'estimation des paramètres de modèles spécialisés.

Par ailleurs, Duncan et Kalton définissent quatre genres d'enquêtes: 1) l'enquête à passages répétés, où rien n'est fait pour veiller à ce que des éléments particuliers de la population fassent partie de l'échantillon plus d'une fois, 2) l'enquête à échantillon constant, où les mêmes éléments sont observés à chaque période, 3) l'enquête avec renouvellement de l'échantillon, où des éléments de la population sont observés pour un nombre déterminé de périodes, puis supprimés de l'échantillon par renouvellement selon un plan déterminé, et 4) l'enquête à panel fractionné, qui est une combinaison de l'enquête à échantillon constant et de l'enquête 1) ou 3). Duncan et Kalton indiquent aussi sous forme de tableau les genres d'enquêtes qui conviennent le mieux aux différents objectifs.

Lorsqu'un établissement réalise une enquête à passages répétés, il doit parer à toutes les difficultés qui accompagnent normalement l'exécution d'une enquête unique sauf que dans ce cas-ci, les problèmes sont amplifiés. Pour assurer la qualité d'une enquête à passages répétés, il est nécessaire de procéder toujours de la même façon sur le terrain et d'appliquer les mêmes méthodes de traitement, de gestion de données et d'estimation pour toutes les périodes. Il est difficile d'obtenir la collaboration constante des répondants pour plusieurs périodes successives et tout aussi difficile de retrouver les répondants qui ont déménagé. L'erreur de réponse est présente dans tous les genres d'enquête sauf que dans le cas des enquêtes à passages répétés, il faut composer avec un phénomène de "conditionnement" lié à la répétition des interviews. De plus, les erreurs de réponse ont pour effet de créer des incohérences dans les données lorsque celles-ci sont recueillies sur une longue période. Enfin, le changement de composition des unités, telles les familles, vient compliquer à la longue l'estimation et l'analyse.

Nous n'aborderons ici que quelques-unes des questions qui se rattachent aux enquêtes à passages répétés. Notre analyse est fondée sur une grande enquête réalisée par le U.S. Soil Conservation Service en collaboration avec l'université Iowa State. Dans la section 2, nous examinons quelques-unes des méthodes d'estimation utilisées dans les enquêtes à passages répétés. Cette analyse se prolonge dans la section 3, où il est surtout question de l'estimation de paramètres longitudinaux dans les enquêtes par panel. Dans la section 4, nous exposons brièvement les méthodes d'estimation utilisées dans l'enquête du U.S. Soil Conservation Service. Enfin, la section 5 renferme une brève description des effets de l'erreur de mesure sur les estimations de la variation brute.

2. ESTIMATION

Dans cette section, nous allons exposer à grands traits la méthode d'estimation par les moindres carrés généralisés appliquée à des enquêtes où seul un sous-ensemble des éléments de l'échantillon est observé pendant des périodes consécutives. La méthode des moindres carrés généralisés est la première méthode à laquelle se sont intéressés les auteurs qui étudiaient l'estimation dans les enquêtes à passages répétés. Sur les traces de Cochran (1942), Jessen (1942) fut le premier à envisager la construction de poids à variance minimum pour une série d'estimateurs non biaisés établis pour chaque période visée par l'enquête.

Jessen (1942) a analysé le cas particulier de l'échantillonnage effectué à deux reprises où le nombre d'observations diffère d'un échantillon à l'autre et s'est intéressé à la répartition

optimale des unités entre les groupes d'échantillons chevauchants et non chevauchants. Patterson (1950) a examiné le cas de T sondages successifs avec plusieurs modes de renouvellement partiel des unités. Le plan d'échantillonnage le plus simple prévoyait le renouvellement d'une proportion déterminée des unités d'échantillonnage à chaque nouveau sondage. En outre, Patterson (1950) avait supposé que, pour une unité i donnée, les écarts $x_{ti} - x_t$, $t = 1, 2, \ldots$, suivaient un processus autorégressif du premier ordre, x_{ti} étant la valeur de l'unité de population i au temps t, et x_t, la moyenne de la population finie correspondante. Suivant le modèle d'erreur qui en a découlé, il a défini des estimateurs optimaux des valeurs fixes x_t et des écarts $x_t - x_{t-1}$. Il s'est également penché sur l'estimation optimale de x_t suivant des formes généralisées du plan de renouvellement partiel, la détermination de la taille optimale de l'échantillon et l'estimation avec erreurs non autorégressives.

La méthode des moindres carrés a été approfondie par Eckler (1955), Gurney et Daly (1965) et Jones (1980). On en est venu aussi à parler d'estimateurs composites; voir à ce sujet Rao et Graham (1964), Graham (1973) et Wolter (1979). Battese, Hasabelnaby et Fuller (1989) décrivent comment le Département de l'agriculture des É.-U. applique la méthode des moindres carrés dans son enquête sur les activités des exploitations agricoles.

Il semble juste d'affirmer que ces auteurs se sont intéressés surtout à des moyennes ou à des totaux pour des périodes précises. Autrement dit, ils n'ont pas étudié explicitement des paramètres longitudinaux comme la proportion d'individus appartenant à une classe particulière à la période 1 et à la période 2. Nous verrons toutefois que la méthode des moindres carrés s'applique à des paramètres de ce genre.

Une caractéristique intéressante de la méthode des moindres carrés linéaires est que les estimateurs relatifs à un certain nombre de caractères ont la propriété d'additivité, c'est-à-dire que la somme de l'estimateur par les moindres carrés de Y et de l'estimateur par les moindres carrés de Z est égale à l'estimateur par les moindres carrés de $Y + Z$. Toutefois, si l'on se sert d'autres vecteurs d'observations pour construire des estimateurs, la propriété d'additivité disparaît.

