

## Une comparaison d'estimateurs d'un ensemble de totaux de population

DON ROYCE<sup>1</sup>

### RÉSUMÉ

Les données du Programme des estimations démographiques (PED) de Statistique Canada sont toujours étalonnées en fonction du recensement le plus récent sans qu'il soit tenu compte de l'erreur de couverture dans ce recensement. Or, depuis qu'ils ont constaté une forte hausse du niveau de sous-dénombrement lors du recensement de 1986, Statistique Canada étudie la possibilité de redresser l'effectif de la population de base du PED pour tenir compte du sous-dénombrement net. Dans cet article, on définit et compare quatre estimateurs de l'effectif de la population de base: les chiffres non redressés du recensement, les chiffres redressés, un estimateur de test préliminaire et un estimateur composite. L'élément de comparaison, en l'occurrence, est une généralisation des fonctions de risque proposées antérieurement connue sous le nom d'erreur quadratique moyenne pondérée (EQMP). L'EQMP s'applique non seulement aux totaux de population, mais aux fonctions de totaux de population, comme les proportions de population et les taux de croissance d'un recensement à l'autre. Il est aussi question de l'utilisation de l'EQMP dans l'élaboration et l'évaluation des estimateurs pour petites régions dans le contexte du redressement des chiffres du recensement.

MOTS CLÉS: Ajustement du recensement; sous-dénombrement; estimation pour les petites régions.

### 1. INTRODUCTION

Le Programme des estimations démographiques (PED) de Statistique Canada fournit une multitude de données détaillées sur les caractéristiques et la répartition de la population canadienne pour la période quinquennale qui sépare deux recensements. Les estimations intercensitaires de la population servent à de nombreuses opérations importantes comme le calcul des paiements de transfert – qui se chiffrent à plusieurs milliards de dollars – entre l'administration fédérale et les administrations provinciales, l'estimation de variables démographiques importantes comme le taux de natalité et le taux de mortalité, la détermination des niveaux d'immigration futurs et la pondération dans les enquêtes démographiques courantes comme l'enquête mensuelle sur la population active.

Le PED repose depuis toujours sur les données du recensement le plus récent sans que l'erreur de couverture n'ait jamais été prise en considération. En 1986 toutefois, on a observé une hausse appréciable du niveau de sous-dénombrement par rapport aux recensements précédents et ce niveau était toujours aussi variable d'une région à l'autre et d'un groupe démographique à l'autre. Cela a perturbé le PED et les nombreux autres programmes qui utilisent des estimations démographiques. C'est pourquoi on a mis sur pied au début de 1989 un projet qui visait à discuter de l'opportunité de redresser les estimations démographiques postcensitaires en fonction de l'erreur de couverture estimée dans le recensement et de la manière d'effectuer ce redressement. La recherche décrite dans cet article s'inscrivait dans le cadre de ce projet. Pour une description plus générale du projet, voir Royce (1992).

Il convient de souligner que ce redressement ne toucherait que les estimations démographiques. Les données du recensement de 1991 seront publiées sans qu'il y ait de redressement pour le sous-dénombrement, exception faite des redressements mineurs normalement effectués pour tenir compte du sous-dénombrement chez les résidents temporaires et chez les personnes dont

<sup>1</sup> Don Royce, chef de la Section de la qualité des données du recensement, Statistique Canada, Ottawa (Ontario), Canada K1A 0T6.

le logement a été classé par erreur parmi les logements inoccupés. Du point de vue technique toutefois, la question étudiée se rapproche sensiblement de la question du redressement des chiffres du recensement qui retient l'attention de nombreux organismes statistiques depuis quelques années.

Les deux questions fondamentales qui se posent relativement au redressement sont de savoir 1) dans quelle mesure un redressement améliore les chiffres du recensement et 2) quelles sont les meilleures méthodes de redressement. Dans cet article, nous comparons la précision de divers estimateurs d'une série de totaux de population en utilisant comme critère l'erreur quadratique moyenne pondérée (EQMP).

Dans la section 2, nous commençons par étudier le cas d'un total de population unique. Nous calculons puis comparons les erreurs quadratiques moyennes de quatre estimateurs possibles: estimateur fondé sur les chiffres non redressés du recensement, estimateur fondé sur les chiffres redressés du recensement, estimateur de test préliminaire et estimateur composite. Dans la section 3, nous étendons l'analyse à des totaux de population multiples et à des fonctions de totaux de population comme les proportions de population et les taux de croissance. La section 4 porte sur l'estimation pour petites régions; nous y étudions des méthodes d'estimation, et notamment l'estimation synthétique et un cas particulier de l'estimation synthétique appelé redressement général. Enfin, dans la section 5, nous proposons des sujets de recherche pour l'avenir.

L'élaboration des estimateurs présentés dans cet article repose sur deux hypothèses. Premièrement, les opérations de redressement doivent produire des estimations qui soient cohérentes entre les divers niveaux d'agrégation géographique et les divers groupes démographiques et cohérentes d'une période à l'autre. Aux yeux des utilisateurs, il est essentiel que la somme des parties égale le tout et qu'il n'y ait pas de bris de continuité majeure dans les séries d'estimations. Deuxièmement, le redressement doit reposer sur les résultats combinés des deux études de mesure de la couverture de Statistique Canada, à savoir la contre-vérification des dossiers, qui sert à mesurer le sous-dénombrement brut, et l'étude du surdénombrement, qui sert à mesurer le surdénombrement brut. Les deux études sont exposées à l'erreur d'échantillonnage et à l'erreur non due à l'échantillonnage.

## 2. TOTAL DE POPULATION UNIQUE

Nous allons tout d'abord définir et comparer quatre estimateurs pour un total de population unique, l'erreur quadratique moyenne (EQM) servant de critère dans la comparaison.

Soient:  $Y$  l'effectif recensé;

$T$  l'effectif réel (inconnu) de la population (à estimer);

$U$  le sous-dénombrement net réel, c.-à-d.  $U = T - Y$ ;

$\hat{U}$  une estimation de  $U$  tirée des études de mesure de la couverture;

$\sigma^2$  la variance de  $\hat{U}$ ;

$R$  le biais relatif de  $\hat{U}$ , c.-à-d.  $R = E(\hat{U})/U - 1$ .

Pour les quatre estimateurs, on peut exprimer la valeur estimée de  $T$  comme la somme de l'effectif recensé et de la valeur estimée de  $U$ . Par conséquent, l'EQM de la valeur estimée de  $T$  sera la même que celle de la valeur estimée de  $U$  correspondante. Les EQM (et EQMP dans les sections subséquentes) sont calculées pour des répétitions hypothétiques des études de mesure de la couverture, où les chiffres du recensement sont considérés comme des quantités fixes.

### 2.1 Estimateur fondé sur les chiffres non redressés du recensement

Dans ce cas particulier, la valeur estimée de  $U$  est nulle. Son biais est égal à  $-U$  et sa variance, égale à zéro. Par conséquent,  $EQM(\hat{U}^c) = U^2$ .

## 2.2 Estimateur fondé sur les chiffres redressés du recensement

Dans ce cas, l'estimateur de  $U$  est  $\hat{U}$ . Son biais est  $UR$  et sa variance,  $\sigma^2$ . Par conséquent,  $\text{EQM}(\hat{U}^A) = \sigma^2 + U^2R^2$ .

## 2.3 Estimateur de test préliminaire

En comparant les EQM des deux estimateurs précédents, on peut penser qu'il sera préférable d'utiliser les chiffres redressés, par opposition aux chiffres non redressés, dès que

$$\sigma^2 < U^2(1 - R^2). \quad (1)$$

Bien que les paramètres de cette inéquation soient inconnus, on peut les estimer (sauf  $R$ ) à l'aide des études de mesure de la couverture, de là la possibilité de se servir de ces estimations pour construire un test statistique qui permettra de vérifier l'hypothèse que l'inégalité est juste. Le résultat du test détermine le choix de l'estimateur (d'où le nom d'estimateur de test préliminaire).

De façon précise, supposons que  $|R| < 1$  (condition indispensable pour que (1) se vérifie) et que  $\hat{U} \sim N(U(1 + R), \sigma^2)$ , où  $\sigma^2$  est connu. Alors,  $\hat{U}^2/\sigma^2$  suit une distribution de  $\chi^2_{(1)}$  non centrale avec comme paramètre de non-centralité  $\lambda = U^2(1 + R)^2/2\sigma^2$ . L'hypothèse nulle  $H_0 : \sigma^2 \geq U^2(1 - R^2)$  équivaut à l'hypothèse  $H_0 : \lambda \leq (1 + R)/2(1 - R)$ . On pourrait donc, par exemple, procéder à un redressement si  $\hat{U}^2/\sigma^2 > c$ , la valeur critique  $c \geq 0$  étant choisie de manière que

$$\alpha = \text{Pr} \left\{ \chi^2_{(1, \frac{1+R}{2(1-R)})} \geq c \right\}, \quad (2)$$

où  $\alpha$  est le niveau de signification du test. Ce test est une version particulière d'un test plus général proposé par Toro-Vizcarrondo et Wallace (1968).

Notons que  $\hat{U}^2/\sigma^2$  est l'inverse du carré du coefficient de variation (CV) estimé de  $\hat{U}$ . Cela revient à dire que pour effectuer un redressement, il faut un CV suffisamment petit (en valeur absolue).

En pratique, il faudrait remplacer  $R$  dans (2) par une estimation provisoire du biais relatif, disons  $r$ . Royce (1991) étudie la sensibilité de  $c$  par rapport à  $R$  dans le cas d'un test unilatéral (on a utilisé en l'occurrence une distribution normale au lieu d'une distribution de  $\chi^2$ ). Par exemple, étant donné un niveau de signification de 2.5%, on a observé que la réduction du CV critique ne dépassait pas six points de pourcentage environ (de 33.8 à 27.1%), même lorsque le biais relatif atteignait 50%.

Si  $\sigma$  n'est pas connu mais qu'il existe une estimation  $\hat{\sigma}$ , on peut construire un test semblable en supposant que

$$\frac{\nu \hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(\nu)} \quad (3)$$

qui est indépendant de  $\hat{U}$ . On obtient alors un test fondé sur une distribution  $F$  non centrale. Pour obtenir plus de détails sur la construction de ces tests, le lecteur est prié de consulter Judge et Bock (1978).

Afin de déterminer l'EQM de l'estimateur de test préliminaire, notons que cet estimateur peut s'écrire  $\hat{U}^P = I\hat{U}$  où

$$I = 1 \quad \text{si} \quad \frac{\hat{U}^2}{\sigma^2} > c$$

$$= 0 \quad \text{si} \quad \frac{\hat{U}^2}{\sigma^2} \leq c. \quad (4)$$

Lorsque  $\sigma^2$  est connu, on peut montrer que l'EQM de cet estimateur est définie (voir, par exemple, Judge et Bock (1978), p. 72)

$$\text{EQM}(\hat{U}^P) = \sigma^2 + U^2 R^2 + (2U^2(1 + R) - \sigma^2) \Pr\{\chi^2_{(3,\lambda)} \leq c\} - U^2(1 + R)^2 \Pr\{\chi^2_{(5,\lambda)} \leq c\}. \quad (5)$$

Notons que lorsque  $c \rightarrow \infty$ , c.-à-d. que la probabilité d'un redressement est de plus en plus minime, l'EQM tend vers  $U^2$ , soit l'EQM de l'estimateur fondé sur les chiffres non redressés du recensement. De même, lorsque  $c \rightarrow 0$ , c.-à-d. que la probabilité d'un redressement est de plus en plus grande, l'EQM tend vers  $\sigma^2 + U^2 R^2$ , soit l'EQM de l'estimateur fondé sur les chiffres redressés du recensement. Par conséquent, les deux premières méthodes d'estimation présentées dans cette section peuvent être considérées comme des versions extrêmes de la troisième méthode.

La figure 1 renferme un graphique où la valeur  $\text{EQM}/\sigma^2$  pour l'estimateur de test préliminaire est exprimée en fonction de  $U^2/\sigma^2$  pour diverses valeurs de  $c$  dans le cas où il n'y a pas de biais ( $R = r = 0$ ). Les valeurs  $\text{EQM}/\sigma^2$  pour les estimateurs fondés sur les chiffres redressés et les chiffres non redressés du recensement sont aussi reproduites dans ce graphique. Pour chaque valeur de  $c$  proposée, la courbe de l'EQM de l'estimateur de test préliminaire débute plus haut sur l'axe des  $y$  par rapport à la courbe de l'EQM de l'estimateur fondé sur les chiffres non redressés, croise la courbe de l'EQM de l'estimateur fondé sur les chiffres redressés, atteint un maximum, puis se rapproche de la courbe de l'EQM de l'estimateur fondé sur les chiffres redressés. À mesure que  $c$  diminue et que, par voie de conséquence, le niveau de signification  $\alpha$  du test augmente, l'EQM de l'estimateur de test préliminaire se rapproche plus rapidement de celle de l'estimateur fondé sur les chiffres redressés; en revanche, elle est plus élevée pour de faibles valeurs de  $U^2/\sigma^2$ . Par conséquent, le rendement de l'estimateur de test préliminaire pour l'intervalle des valeurs de  $U^2/\sigma^2$  dépend du niveau de signification choisi pour le test.

Les figures 2 et 3 renferment des graphiques similaires pour les cas où  $R = .5$  et  $R = -.5$  respectivement (comme nous pourrions croire ne pas disposer de l'information voulue pour établir une valeur estimée de  $R$ , nous avons posé  $r = 0$ ). Là encore, l'EQM de l'estimateur de test préliminaire se rapproche de celle de l'estimateur fondé sur les chiffres redressés à mesure que  $U^2/\sigma^2$  augmente. Elle se rapproche plus rapidement si le biais est positif que s'il n'y a pas de biais mais moins rapidement si le biais est négatif.

Quelle est la "meilleure" valeur de  $c$  pour le test? Idéalement, il faudrait choisir la valeur de  $c$  de telle manière que l'EQM de l'estimateur de test préliminaire soit le plus comparable possible à la moins élevée des EQM de l'estimateur fondé sur les chiffres redressés et de l'estimateur fondé sur les chiffres non redressés. À cet égard, Sawa et Hiromatsu (1973) ont proposé une méthode, élargie par la suite par Brook (1976), qui consiste à minimiser l'écart maximum entre l'EQM de l'estimateur de test préliminaire et la moins élevée des EQM de l'estimateur fondé sur les chiffres redressés et de l'estimateur fondé sur les chiffres non redressés. Lorsqu'il n'y a pas de biais, cette méthode produit une valeur optimale de  $c$  d'environ 1.88. Cela correspond à un CV critique (en valeur absolue) de 73% pour le niveau de sous-dénombrement estimé. L'EQM de cet estimateur est reproduite dans la figure 4.