Dans beaucoup d'enquêtes, on ne peut calculer les estimateurs par moindres carrés optimaux pour toutes les périodes parce qu'on ne peut se servir de toute l'information disponible pour l'estimation. Premièrement, il n'est pas possible d'intégrer toutes les données des enquêtes des périodes antérieures à une analyse par les moindres carrés pour la période courante car le nombre de variables dépassera souvent le nombre d'observations. Deuxièmement, l'organisme qui publie les données peut être tenu de respecter un plafond en ce qui concerne le nombre de fois où il est permis de réviser des estimations antérieures. Smith et Holt (1989) se sont penchés sur ce dernier point.

Afin d'illustrer ces problèmes d'estimation, nous avons voulu utiliser un exemple simple. À cette fin, le tableau 1 représente un tableau de contingence qui montre la division de la même variable en deux classes pour deux périodes données, et dont les observations reposent sur un très grand échantillon. Nous avons identifié les classes de ce tableau en désignant la première comme les personnes avec emploi et la seconde comme les chômeurs. Nous supposons que la population ne varie pas d'une année à l'autre. Si nous devions considérer les naissances et les décès, il nous faudrait alors un tableau 3 × 3. Supposons que nous voulons estimer la

Tableau 1

Proportions hypothétiques pour deux périodes données

PÉRIODE 1	PÉRIODE 2		
	Personnes avec emploi	Chômeurs	Total
Personnes avec emploi	0.91	0.02	0.93
Chômeurs	0.03	0.04	0.07
Total	0.94	0.06	1.00

variation de niveau d'une période à l'autre. Supposons aussi que nous voulons dresser un tableau des variations brutes; cette opération suppose l'estimation des fréquences par case du tableau de contingence. Pour un tableau 2 × 2, il suffit d'estimer la fréquence de la case (1, 1) et les proportions marginales pour obtenir les fréquences des autres cases.

Notre analyse porte sur deux périodes, pour lesquelles le même nombre d'éléments sont observés. Nous supposons que la moitié des éléments observés à la première période le sont aussi à la seconde. Autrement dit, les éléments observés à la seconde période se répartissent en deux groupes égaux: un groupe formé d'éléments observés à la première période et l'autre formé de nouveaux éléments. Le vecteur des observations est composé de la proportion d'éléments de la classe 1 qui font partie de la moitié de l'échantillon observée uniquement à la première période [désignée par $P_{E\cdot 1}$], de la proportion d'éléments de la classe 1 qui font partie de l'autre moitié de l'échantillon de la période 1 [désignée par $P_{E\cdot 2}$], de la proportion d'éléments de la classe 1 aux deux périodes, qui font partie de la moitié d'échantillon observée aux deux périodes [désignée par P_{EE}], de la proportion d'éléments de la classe 1 à la période 2 parmi les éléments observés aux deux périodes [désignée par $P_{\cdot E2}$] et de la proportion d'éléments de la classe 1 à la période 2 parmi les éléments qui n'ont été observés qu'à la période 2 [désignée par $P_{\cdot E3}$].

Nous supposons un échantillonnage aléatoire simple. Comme les statistiques consistent en des proportions d'échantillon, on peut reproduire facilement la matrice des covariances du vecteur de cinq estimateurs. Un multiple de cette matrice est représenté par le tableau 2. Pour obtenir la matrice des covariances pour un échantillon de taille n à chaque période, il suffit de diviser chaque élément de la matrice du tableau 2 par n, puis de multiplier le résultat par deux. Le tableau 3 donne les variances obtenues avec diverses méthodes d'estimation. La première colonne contient les variances obtenues avec la méthode qui n'utilise que les éléments

Tableau 2

Matrice des covariances du vecteur des proportions d'échantillon, deux périodes et échantillons se chevauchant dans une proportion de 50%
(Pour un échantillon de taille n, multiplier chaque élément du tableau par 2, puis diviser par n)

$P_{E\cdot 1}$	$P_{E\cdot 2}$	P_{EE}	$P_{\cdot E2}$	$P_{\cdot E3}$
0.0651	0	0	0	0
0	0.0651	0.0637	0.0358	0
0	0.0637	0.0819	0.0546	0
0	0.0358	0.0546	0.0564	0
0	0	0	0	0.0564

Tableau 3

Variances obtenues avec diverses méthodes d'estimation (Pour un échantillon de taille n à chaque période, multiplier chaque élément du tableau par 2, puis diviser par n)

Paramètre	Méthode		
	Simple	MCG avec contrainte	MCG
$P_{E\cdot}$	0.0326	0.0326	0.0294
P_{EE}	0.0819	0.0397	0.0374
$P_{\cdot E}$	0.0278	0.0258	0.0255
$P_{EE}/P_{\cdot E}$	0.0290	0.0229	0.0220
$P_{\cdot E} - P_{E\cdot}$	0.0429	0.0367	0.0353

de l'échantillon de la première période pour estimer la proportion d'éléments de la classe 1 (personnes avec emploi) à cette période. Pour estimer la proportion d'éléments de la classe 1 aux deux périodes, la méthode ordinaire n'utilise que les éléments communs aux deux échantillons et pour estimer la proportion d'éléments de la classe 1 à la période 2, elle n'utilise que l'échantillon observé à la période 2. Par conséquent, si nous avons un échantillon de 200 éléments à chaque période, l'échantillon de la première période sert à estimer la proportion d'éléments de la classe 1 à cette période, les 100 éléments communs aux deux périodes servent à estimer la proportion d'éléments qui demeurent dans la classe 1 d'une période à l'autre et les 200 éléments observés à la période 2 servent à estimer la proportion d'éléments de la classe 1 pour cette période.