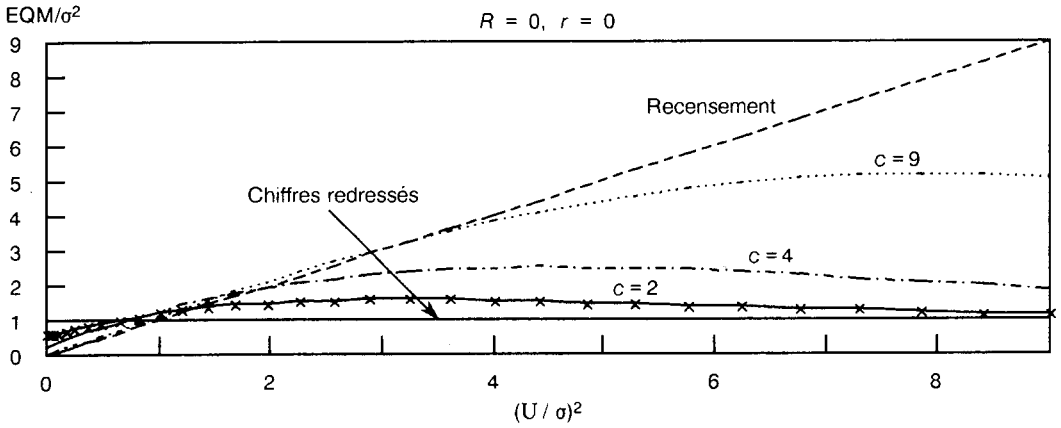


Figure 1 : Comparaison d'EQM

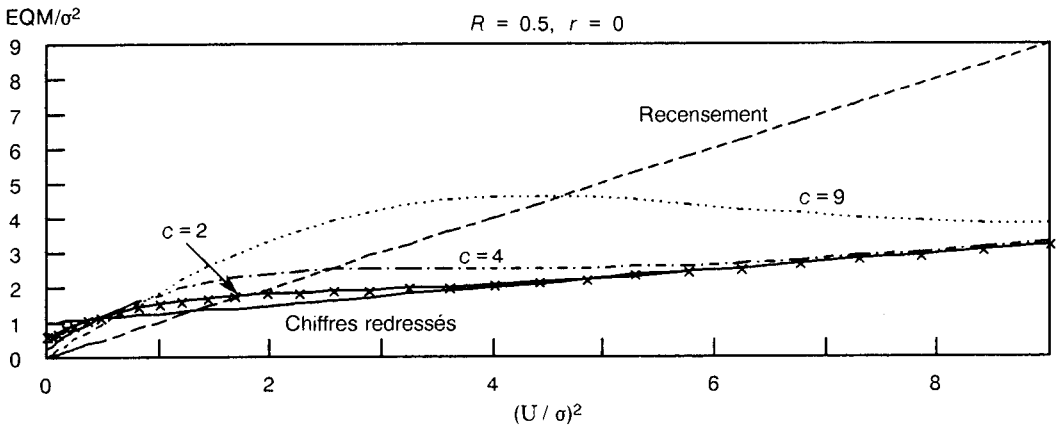


Figure 2 : Comparaison d'EQM

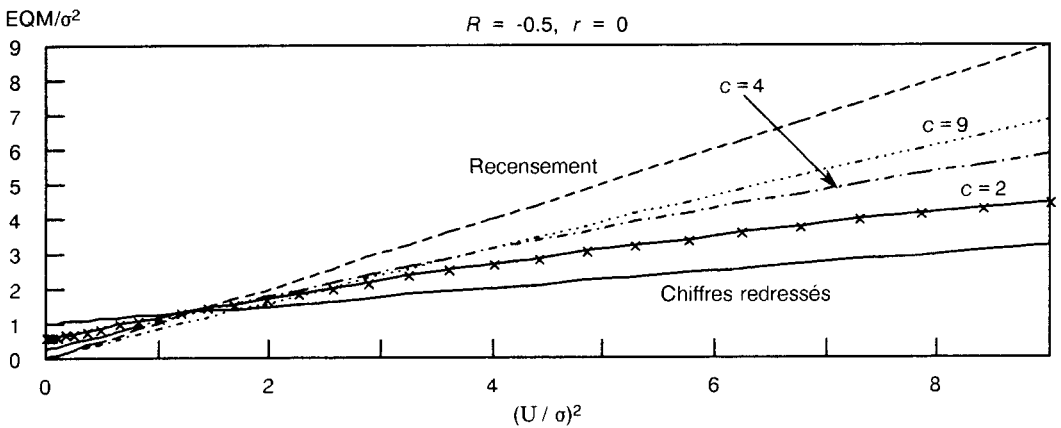


Figure 3 : Comparaison d'EQM

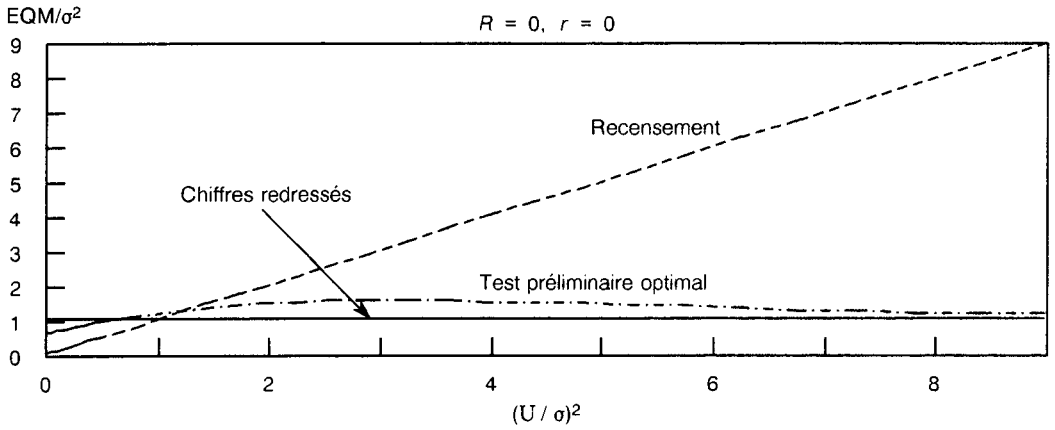


Figure 4 : Comparaison d'EQM

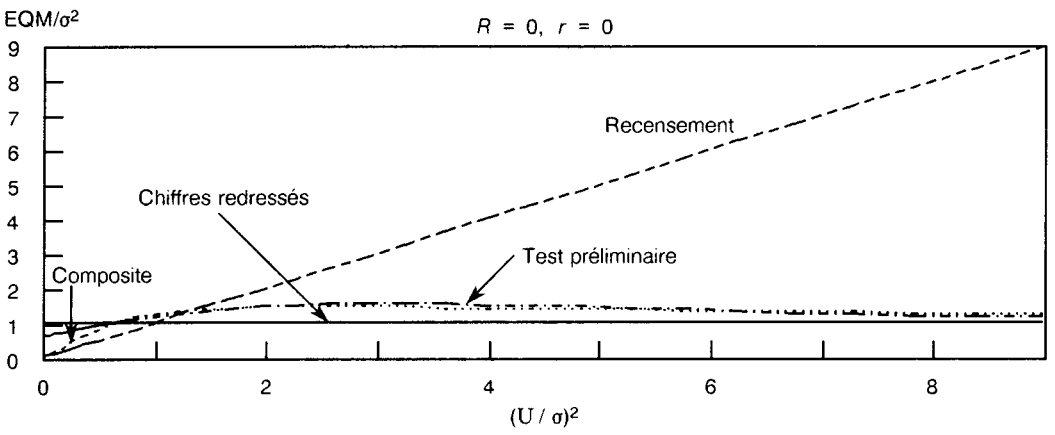


Figure 5 : Comparaison d'EQM

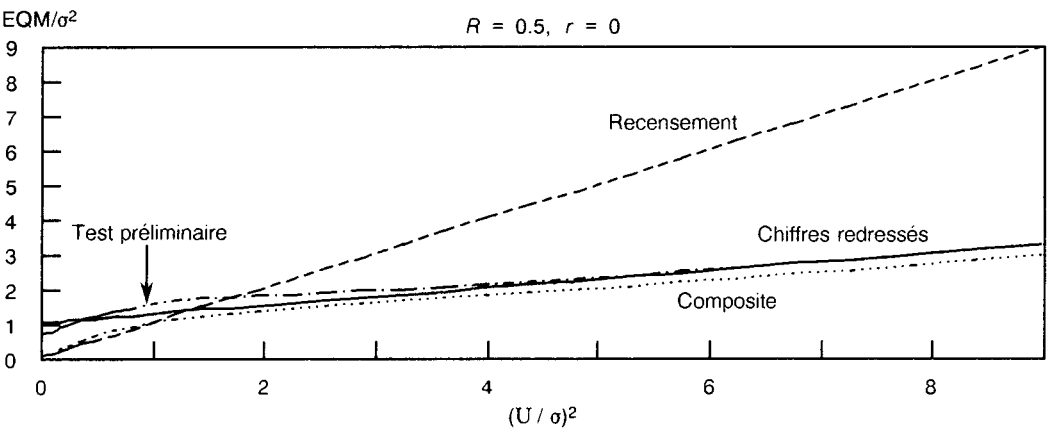


Figure 6 : Comparaison d'EQM