La dernière colonne du tableau 3 contient les variances du meilleur estimateur linéaire sans biais construit à l'aide des moindres carrés généralisés. Cet estimateur est construit à l'aide du vecteur des cinq estimateurs de base et de la matrice des covariances de ce vecteur. Sa formulation est la suivante:

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y, \tag{1}$$

où V est définie dans le tableau 2, $\beta = (P_{E\cdot}, P_{\cdot E}, P_{EE})$,

$$X' = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

et Y est le vecteur quinquidimensionnel des estimations directes,

$$Y' = (\tilde{P}_{E\cdot 1}, \tilde{P}_{E\cdot 2}, \tilde{P}_{EE}, \tilde{P}_{\cdot E2}, \tilde{P}_{\cdot E3}).$$

La seconde colonne du tableau 3 contient les variances de l'estimateur par les moindres carrés restreint, la contrainte étant que l'estimateur pour la période 1 doit être l'estimateur obtenu à l'aide de l'échantillon initial. Cette méthode conviendrait si les organismes statistiques ne révisaient jamais les estimations déjà publiées. Par exemple, le Bureau of Labour Statistics des États-Unis ne révise jamais les statistiques du chômage. Une fois publiées, ces statistiques tiennent lieu d'estimations officielles. Il est vrai toutefois que, par rapport à notre exemple, ces statistiques reposent sur un échantillon plus complexe et une enquête qui s'étend sur une plus longue période.

Pour décrire l'estimateur par les moindres carrés généralisés restreint du tableau 3, définissons le modèle

$$Y = X\beta + e,$$

où X est une matrice fixe $n \times k$ et

$$E\{ee'\} = V.$$

L'estimateur par les moindres carrés généralisés de β, dont certains éléments sont contraints à être des combinaisons linéaires de Y, peut être construit de la façon suivante. Considérons la fonction lagrangienne

$$(Y - X\beta)'V^{-1}(Y - X\beta) - 2\sum_{i=1}^{b} \lambda_i(\Gamma_i\beta - g_i),$$

où Γ_i est un vecteur ligne fixe et b est le nombre de contraintes. La solution à ce problème de minimisation est définie

$$\begin{pmatrix} X'V^{-1}X & \Gamma' \\ \Gamma & 0 \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \lambda \end{pmatrix} = \begin{pmatrix} X'V^{-1}Y \\ g \end{pmatrix},$$

où $\lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_b)$, $\Gamma' = (\Gamma_1', \Gamma_2', \ldots, \Gamma_b')$ et $g' = (g_1, g_2, \ldots, g_b)$. Si nous remplaçons g par la combinaison linéaire GY, l'équation ci-dessus devient

$$\begin{pmatrix} X'V^{-1}X & \Gamma' \\ \Gamma & 0 \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \lambda \end{pmatrix} = \begin{pmatrix} X'V^{-1} \\ G \end{pmatrix} Y.$$

Cette équation définit l'estimateur restreint de β comme une fonction linéaire de Y. Par conséquent, la variance de l'estimateur de β correspondra à la partie supérieure $k \times k$ de

$$\begin{pmatrix} X'V^{-1}X & \Gamma' \\ \Gamma & 0 \end{pmatrix}^{-1} \begin{pmatrix} X'V^{-1} \\ G \end{pmatrix} V \left(\begin{pmatrix} X'V^{-1}X & \Gamma' \\ \Gamma & 0 \end{pmatrix}^{-1} \begin{pmatrix} X'V^{-1} \\ G \end{pmatrix} \right)'.$$

Cette méthode n'est pas la seule façon de calculer l'estimateur par les moindres carrés généralisés restreint. Il y a aussi l'estimateur composite, qui est un autre estimateur de niveau et de variation qui ne modifie en rien l'estimateur précédent. Voir par exemple Wolter (1979).

Ce court exemple met en lumière plusieurs points. Premièrement, compte tenu d'une corrélation de 0.591 entre les niveaux d'emploi des deux périodes, l'application des moindres carrés généralisés améliore modérément (environ 10 %) l'estimation du niveau d'emploi pour la période courante. En revanche, la même méthode améliore sensiblement la variance de la valeur estimée de P_{EE}. En effet, cette variance équivaut à environ 45 % de la variance de l'estimateur simple. Le second point à retenir est que l'application des moindres carrés généralisés avec contrainte donne des estimations de P_{EE} et $P_{.E}$ qui sont presque aussi efficientes que celles obtenues par les moindres carrés généralisés sans contrainte. Dans le cas de $P_{.E}$, la perte d'efficience est d'environ 1% et dans le cas de P_{EE}, elle est d'environ 6%.

3. ESTIMATEURS LONGITUDINAUX

Nous avons défini plus haut l'enquête à échantillon constant comme une enquête où les mêmes éléments sont observés à chaque période de collecte des données. L'enquête à échantillon constant se prête bien à l'observation de certaines unités physiques, comme des parcelles de terrain. Par contre, en ce qui a trait à l'observation de populations humaines, l'enquête à échantillon constant n'est rien de plus qu'une vue de l'esprit. Dans la réalité, l'enquêteur perd toujours une partie du groupe de répondants entre deux passages d'une enquête. Lepkowski (1989) et Little et Su (1989) font une bonne analyse des méthodes de traitement de la non-réponse. Voir aussi à ce sujet Little et Rubin (1987), Kalton (1983) et Madow et coll. (1983).

Nous avons aussi défini l'enquête avec renouvellement de l'échantillon, où certains éléments de l'échantillon font place à de nouveaux éléments à chaque passage de l'enquête. Dans ce cas, nous pouvons dire qu'il existe une forme de planification de la non-réponse pour les éléments qui sont supprimés de l'échantillon. Il faut donc voir un lien entre l'estimation en situation de non-réponse et l'estimation dans les enquêtes avec renouvellement partiel de l'échantillon.

Comme il est difficilement concevable qu'un enquêteur obtienne des réponses de tous les membres de l'échantillon à chaque passage de l'enquête, il faut s'attendre à recourir à une méthode qui permettra de compenser la non-réponse (prévue ou non prévue). Il existe deux méthodes simples et courantes. Si l'intention première de l'enquêteur est de suivre l'évolution d'un groupe de personnes dans le temps, très souvent il ne considérera dans son échantillon que les personnes qui ont participé à tous les passages de l'enquête. Dans ces circonstances, il dispose d'une méthode de pondération par laquelle il peut redresser les données de l'enquête à l'aide des caractéristiques du groupe initial des répondants ou de données supplémentaires ou les deux. On procède souvent de cette façon dans les enquêtes spéciales portant sur une population spécifique. Dans ce cas, les résultats ne sont publiés qu'une fois l'enquête terminée.

La seconde méthode consiste à établir des estimations pour chaque période à l'aide des données dont on dispose pour la période en question. Cette méthode est souvent utilisée pour les enquêtes périodiques; les résultats sont publiés à la fin de chaque enquête et ne sont pas révisés par la suite et aucune estimation longitudinale n'est produite. Un des avantages de cette méthode est qu'il est très facile d'établir des estimations pour la période t puisqu'on ne se sert pas des données de la période précédente pour calculer les estimations des valeurs courantes. Avec cette méthode, on obtient habituellement des estimations acceptables (non optimales) des valeurs courantes mais des estimations de la variation qui laissent à désirer.

Par ailleurs, on peut utiliser les deux méthodes dans une même enquête. La Survey of Income and Program Participation (SIPP), réalisée par le U.S. Bureau of the Census, est une enquête par panel avec rotation du moment d'interview et une période de rappel de quatre mois. À chaque passage de l'enquête, le U.S. Bureau of the Census produit une série de poids qui peuvent servir à établir des estimations pour la période en question à l'aide des données fournies par l'ensemble des personnes qui ont participé à l'enquête à cette occasion. L'organisme américain produit aussi a) l'échantillon de personnes qui ont participé aux huit passages de l'enquête en 1984-1985 de même que les poids relatifs à ces personnes, b) l'échantillon de personnes qui ont participé aux quatre passages de l'enquête en 1984 de même que les poids correspondants et c) l'échantillon de personnes qui ont participé aux quatre passages en 1985 de même que les poids correspondants.

Nous allons maintenant décrire une méthode d'estimation pour une enquête par panel où il y a des cas de non-réponse et où l'analyse n'est réalisée qu'à la toute fin. Nous supposons qu'une proportion raisonnable des unités participent à tous les passages de l'enquête et que l'analyse longitudinale revêt de l'intérêt. Il s'agit ici de construire des poids pour les unités dont les enregistrements sont complets. Les données fournies par les répondants dont les enregistrements sont incomplets servent d'information supplémentaire.

La première étape de l'analyse consiste à choisir quelques variables qui ont une grande importance pour l'enquête. Le nombre de variables que l'on peut utiliser dépendra de la taille de l'échantillon. Dans un deuxième temps, on calcule la structure des covariances du vecteur des estimations, qui est composé des estimations simples calculées pour chacune des variables pour chaque genre de schéma de réponse et chaque période pertinente. La structure des covariances est une fonction du schéma de réponse et de non-réponse. Il existe plusieurs définitions des estimateurs simples. Dans le cas de l'échantillonnage aléatoire simple, les estimateurs simples sont des moyennes. Dans le cas de l'échantillonnage stratifié, le vecteur initial peut être défini de manière à inclure les estimations pour chaque strate. Par ailleurs, l'estimateur simple pourrait servir à pondérer les réponses dans chaque strate pour compenser la non-réponse. Le vecteur Y figurant dans l'équation (1) est justement un vecteur d'estimations simples.

Étant donné le vecteur d'estimations simples et la matrice des covariances estimée de ce vecteur, nous pouvons construire à l'aide des moindres carres généralisés de meilleurs estimateurs pour chacune des périodes. Par exemple, si nous avons une enquête par panel avec trois passages, il y a sept schémas de réponse possibles, soit XXX, $0XX$, $X0X$, $XX0$, $X00$, $0X0$, $00X$, où X signifie réponse et 0, non-réponse. Si nous choisissons deux variables d'intérêt, le vecteur des estimations simples contiendra $12 \times 2 = 24$ estimations puisqu'il y a 12 réponses - groupes associés aux sept schémas de réponse. Dans cet exemple, les moindres carrés généralisés serviraient à produire six estimations, soit les valeurs estimées des deux variables d'intérêt pour chacune des trois périodes.

L'estimateur par les moindres carrés généralisés des caractères choisis sert de variable de contrôle à une étape ultérieure. À l'aide de méthodes de régression, nous construisons des poids pour les personnes qui ont participé à tous les passages. Ces poids sont construits de telle manière que les estimations par les moindres carrés généralisés pour chaque période visée soient reproduites par l'échantillon pondéré des personnes qui ont participé aux trois passages. En d'autres termes, les valeurs estimées des variables choisies pour les diverses périodes servent de données de contrôle.

L'efficacité de cette méthode dépend de la corrélation entre les variables de contrôle choisies et la variable d'analyse. Si l'une des variables de contrôle est aussi la variable d'analyse, la méthode sera très efficace. La seule raison pour laquelle cette méthode n'est pas parfaitement efficace pour les variables de contrôle est que la méthode des moindres carrés généralisés n'utilise qu'une quantité limitée de données.

Son principal avantage est qu'elle produit une seule série de données de totalisation à partir de laquelle on peut construire des estimateurs ayant la propriété d'additivité pour toutes les périodes visées par l'enquête et tous les tableaux de variation brute. L'inconvénient est que les estimations pour des périodes précises ne sont pas parfaitement efficaces.