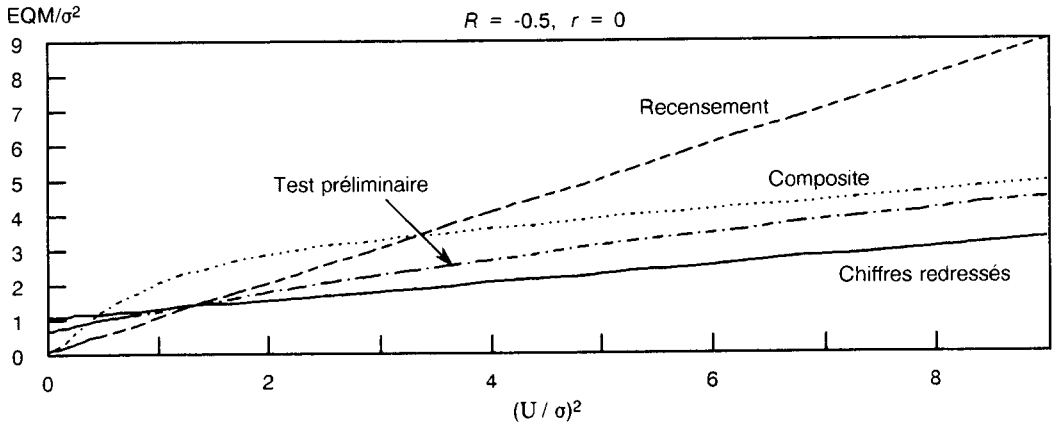


Figure 7 : Comparaison d'EQM

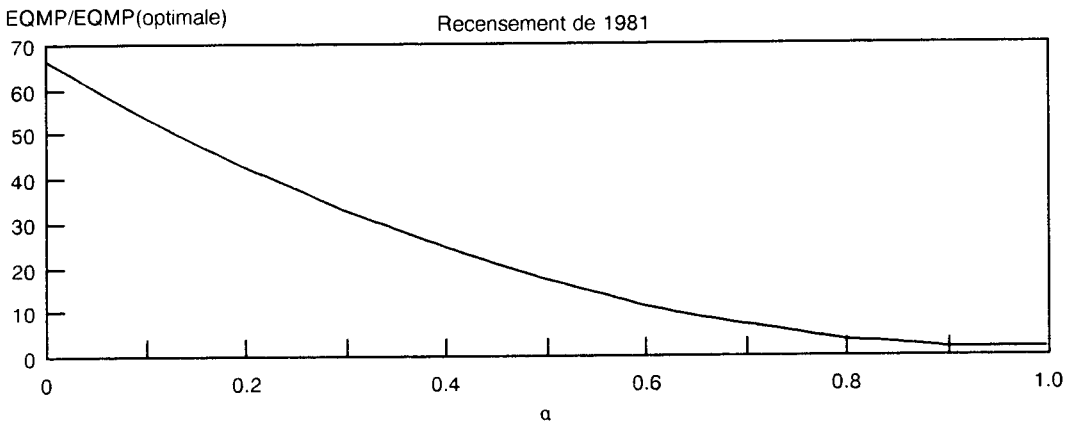


Figure 8 : EQMP pour totaux de population

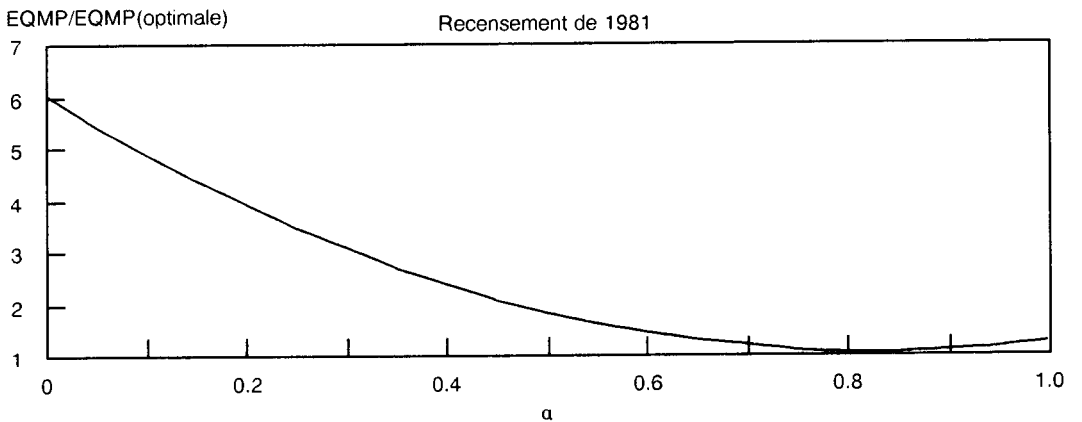


Figure 9 : EQMP pour proportions de population

Dans Judge et Bock (1978), on trouve d'autres méthodes pour choisir la valeur optimale de  $c$ , par exemple la minimisation de la distance moyenne (au lieu de l'écart maximum) et des méthodes bayésiennes.