La variance de cette méthode peut être calculée de la même manière que celle de la méthode utilisée pour l'échantillonnage double. Soit Y le caractère d'intérêt. Pour plus de simplicité, nous allons supposer un échantillonnage aléatoire simple à chaque fois. Nous définissons comme suit le modèle servant à l'estimation:

$$Y_i = \mu_Y + (X_i - \mu_X)\theta + e_i$$

$$\mu_X = E\{X\},$$

$$e_i \sim \mathrm{Ind}(0, \sigma_e^2).$$

Soit $\hat{\mu}_X$ l'estimateur par les moindres carrés généralisés de μ_X. Alors, l'estimateur de la moyenne de Y s'écrit

$$\hat{\mu}_Y = \bar{y} + (\hat{\mu}_X - \bar{x})\hat{\theta},$$

où $\hat{\theta}$ est le vecteur des coefficients de régression que nous avons calculés en faisant la régression de Y_i par rapport à X_i à l'aide de la série d'enregistrements complets et $(\bar{y},\bar{x})$ est le vecteur de moyennes pour les éléments observés à tous les passages. Posons m comme le nombre de ces enregistrements. Alors, la variance de l'estimateur est approximativement

$$V\{\hat{\mu}_Y\} = m^{-1}\sigma_e^2 + \theta' V\{\hat{\mu}_X\}\theta,$$

où $V\{\hat{\mu}_X\}$ est la matrice des covariances de $\hat{\mu}_X$.

L'estimateur que nous venons de décrire sera efficace dans la plupart des circonstances. Cependant, il se peut qu'il produise des valeurs estimées négatives pour des quantités réputées non négatives parce qu'il est linéaire et que certains poids peuvent être négatifs. Des méthodes ont été mises au point pour corriger cette lacune. Voir Huang et Fuller (1978).

4. INVENTAIRE DES RESSOURCES NATIONALES DES É.-U.

Le Iowa State Statistical Laboratory collabore avec le U.S. Soil Conservation Service à la réalisation d'une enquête d'envergure sur l'utilisation du sol aux États-Unis. Des enquêtes ont déjà eu lieu en 1958, 1967, 1975, 1977, 1982 et 1987. On en prévoit une autre en 1992.

Cette enquête permet de recueillir des données sur la nature et l'utilisation du sol, le couvert végétal, la possibilité de transformer des terres qui ne servent pas actuellement à la culture en terres labourables, l'érosion hydrique et les méthodes de conservation. La collecte des données est confiée à des employés du U.S. Soil Conservation Service tandis que l'université Iowa State s'occupe de l'élaboration du plan de sondage et de l'estimation.

L'échantillon est un échantillon stratifié des terres non fédérales de 49 États (l'Alaska étant exclu) et de Porto Rico. Les unités d'échantillonnage sont des portions de terrain appelées segments. La superficie de ces segments varie de 40 à 640 acres. Des données sont recueillies

pour tout le segment en ce qui concerne des aspects comme le sol urbain et les plans d'eau. En revanche, des données détaillées sur la nature et l'utilisation du sol sont recueillies à certains endroits dans le segment, choisis aléatoirement. En règle générale, on compte trois points d'observation par segment; les segments de 40 acres n'en comptent que deux et les segments des échantillons de deux États n'en comptent qu'un. Certaines données, comme la superficie totale et la superficie représentée par les routes, sont recueillies au moyen d'un recensement qui n'a rien à voir avec l'enquête précitée.

En 1982, l'échantillon comprenait environ 350,000 segments et près d'un million de points d'observation. En 1987, il comptait environ 100,000 segments, dont la majeure partie provenait de l'échantillon de 1982. Néanmoins, environ 1,500 nouveaux segments, prélevés dans des régions à forte croissance urbaine, ont été inclus dans l'échantillon de 1987. Ainsi, celui-ci comptait environ 280,000 points d'observation.

L'enquête de 1987 a été la première où on a décidé de faire une analyse de données longitudinales; cette analyse allait porter sur la période 1982-1987. Par la même occasion, on a décidé que les données de l'enquête allaient désormais être mises à la disposition du Soil Conservation Service de chaque État pour qu'il puisse faire ses propres analyses.

En 1987, les membres du personnel sur le terrain se sont vu remettre une feuille de travail qui contenait les données des segments pour 1982. Ils devaient y inscrire les données pour 1987 en se fondant sur les résultats d'observations sur le terrain et de la photographie aérienne. Ils étaient autorisés à corriger les données de 1982 si celles-ci étaient inexactes. Des méthodes de contrôle et de vérification ont été appliquées durant la phase de traitement.

On a conçu l'échantillon de manière à obtenir des estimations acceptables pour des unités appelées "zones principales de sols" (Major Land Resource Areas - MLRA). Ces zones sont définies en fonction de la nature du sol et du couvert végétal. On en compte environ 180 sur le territoire visé par l'enquête. Par ailleurs, la superficie estimée pour chaque comté doit concorder avec la superficie totale du comté. L'échantillon de l'enquête comprend environ 3,100 comtés. Comme il doit y avoir concordance entre les estimations de superficie relatives aux comtés et celles relatives aux zones principales de sols, l'unité de totalisation fondamentale est la portion d'une zone principale de sols comprise dans les limites d'un comté. Ces unités de base sont au nombre de 5,530 et sont désignées par le sigle MLRAC.

Le plan de sondage équivaut à la forme la plus élémentaire d'une enquête par panel puisque l'échantillon de 1987 est à peu de choses près un sous-ensemble de l'échantillon de 1982. On a choisi d'utiliser comme variables de contrôle les superficies représentées en 1982 par 14 modes d'utilisation du sol parmi les plus importants (par ex.: terrain labourable, terrain de parcours, terrain forestier et terrain urbain). De plus, les données externes, comme la superficie représentée par les routes en 1987, et les données des segments, comme la superficie du terrain urbain en 1987, constituent de l'information supplémentaire au même titre que les données tirées des enregistrements incomplets.

Le tableau 4 est la version condensée d'un tableau d'estimations pour un des États visés par l'enquête. On n'y retrouve que 4 des 14 modes d'utilisation du sol considérés pour l'estimation. Les chiffres figurant dans la colonne de droite sont les estimations pour 1982. Les totaux des colonnes 3 (terrain urbain) et 4 (routes) sont tirés respectivement des données des segments et des sources externes. Le vecteur formé des quatre premiers totaux de la colonne de droite et des deux derniers totaux de la ligne du bas (superficie de terrain urbain en 1987 et superficie représentée par les routes en 1987) est un vecteur de totaux qui correspond au vecteur des moyennes estimées, $\hat{\mu}_X$, de la section 3.

Les autres chiffres du tableau sont essentiellement des estimations obtenues par les moindres carrés, qui satisfont les six totaux de contrôle. Au cours du processus d'estimation, on a dû parfois recourir à des méthodes d'imputation, par exemple lorsqu'un changement observé dans les données des segments ne se reflétait pas dans les données des points d'observation.

Le plan de l'enquête s'est traduit par des variances élevées pour les estimations directes de la variation de superficie dans le cas des modes d'utilisation du sol relativement moins

Tableau 4
Illustration de la méthode d'estimation

1982	1987				TOTAL
	Terrain labourable	Autres	Terrain urbain	Routes	
Terrain labourable	26,243	179	13	6	26,441
Autres	771	7,114	6	2	7,893
Terrain urbain	0	0	623	0	623
Routes	17	4	0	1,038	1,059
TOTAL pour 1987	27,031	7,297	642	1,046	36,016

importants comme le terrain urbain, le terrain labourable et les petites nappes d'eau. On s'est donc servi d'une méthode d'estimation pour petites régions pour établir des estimations de variation de superficie pour les zones principales de sols comprises dans les limites d'un comté (MLRAC). À cette fin, on a utilisé un programme d'ordinateur élaboré à l'Université Iowa State. Fuller (1986) expose la théorie qui est à la base de la méthode d'estimation pour petites régions. Le programme d'ordinateur a permis d'établir des estimations de la variation de superficie pour cinq modes d'utilisation du sol secondaires pour chacune des 5,500 MLRAC. Il s'agit là essentiellement d'une opération de répartition en ce sens que la somme des estimations pour les MLRAC équivaut à l'estimation globale pour l'État. Par ailleurs, on a établi des estimations pour les éléments du tableau 4 (élargi à 14 modes d'utilisation du sol) pour chaque MLRAC.

À cette occasion, les estimations régionales pour les MLRAC, les estimations de la superficie représentée par les routes et les estimations de la superficie totale de terrain labourable pour l'État ont servi de totaux de contrôle. Le processus d'estimation s'est terminé par la pondération des données des points d'observation de manière à obtenir les estimations du tableau 4 pour chaque MLRAC.

En résumé, le processus d'estimation aboutit à une série de données de totalisation qui se rapportent à des points d'observation et qui permettent d'estimer tous les éléments d'un tableau à double entrée décrivant, pour n'importe quelle région identifiable, l'évolution de la superficie associée à divers modes d'utilisation du sol pour la période 1982-1987. Les estimations ainsi obtenues concordent avec les estimations établies antérieurement à l'échelle de l'État pour les principaux modes d'utilisation du sol et concordent aussi avec les données provenant de sources autres que les points d'observation.

En règle générale, l'échantillon de totalisation ne produit pas de bonnes estimations de la variance même si les segments et les strates sont bien identifiés dans la série de données. À cause du contrôle exercé sur l'échantillon de 1982, qui était plus grand, les données des points d'observation relatives aux principaux modes d'utilisation du sol, comme le terrain labourable, produiront des estimations de la variance trop élevées. Pour obtenir des estimations justes, il faudrait recourir aux formules de l'échantillonnage double.

5. ERREUR DE MESURE

L'erreur de mesure peut avoir une incidence notable sur l'analyse des données dans le temps. Cette incidence peut être modérée dans le cas de moyennes observées périodiquement mais peut aussi être très appréciable dans le cas de l'estimation de la variation brute ou de l'estimation par régression.

Pour montrer jusqu'à quel point l'erreur de mesure peut biaiser les estimateurs de la variation brute, nous allons reprendre l'exemple du tableau 1. Chua et Fuller (1987) montrent que les proportions estimées qui figurent dans les cases du tableau à double entrée seront fortement biaisées si les données sont recueillies au moyen d'une méthode comme celle qu'utilise le U.S. Bureau of the Census. Voir aussi à ce sujet Abowd et Zellner (1985), Poterba et Summers (1985) et Singh et Rao (1990). Le modèle de Chua-Fuller suppose que les erreurs de réponse aux deux périodes sont indépendantes. Il suppose aussi que, pour chaque période,

$$P\{\text{réponse} = E|\text{situation réelle} = E\} = 1 - \alpha + \alpha P_E,$$
$$P\{\text{réponse} = U|\text{situation réelle} = E\} = \alpha P_U,$$
$$P\{\text{réponse} = U|\text{situation réelle} = U\} = 1 - \alpha + \alpha P_U,$$
$$P\{\text{réponse} = E|\text{situation réelle} = U\} = \alpha P_E,$$

où α est le paramètre du mécanisme de réponse. Selon ce modèle, l'espérance de la proportion d'employés à n'importe quelle période est égale à la proportion réelle. Un estimateur convergent de P_{EE} selon le modèle de Chua-Fuller est

$$\hat{\pi}_{EE} = (1 - \alpha)^{-2}\{\hat{P}_{EE} - \hat{P}_{E\cdot}\hat{P}_{\cdot E}[1 - (1 - \alpha)^2]\},$$

où $\hat{P}_{EE}$, $\hat{P}_{E\cdot}$ et $\hat{P}_{\cdot E}$ sont les estimateurs directs et α est un paramètre du mécanisme de réponse. Voir aussi Battese et Fuller (1973). Compte tenu des résultats de la réinterview, une valeur $\alpha = 0.10$ n'est pas exagérée. Par conséquent, nous avons

$$\pi_{EE} = (0.90)^{-2}\{0.91 - 0.93(0.94)(0.19)\}$$
$$= 0.9184.$$

Le tableau des proportions corrigées en fonction de l'erreur de réponse est

$$\begin{pmatrix} 0.9184 & 0.0116 \\ 0.0216 & 0.0484 \end{pmatrix}.$$

Dans cet exemple, le biais de l'estimateur direct de P_{EE} est 0.0084. Chua et Fuller estiment le même biais à 0.0168 dans le tableau de contingence qui comprend aussi les inactifs. Dans le tableau 5, nous comparons des méthodes d'estimation pour P_{EE}. Nous supposons un échantillon de 10,000 unités. Les trois méthodes de la portion gauche du tableau sont celles