## 2.4 Estimateur composite

Nous avons vu que l'estimateur de test préliminaire s'écrit  $\hat{U}^P = I\hat{U}$ , où  $I$  ne peut prendre que les valeurs 0 ou 1. Or, on a montré que cet estimateur est inacceptable (Cohen, 1965) à cause justement de cette discontinuité inhérente. Dans ces circonstances, le multiplicateur de  $\hat{U}$  pourrait peut-être prendre n'importe quelle valeur entre 0 et 1. Autrement dit, au lieu de nous servir des données pour savoir **s'il faut procéder ou non** à un redressement, nous nous en servons pour savoir quelle doit être la **mesure** du redressement. Spencer (1980) et, plus récemment, Andrews (1991) ont déjà parlé de ce type d'estimateur. Posons  $\hat{U}^\alpha = \alpha\hat{U}$  où  $0 \leq \alpha \leq 1$ . Pour une valeur  $\alpha$  donnée, l'estimateur ci-dessus a une EQM égale à

$$\text{EQM}(\alpha\hat{U}) = \alpha^2\sigma^2 + U^2(\alpha(1+R) - 1)^2, \quad (6)$$

laquelle expression est minimisée lorsque

$$\alpha = \frac{U^2(1+R)^2}{(1+R)(\sigma^2 + U^2(1+R)^2)}. \quad (7)$$

Si  $\sigma$  est supposé connu, un estimateur possible de  $\alpha$  est

$$\hat{\alpha} = \frac{\hat{U}^2}{(1+r)(\sigma^2 + \hat{U}^2)} \quad (8)$$

et par conséquent,

$$\hat{U}^{\hat{\alpha}} = \frac{\hat{U}^3}{(1+r)(\sigma^2 + \hat{U}^2)}. \quad (9)$$

On peut calculer l'EQM approximative de cet estimateur au moyen d'une approximation par série de Taylor. En posant

$$h(U, \sigma^2) = \frac{U^3}{(1+r)(\sigma^2 + U^2)} \quad (10)$$

nous obtenons (en supprimant les termes dont l'ordre est supérieur à celui des termes de dérivée première)

$$\begin{aligned} \text{EQM}(\hat{U}^{\hat{\alpha}}) &\doteq (h(U, \sigma^2) - U)^2 + \left(\frac{\partial h(U, \sigma^2)}{\partial U}\right)^2 (U^2 R^2 + \sigma^2) \\ &+ 2(h(U, \sigma^2) - U) \left(\frac{\partial h(U, \sigma^2)}{\partial U}\right) UR. \end{aligned} \quad (11)$$

On peut aussi utiliser cette approximation lorsque  $\sigma$  est inconnu en posant l'hypothèse formulée en (3). La formule de l'EQM comprend alors le terme additionnel:

$$\left(\frac{\partial h(U, \sigma^2)}{\partial \sigma^2}\right)^2 \frac{2\sigma^4}{\nu}. \quad (12)$$



Les figures 5, 6 et 7 contiennent un graphique où l'EQM de l'estimateur composite est exprimée en fonction de  $U^2/\sigma^2$ ; dans ces mêmes graphiques sont reproduites aussi les EQM de l'estimateur fondé sur les chiffres non redressés du recensement, de l'estimateur fondé sur les chiffres redressés et de l'estimateur de test préliminaire optimal de la section 2.3. Lorsqu'il n'y a pas de biais (figure 5) ou que le biais est positif (figure 6), l'estimateur composite est supérieur à l'estimateur de test préliminaire optimal. Par contre, lorsque le biais est négatif (figure 7), l'EQM de l'estimateur composite peut être beaucoup plus élevée que celle des autres estimateurs pour une portion considérable de l'intervalle de valeurs de  $U^2/\sigma^2$ .

### 3. ESTIMATEURS À CARACTÈRE PLUS GÉNÉRAL

Dans cette section, nous généralisons de deux manières les quatre estimateurs étudiés dans la section précédente. Premièrement, au lieu de ne considérer que des totaux de population uniques, nous allons étudier des vecteurs de totaux de population, désignés par  $\underline{T} = (T_1, T_2, \dots, T_N)$ . Deuxièmement, nous ne nous limitons pas aux totaux de population; nous considérons aussi les fonctions de ces totaux, désignées par  $\underline{g}(\underline{T}) = (g_1(\underline{T}), g_2(\underline{T}), \dots, g_K(\underline{T}))$ , où, de façon générale,  $K \neq N$ . Les principales fonctions de totaux de population comprennent les proportions de population, utilisées pour les transferts de fonds entre l'administration fédérale et les administrations provinciales, les taux de croissance entre les recensements, les différences de taux de croissance entre les provinces, et ainsi de suite.

En évaluant la précision globale d'une estimation  $\underline{g}(\hat{\underline{T}}^*)$  de  $\underline{g}(\underline{T})$ , nous allons recourir à une fonction de perte. L'utilisation des fonctions de perte dans le but d'évaluer les effets du redressement des chiffres du recensement est traitée dans Fellegi (1980), Citro et Cohen (1985), Spencer (1986) et Wolter et Causey (1991), pour n'en nommer que quelques-uns. La fonction de perte qui est utilisée ici est une généralisation des fonctions de perte proposées antérieurement pour des totaux et des proportions de population. De façon plus précise, le risque (perte moyenne espérée) de l'estimateur  $\underline{g}(\hat{\underline{T}}^*)$  est l'erreur quadratique moyenne pondérée (EQMP), qui est définie par l'expression

$$\text{EQMP}(\underline{g}(\hat{\underline{T}}^*)) = E \left\{ \sum_{k=1}^K w_k (g_k(\hat{\underline{T}}^*) - g_k(\underline{T}))^2 \right\}, \quad (13)$$

où  $w_k$  est un poids défini par l'utilisateur qui indique l'importance de la  $k$ -ième composante de la fonction de perte.

Comme  $\underline{g}$  peut être complexe dans la pratique, il est utile de disposer aussi d'une approximation de l'EQMP obtenue en développant  $\underline{g}(\hat{\underline{T}}^*)$  en une série de Taylor par rapport à  $\underline{T}$ . On a ainsi:

$$\text{EQMP} \underline{g}(\hat{\underline{T}}^*) \doteq \sum_{i=1}^N \sum_{j=1}^N \omega_{ij} [\text{Cov}(\hat{U}_i^*, \hat{U}_j^*) + \text{Biais}(\hat{U}_i^*) \text{Biais}(\hat{U}_j^*)] \quad (14)$$

où le poids  $\omega_{ij}$  est défini par l'expression

$$\omega_{ij} = \sum_{k=1}^K w_k \frac{\partial g_k}{\partial T_i} \frac{\partial g_k}{\partial T_j}. \quad (15)$$

(Notons que l'approximation de l'EQMP peut aussi être formulée comme l'espérance mathématique de la forme quadratique  $(\underline{\hat{U}}^* - \underline{U})' \Omega (\underline{\hat{U}}^* - \underline{U})$ , où  $\omega_{ij}$  est l'élément  $ij$  de  $\Omega$ .)

Cette formule a ceci d'utile qu'elle exprime chaque composante de la fonction de risque en deux parties: un poids,  $\omega_{ij}$ , qui dépend uniquement de  $w_k$  et de la fonction  $g$ , et l'expression entre crochets, qui dépend uniquement de l'estimateur utilisé.

Tandis que le choix de  $w_k$  peut être arbitraire, des considérations d'équité ont souvent amené les expérimentateurs à choisir  $w_k = 1/T_k$ . En ce qui concerne les totaux et les proportions de population par exemple, la fonction de risque devient alors l'équivalent des fonctions proposées par Fellegi (1980) et utilisées par Wolter et Causey (1991), pour ne nommer que ceux-ci. Les autres valeurs proposées pour  $w_k$  dans les ouvrages statistiques sont  $w_k = 1/Y_k$ ,  $w_k = 1/\hat{T}_k$ , et  $w_k = 1$ . Pour une analyse plus détaillée des avantages de ces divers poids, le lecteur est prié de consulter les ouvrages mentionnés plus haut. Le tableau 1 donne des exemples de poids  $\omega_{ij}$  pour diverses fonctions.

En ce qui concerne les taux de croissance démographique, la première paire d'indices désigne la population étudiée (par ex. : population d'une province), tandis que la seconde paire d'indices désigne le recensement de la période 1 et de la période 2 respectivement. Le second indice rattaché à  $T_i$  désigne aussi le recensement de la période 1 ou de la période 2.

Dans le reste de cette section, nous montrons comment l'EQMP sert à l'élaboration et à l'évaluation de nos quatre estimateurs: estimateur fondé sur les chiffres non redressés du recensement, estimateur fondé sur les chiffres redressés, estimateur de test préliminaire et estimateur composite.

**Tableau 1**

Exemples de poids  $\omega_{ij}$  dans la formule d'approximation de l'EQMP pour diverses fonctions

Fonction	$\omega_{ij}$
Série de totaux de population	$\omega_{ii} = w_i$ $\omega_{ij} = 0 \quad i \neq j$
Série de proportions de population	$\omega_{ii} = \frac{1}{T^4} \left( \sum_k w_k T_k^2 + w_i T^2 - 2w_i T T_i \right)$ $\omega_{ij} = \frac{1}{T^4} \left( \sum_k w_k T_k^2 - T(w_i T_i + w_j T_j) \right) \quad i \neq j$
Série de taux de croissance	$\omega_{i11} = \frac{w_i T_{i2}^2}{T_{i1}^4}$ $\omega_{i12} = - \frac{w_i T_{i1} T_{i2}}{T_{i1}^4} = \omega_{i21}$ $\omega_{i22} = \frac{w_i T_{i1}^2}{T_{i1}^4}$ $\omega_{ij11} = \omega_{ij12} = \omega_{ij21} = \omega_{ij22} = 0 \quad i \neq j$

### 3.1 Estimateur fondé sur les chiffres non redressés du recensement

L'EQMP de l'estimateur fondé sur les chiffres non redressés du recensement est  $EQMP(\hat{U}^C) = \sum_{ij} \omega_{ij} U_i U_j$ .

### 3.2 Estimateur fondé sur les chiffres redressés du recensement

L'EQMP de l'estimateur fondé sur les chiffres redressés est  $EQMP(\hat{U}^A) = \sum_{ij} \omega_{ij} [\sigma_{ij} + b_i b_j]$  où  $\sigma_{ij} = Cov(\hat{U}_i, \hat{U}_j)$  et  $b_i = Biais(\hat{U}_i)$ .

### 3.3 Estimateur de test préliminaire

Comme dans la section 2.3, nous préférons l'estimateur fondé sur les chiffres redressés à l'estimateur fondé sur les chiffres non redressés si l'EQMP du premier est moindre que celle du second, c.-à-d. si

$$D = \sum_{ij} \omega_{ij} [U_i U_j - \sigma_{ij} - b_i b_j] > 0. \quad (16)$$

Fellegi (1980) a proposé des tests pour ce genre d'hypothèses dans le cas particulier des totaux de population et des proportions de population; cependant, ces notions peuvent s'appliquer facilement à n'importe quelle fonction  $g$ . Le membre de gauche de l'inéquation (16) est estimé par  $\hat{D} = \sum_{ij} \omega_{ij} [\hat{U}_i \hat{U}_j - 2\sigma_{ij}]$  où les  $\omega_{ij}$  sont supposés connus. En pratique, on estime les  $\omega_{ij}$  en substituant dans l'équation (13) les chiffres bruts du recensement ou les chiffres redressés. Fellegi soutient qu'une faible variation des poids sera peu susceptible de modifier sensiblement les résultats des tests. On peut alors montrer facilement que  $E(\hat{D}) = D + 2 \sum_{ij} \omega_{ij} b_i (U_j + b_j)$ . En ce qui concerne les totaux et les proportions, Fellegi a présenté des arguments qui expliquent pourquoi on peut supposer que le second terme du membre de droite est non positif, c.-à-d.  $\sum_{ij} \omega_{ij} b_i (U_j + b_j) \leq 0$  de sorte que  $\hat{D}$  tendrait à sous-estimer  $D$ . Fellegi a aussi calculé une variance approximative pour  $\hat{D}$ . Grâce à cela et à l'hypothèse que  $\hat{D}$  est distribué normalement, on a pu construire un test pour l'hypothèse posée en (16).

Dans le cas plus général, la variance approximative de  $\hat{D}$  est définie  $Var(\hat{D}) \doteq 4 \sum_{ij} \omega_{ij} \sigma_{ij} (\sum_{i'j'} \omega_{i'j'} \omega_{ij} U_{i'} U_{j'})$ . On peut alors calculer une estimation de  $Var(\hat{D})$  en remplaçant  $U_i$  et  $\sigma_{ij}$  dans l'équation ci-dessus par leurs valeurs estimées.

Dans le cas des totaux par exemple, la variable à tester (valeur  $z$ ) est définie

$$z = \frac{\hat{D}}{\sqrt{\hat{Var}(\hat{D})}} = \frac{\sum_i \frac{\hat{U}_i^2 - 2\sigma_i^2}{Y_i}}{2\sqrt{\sum_i \frac{\hat{U}_i^2 \hat{\sigma}_i^2}{Y_i^2}}}, \quad (17)$$

où, en l'occurrence, l'inverse de l'effectif recensé sert de poids. Une expression semblable existe pour les proportions de population.

Le tableau 2 donne les valeurs  $z$  calculées pour les totaux et les proportions de population par province pour les recensements de 1976, 1981 et 1986. Les données sont tirées des contre-vérifications de dossiers effectuées lors de ces recensements.

D'après ce tableau, le redressement des totaux de population prime largement le redressement des proportions de population, ce qui dénote que les estimations de différences de taux de sous-dénombrement entre les provinces sont moins précises que les estimations de taux de sous-dénombrement. D'autres résultats numériques figurent dans Royce et Luc (1990).

Tableau 2

Valeurs  $z$  des tests de Fellegi pour le redressement des totaux et des proportions de population par province, Contre-vérification des dossiers, 1976, 1981 et 1986

Fonction	1976	1981	1986
Totaux	9.3	10.1	13.1
Proportions	3.1	1.8	1.5

### 3.4 Estimateur composite

À première vue, nous serions portés à croire que  $\alpha_i \hat{U}_i$  est une extension naturelle de l'estimateur composite de la section 2.4. Or, la variation du degré de redressement selon la valeur de  $i$  cause des problèmes de cohérence. Par exemple, nous pourrions devoir opérer un redressement plus profond au niveau national qu'au niveau provincial du fait que les estimations du sous-dénombrement pour les provinces sont moins précises que celles pour le Canada. Si nous devons faire cela, la somme des totaux provinciaux ne concorderait pas avec le total pour le Canada.