Tableau 5

Erreur quadratique moyenne de divers estimateurs pour un échantillon de 10,000 unités à chaque période et un taux de chevauchement de 50%
(erreur quadratique moyenne de l'estimateur par les MCG redressé en fonction de l'erreur de mesure = 100)

Paramètre	Méthode d'estimation					
	Ordinaire			Erreur de mesure		
	Simple	MCG avec contrainte	MCG	Simple	MCG avec contrainte	MCG
$P_{E\cdot}$	111	111	100	111	111	100
$P_{\cdot E}$	111	101	100	111	101	100
P_{EE}	1071	967	961	250	106	100

du tableau 3. Les trois autres représentent les estimateurs redressés en fonction du biais dû à l'erreur de mesure. Dans le calcul de la variance, on suppose que α a une erreur type de 0.01. La correction effectuée pour tenir compte du biais dû à l'erreur de mesure ne modifie pas les estimateurs de $P_{E\cdot}$ et $P_{\cdot E}$. Dans cet exemple, le carré du biais de l'estimateur ordinaire de P_{EE} équivaut à environ neuf fois la variance de l'estimateur par les moindres carrés généralisés. Le biais dû à l'erreur de mesure explique donc en très grande partie l'erreur quadratique moyenne de l'estimateur de P_{EE}.

Ces résultats ont des conséquences importantes pour ce qui a trait à la conception du plan de sondage. Afin d'illustrer cela, revenons au problème de la variation brute. Supposons que nous voulions estimer la probabilité qu'une personne appartienne à la catégorie des personnes avec emploi pendant deux périodes données, P_{EE}. Nous supposons qu'il est possible de réaliser des réinterviews indépendantes aux deux périodes et que les interviews réalisées à deux périodes quelconques sont indépendantes. Nous supposons aussi qu'il n'y a que deux scénarios d'interview possibles:

A. Interview et réinterview à l'une des deux périodes.

B. Interview à la période 1 et interview à la période 2.

Enfin, nous supposons que l'erreur de réponse est non biaisée et qu'un modèle à deux catégories (personnes avec emploi et chômeurs) est approprié dans les circonstances. Nous supposons aussi que la probabilité qu'une réponse soit exacte dépend uniquement de la catégorie à laquelle appartient le répondant dans la période courante. Soient les probabilités de réponse définies en fonction de α et soit

$$\gamma = (1 - \alpha)^{-2}.$$

Désignons par θ_{ij} l'élément ij de la matrice de probabilités 2×2 observée par suite de la réinterview. Par conséquent, θ_{ij} est la probabilité qu'une personne réponde i à l'interview et j à la réinterview. Pour ce modèle simple, il existe des formules explicites pour les estimateurs. Ainsi,

$$\hat{\gamma} = \left(\hat{\theta}_{11} - \hat{\theta}_{1}^{2}\right)^{-1}\left(\hat{\theta}_{1} - \hat{\theta}_{1}^{2}\right)$$

et

$$\hat{P}_{11} = \hat{\gamma}(\tilde{P}_{11} - \tilde{P}_{1\cdot}\tilde{P}_{\cdot 1}) + \tilde{P}_{1\cdot}\tilde{P}_{\cdot 1}$$

où

$$\theta_1 = \theta_{11} + \theta_{12} = \theta_{11} + \theta_{21},$$

$\hat{\theta}_{ij}$ sont les estimations découlant de la réinterview et $\tilde{P}_{ij}$, les estimations découlant des interviews réalisés aux deux périodes.

Dans la construction de l'estimateur, les résultats de la réinterview ne servent qu'à estimer le paramètre de l'erreur de mesure. En fait, ces résultats pourraient être utilisés dans une méthode des moindres carrés généralisés pour améliorer les valeurs estimées de P_{11}, $P_{1\cdot}$ et $P_{\cdot 1}$. En supposant qu'il en coûte la même chose pour chaque interview, nous pouvons montrer qu'environ le quart des ressources devraient être consacrées à la réinterview. Le tableau 6 indique l'efficacité des estimateurs redressés en fonction de l'erreur de mesure par rapport aux estimateurs directs biaisés.

Tableau 6

Efficacité relative des estimateurs redressés en fonction de l'erreur de mesure

	Taille de l'échantillon (n)			
	500	1,000	5,000	10,000
EQM (direct)/EQM (erreur de mesure)	0.87	1.13	3.22	5.84

Pour de petits échantillons, l'erreur quadratique moyenne est moins élevée dans le cas de l'estimateur direct parce que la variance de cet estimateur est moins élevée. Rappelons-nous que seulement les trois quarts des observations nous renseignent sur $P_{EE} = P_{11}$. Toutefois, pour des échantillons dont la taille est supérieure à 750, le carré du biais de l'estimateur direct représente la très grande partie de l'erreur quadratique moyenne de cet estimateur, laquelle est plus élevée que celle de l'estimateur redressé en fonction de l'erreur de mesure. Ce court exemple illustre bien l'efficacité des plans de sondage qui permettent d'estimer les paramètres du processus d'erreur de mesure.