Par conséquent, nous nous limitons dans la pratique à une seule valeur de alpha, c.-à-d.  $\hat{U}^\alpha = \alpha \hat{U}$ , où, là encore,  $0 \leq \alpha \leq 1$ . L'EQMP de cet estimateur est

$$\text{EQMP}(\hat{U}^\alpha) = \sum_{ij} \omega_{ij} [\alpha^2(\sigma_{ij} + b_i b_j) + (\alpha - 1)^2 U_i U_j + 2\alpha(\alpha - 1) U_i b_j], \quad (18)$$

laquelle expression est minimisée lorsque

$$\alpha = \frac{\sum_{ij} \omega_{ij} U_i (U_j + b_j)}{\sum_{ij} \omega_{ij} [\sigma_{ij} + (U_i + b_i)(U_j + b_j)]}. \quad (19)$$

Si, comme dans la section 3.3, nous posons l'hypothèse que  $\sum_{ij} \omega_{ij} b_i (U_j + b_j) \leq 0$  une borne inférieure pour l'alpha optimal est définie

$$\alpha_L = \frac{\sum_{ij} \omega_{ij} (U_i + b_i)(U_j + b_j)}{\sum_{ij} \omega_{ij} [\sigma_{ij} + (U_i + b_i)(U_j + b_j)]}, \quad (20)$$

que nous estimons par

$$\hat{\alpha}_L = \frac{\sum_{ij} \omega_{ij} \hat{U}_i \hat{U}_j}{\sum_{ij} \omega_{ij} [\hat{\sigma}_{ij} + \hat{U}_i \hat{U}_j]}, \quad (21)$$

en supposant que les  $\omega_{ij}$  sont connus. En pratique, on estime les  $\omega_{ij}$  en substituant dans l'équation (15), comme on l'a fait pour l'estimateur de test préliminaire, les chiffres bruts du recensement ou les chiffres redressés.

Dans le cas des totaux de population par exemple, le degré de redressement estimé est

$$\hat{\alpha}_L = \frac{\sum_i \hat{T}_i \hat{U}_i^2}{\sum_i \hat{T}_i [\hat{\sigma}_i^2 + \hat{U}_i^2]}, \quad (22)$$

où  $\hat{U}_i$  est le taux de sous-dénombrement estimé, c.-à-d.  $\hat{U}_i / (Y_i + \hat{U}_i)$ , et  $\hat{\sigma}_i^2$ , la variance estimée correspondante.

En ce qui concerne les proportions de population, le degré de redressement est calculé au moyen de la formule

$$\hat{\alpha}_L = \frac{\sum_i \hat{T}_i \hat{U}_i^2 - \hat{T} \hat{U}^2}{\sum_i \hat{T}_i [\hat{\sigma}_i^2 + \hat{U}_i^2] - \hat{T} (\hat{\sigma}^2 + \hat{U}^2)}, \quad (23)$$

où  $\hat{U}$  est le taux de sous-dénombrement estimé pour la population globale, c.-à-d.  $\sum_i \hat{U}_i / \sum_i (Y_i + \hat{U}_i)$ , et  $\hat{\sigma}^2$ , la variance estimée correspondante. L'inverse de l'effectif recensé corrigé a servi de poids dans ces deux exemples.

### 3.5 Comparaisons numériques

Pour les totaux de population uniques, il a été possible de définir des formules exactes ou approximatives pour l'EQM des quatre estimateurs étudiés, en fonction de  $U^2/\sigma^2$ , de  $R$ , de  $r$  et (dans le cas de l'estimateur de test préliminaire) de la valeur critique du test. Malheureusement, il n'est pas encore possible de faire de même pour l'EQMP de fonctions complexes d'un vecteur de totaux de population.

Néanmoins, on peut estimer l'EQMP de trois estimateurs – celui fondé sur les chiffres non redressés du recensement, celui fondé sur les chiffres redressés et l'estimateur composite – en substituant dans l'équation (18) des estimations du sous-dénombrement ainsi que les variances estimées correspondantes (s'il existe des estimations du biais, on peut aussi s'en servir; cependant, nous supposons ici qu'il n'y a pas de biais). À titre d'exemple, les figures 8 et 9 contiennent un graphique où le rapport estimé de l'EQMP à l'EQMP optimale pour le recensement de 1981 est exprimé en fonction d' $\alpha$ ; là encore, les provinces sont les unités désignées par  $i$ . Les valeurs extrêmes de  $\alpha = 0$  et  $\alpha = 1$  correspondent respectivement au cas des chiffres non redressés du recensement et au cas des chiffres redressés, tandis que le point minimum de la courbe correspond à l'alpha optimal. La figure 8 porte sur les totaux et la figure 9 sur les proportions. Les valeurs optimales d' $\alpha$  ont été calculées au moyen des formules (22) et (23).

Dans chaque cas, le degré de redressement optimal est proche de 1.0 et donne une EQMP beaucoup moins élevée que celle qui correspond à l'absence de redressement (par un facteur de près de 70, par exemple, en ce qui concerne les totaux de population). Le degré de redressement optimal est moins élevé pour les proportions que pour les totaux, ce qui dénote une fois de plus que les estimations de différences de taux de sous-dénombrement entre les provinces sont moins précises que les estimations de taux de sous-dénombrement. Il est intéressant aussi de constater que l'EQMP qui correspond au redressement complet n'est que légèrement plus élevée que celle qui correspond au degré de redressement optimal. Cette similitude peut avoir des conséquences pratiques importantes car il est beaucoup plus facile de justifier un redressement complet qu'un redressement partiel devant les utilisateurs de données.

#### 4. ESTIMATION POUR PETITES RÉGIONS

Dans les deux sections précédentes, nous avons vu qu'il était possible d'obtenir des estimations directes du sous-dénombrement, ainsi que les variances estimées correspondantes, grâce aux études de mesure de la couverture. Cela est possible notamment pour les provinces, pour certaines grandes régions métropolitaines de recensement et pour de grands groupes démographiques (ex.: âge-sexe, âge-état matrimonial) au niveau national. Or, le Programme des estimations démographiques (PED) produit des estimations à des niveaux de détail très poussés (ex.: par année d'âge, sexe et état matrimonial pour quelque 260 divisions de recensement). En règle générale, il n'existe pas d'estimations directes du sous-dénombrement à ces niveaux.

Néanmoins, à cause de la nécessité d'assurer une cohérence dans les estimations, il faut veiller à ce que tout redressement effectué aux niveaux supérieurs "se répercute" jusqu'aux niveaux les plus bas utilisés dans le PED. Dans cette section, nous voyons comment l'estimation synthétique peut être utile dans cette perspective et comment on peut encore se servir de l'EQMP pour construire des estimateurs de test préliminaire et des estimateurs composites.

L'estimateur synthétique repose sur l'hypothèse que le sous-dénombrement net est uniforme à l'intérieur d'un certain nombre de "groupes de redressement" désignés par l'indice  $a$ . L'estimation synthétique est calculée au moyen de la formule  $\hat{U}_i^S = \sum_a \lambda_{ia} \hat{U}_a$  où  $\lambda_{ia} = Y_{ia}/Y_a$ . Par exemple, les groupes de redressement pourraient être les groupes d'âge-sexe, pour lesquels il existe des estimations du sous-dénombrement  $\hat{U}_a$  à un niveau supérieur.