6. RÉSUMÉ ET CONCLUSIONS

Dans cet article, nous avons fait un survol des questions qui ont trait à l'analyse des données d'enquêtes à passages répétés. Nous avons vu qu'il est possible d'accroître sensiblement l'efficience d'estimations à l'aide de méthodes fondées sur les moindres carrés. Cependant, la complexité des échantillons et les délais fixés font qu'on ne peut tenir compte de toutes les données disponibles dans la construction des estimateurs par les moindres carrés. Le statisticien doit donc, en pratique, se limiter à un sous-ensemble de variables pour la construction des poids pertinents. En deuxième lieu, nous avons étudié la méthode d'estimation utilisée dans une enquête à deux passages réalisée par le U.S. Soil Conservation Service.

Nous avons montré que l'erreur de mesure pouvait introduire un biais appréciable dans des estimations longitudinales comme les estimations de variations brutes. Nous avons vu qu'il était possible de construire des estimateurs convergents à l'aide de méthodes fondées sur l'erreur de mesure. En outre, il est justifié de consacrer le quart des ressources disponibles à l'estimation de la variance de l'erreur de mesure dans le but explicite d'utiliser des méthodes d'estimation fondées sur la mesure de ce genre d'erreur.

REMERCIEMENTS

Cette étude a été rendue possible en partie grâce à un contrat de coopération (n° 68-3A75-8-12) passé avec le Soil Conservation Service du Département de l'agriculture des É.-U. L'auteur remercie Margot Tollefson pour les calculs.

BIBLIOGRAPHIE

ABOWD, J.M., et ZELLNER, A. (1985). Estimating gross labor force flows. *Journal of Business and Economic Statistics*, 3, 254-283.

BATTESE, G.E., et FULLER, W.A. (1973). An unbiased response model for analysis of categorical data. *Proceedings of the Section on Social Statistics, American Statistical Association*, 202-207.

BATTESE, G.E., HASABELNABY, N.A., et FULLER, W.A. (1989). Estimation des cheptels à l'aide de plusieurs estimateurs à base de sondage aréolaire et à base de sondage multiple. *Techniques d'enquête*, 15, 15-29.

CHUA, T.C., et FULLER, W.A. (1987). A model for multinomial response error applied to labor flows. *Journal of the American Statistical Association*, 82, 46-51.

COCHRAN, W.G. (1942). Sampling theory when the sampling units are of unequal sizes. *Journal of the American Statistical Association*, 37, 199-212.

DUNCAN, G.J., et KALTON, G. (1987). Issues of design and analysis of surveys across time. *International Statistical Review*, 55, 97-117.

ECKLER, A.R. (1955). Rotation sampling. *The Annals of Mathematical Statistics*, 26, 664-685.

FULLER, W.A. (1986). Small area estimation as a measurement error problem. *Proceedings of the Conference on Survey Research Methods in Agriculture*,(éd. D. Faulkenberry), American Statistical Association and NASS, U.S. Department of Agriculture, Washington, D.C.

GARCIA, P.A., BATTESE, G.E., et BREWER, W.D. (1975). Longitudinal study of age and cohort influences on dietary patterns. *Journal of Gerontology*, 30, 349-356.

GRAHAM, J.E. (1973). Composite estimation in two cycle rotation sampling designs. *Communications in Statistics*, 1, 419-431.

GURNEY, M., et DALY, J.F. (1965). A multivariate approach to estimation in periodic sample surveys. *Proceedings of the Section on Social Statistics, American Statistical Association*, 242-257.

HUANG, E.T., et FULLER, W.A. (1978). Nonnegative regression estimation for sample survey data. *Proceedings of the Section on Social Statistics, American Statistical Association*, 300-303.

JESSEN, R.J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agricultural Experiment Station Research Bulletin*, 304, 54-59.

JONES, R.G. (1980). Best linear unbiased estimators for repeated surveys. *Journal of the Royal Statistical Society*, séries B, 42, 221-226.

KALTON, G. (1983). *Compensating for Missing Survey Data*. University of Michigan, Survey Research Center.

KASPRZYK, D., DUNCAN, G.J., KALTON, G., et SINGH, M.P. (1989). *Panel Surveys*. New York: John Wiley.

KASPRZYK, D., et McMILLEN, D.B. (1987). SIPP: Characteristics of the 1984 Panel. *Proceedings of the Section on Social Statistics, American Statistical Association*, 181-186.

LAZARSFELD, P.F., et FISKE, M. (1938). The panel as a new tool for measuring opinion. *Public Opinion Quarterly*, 2, 596-612.

LEPKOWSKI, J.M. (1989). Treatment of wave nonresponse in panel surveys. Dans *Panel Surveys* (éds. D. Kasprzyk, G. Duncan, G. Kalton et M.P. Singh). New York: John Wiley.

LITTLE, R.J.A., et RUBIN, D.B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley.

LITTLE, R.J.A., et SU, H.L. (1989). Item Nonresponse. Dans *Panel Surveys* (éds. D. Kasprzyk, G. Duncan, G. Kalton et M.P. Singh). New York: John Wiley.

MADOW, W.G., OLKIN, I., NISSELSON, H., et RUBIN, D.B. (1983). *Incomplete Data in Sample Surveys*. (Trois volumes) New York: Academic Press.

PATTERSON, H.D. (1950). Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society*, séries B, 12, 241-255.

POTERBA, J.M., et SUMMERS, L.H. (1985). Adjusting the gross change data: Implications for Labor Market Dynamics. *Proceedings of the Conference on Gross Flows in Labor Force Statistics*, U.S. Bureau of the Census et U.S. Bureau of Labor Statistics, 81-95.

RAO, J.N.K., et GRAHAM, J.E. (1964). Rotation designs for sampling on repeated occasions. *Journal of the American Statistical Association*, 59, 492-509.

SINGH, A.C., et RAO, J.N.K. (1990). Adjustments for classification error in gross flows. Document non publié, Statistique Canada, Ottawa, Canada.

SMITH, T.M.F., et HOLT, D. (1989). Some inferential problems in the analysis of surveys over time. Communication présentée à la 47ième session de l'Institut Internationale de Statistique, Paris.

WOLTER, K. (1979). Composite estimation in finite populations. *Journal of the American Statistical Association*, 74, 604-613.