Lorsqu'il y a un seul groupe de redressement, nous sommes devant un cas particulier de l'estimateur synthétique que Wolter et Causey (1991) appellent l'estimateur général. Il est défini par la formule  $\hat{U}_i^{ATB} = \lambda_i \hat{U}$  où  $\lambda_i = Y_i/Y$ . On peut calculer l'EQMP de l'estimateur général et de l'estimateur synthétique au moyen de l'équation (14). Comme  $\omega_{ij}$  ne dépend pas de l'estimateur utilisé, seule l'expression entre crochets varie. Dans le tableau 3, nous comparons les estimateurs de  $U_i$  ainsi que les termes de covariance et de biais correspondants pour les quatre estimateurs suivants: estimateur fondé sur les chiffres bruts du recensement, estimateur fondé sur les chiffres redressés, estimateur général et estimateur synthétique.

**Tableau 3**  
Exemples de covariances et de biais dans la formule d'approximation de l'EQMP pour divers estimateurs

Estimateur	$\hat{U}_i^*$	$\text{Cov}(\hat{U}_i^*, \hat{U}_j^*)$	Biais( $\hat{U}_i^*$ )
Chiffres bruts du recensement	0	0	$- U_i$
Chiffres redressés du recensement	$\hat{U}_i$	$\sigma_{ij}$	$b_i$
Général	$\lambda_i \hat{U}$	$\lambda_i \lambda_j \sigma^2$	$\lambda_i (U + b) - U_i$
Synthétique	$\sum_a \lambda_{ia} \hat{U}_a$	$\sum_{aa'} \lambda_{ia'} \lambda_{ja'} \sigma_{aa'}$	$\sum_a \lambda_{ia} (U_a + b_a) - U_i$

où  $b = \sum_i b_i$  et de même,  $b_a$  est le biais de  $\hat{U}_a$ .

#### 4.1 Estimateurs de test préliminaire

Tout comme ce fut le cas dans les sections 2 et 3, nous pouvons nous servir de l'EQMP pour construire des tests statistiques qui permettront de choisir l'un ou l'autre de deux estimateurs. Supposons, par exemple, que nous devons choisir entre l'estimateur fondé sur les chiffres non redressés du recensement et l'estimateur général pour des totaux de population (le redressement "général" n'a évidemment aucun effet sur les proportions de population). Si nous comparons les EQMP de ces deux estimateurs, nous constatons que l'estimateur général sera préférable à l'estimateur fondé sur les chiffres du recensement si

$$\sigma^2 < U^2(1 - R^2) \left[ 1 - \frac{2TB}{U(1 - R)} \right], \quad (24)$$

où

$$B = 1 - \frac{1}{\sum_i \frac{\lambda_i^2}{\tau_i}} \quad (25)$$

et  $\tau_i = T_i/T$ . Cette condition est définie sous une forme différente dans Wolter et Causey (1991).  $B$  est un indice de l'hétérogénéité du sous-dénombrement; il est non négatif, et est égal à zéro si et seulement si le sous-dénombrement est parfaitement uniforme.

L'inéquation (24) étant semblable à l'inéquation (1) (si l'on fait abstraction du terme entre crochets), il est possible de construire un test très comparable à celui qui est décrit dans la section 2.3. La valeur critique du coefficient de variation dépendra, comme avant, du niveau de signification choisi et du biais relatif mais aussi de  $B/\bar{U}$ , le rapport entre l'indice d'hétérogénéité du sous-dénombrement et le taux de sous-dénombrement global.

Royce (1991) a montré que ce facteur additionnel a un effet plutôt négligeable en pratique sur le CV critique. Par conséquent, si le redressement est justifié aux niveaux supérieurs, nous ne voyons pas pourquoi il ne le serait pas à des niveaux inférieurs. Wolter et Causey (1991) arrivent à des conclusions semblables dans une étude de simulation.

#### 4.2 Estimateurs composites

Dans les sections 2 et 3, nous avons examiné des estimateurs composites dont les deux versions extrêmes correspondaient l'une, à l'estimateur fondé sur les chiffres non redressés du recensement et l'autre, à l'estimateur fondé sur les chiffres redressés du recensement. Avec l'addition de l'estimateur synthétique et de l'estimateur général, l'éventail des estimateurs composites s'est accru sensiblement. Par exemple, on pourrait envisager des estimateurs composites formés d'un estimateur fondé sur les chiffres non redressés et d'un estimateur synthétique, d'un estimateur fondé sur les chiffres redressés et d'un estimateur général, d'un estimateur général et d'un estimateur synthétique, et ainsi de suite. Par conséquent, nous décrivons ci-dessous une méthode qui permet de construire un estimateur composite formé de deux estimateurs quelconques.

L'estimateur composite général est défini par l'expression  $\hat{U}^* = \alpha \hat{U}_1 + (1 - \alpha) \hat{U}_2$  où  $\hat{U}_1$  et  $\hat{U}_2$  sont deux estimateurs. L'EQMP de cet estimateur est

$$\begin{aligned} \text{EQMP}(g(\hat{T}^*)) &= \alpha^2 \text{EQMP}(g(\hat{T}_1)) + (1 - \alpha)^2 \text{EQMP}(g(\hat{T}_2)) \\ &+ 2\alpha(1 - \alpha) \text{EPCMP}(g(\hat{T}_1, g(\hat{T}_2))), \end{aligned} \quad (26)$$

où

$$\text{EPCMP}(g(\hat{T}_1), g(\hat{T}_2)) = \sum_{ij} \omega_{ij} [\text{Cov}(\hat{U}_{1i}, \hat{U}_{2j}) + \text{Biais}(\hat{U}_{1i})\text{Biais}(\hat{U}_{2j})] \quad (27)$$

est défini comme l'erreur de produit croisé moyenne pondérée de  $g(\hat{T}_1)$  et  $g(\hat{T}_2)$ . L'EQMP de notre estimateur composite est minimisée lorsque

$$\alpha = \frac{\text{EQMP}(g(\hat{T}_2)) - \text{EPCMP}(g(\hat{T}_1), g(\hat{T}_2))}{\text{EQMP}(g(\hat{T}_1)) + \text{EQMP}(g(\hat{T}_2)) - 2\text{EPCMP}(g(\hat{T}_1), g(\hat{T}_2))}. \quad (28)$$

Pour obtenir une estimation de  $\alpha$ , on remplace l'EQMP et l'EPCMP dans l'équation ci-dessus par leurs valeurs estimées.

Pour illustrer l'utilisation de cette méthode, supposons qu'on a décidé de redresser l'effectif de la population d'une province. Pour effectuer le redressement à tous les niveaux d'agrégation, on a le choix entre le redressement général (c.-à-d. redresser toutes les quantités infra-provinciales au moyen du même facteur) et le redressement synthétique, où les corrections se font séparément dans plusieurs groupes d'âge-sexe. L'avantage de la première méthode est qu'elle utilise uniquement l'estimation du sous-dénombrement pour l'ensemble de la province, laquelle estimation est vraisemblablement plus fiable que les estimations du sous-dénombrement par groupe d'âge-sexe au niveau provincial. En revanche, si le sous-dénombrement varie beaucoup entre les groupes d'âge-sexe et si les quantités infra-provinciales désignées par l'indice  $i$  varient elles aussi selon la composition âge-sexe, il sera préférable d'opter pour l'estimateur synthétique.

S'il est possible d'obtenir les valeurs estimées de  $U_i$  d'une source quelconque, on peut estimer toutes les composantes (covariance et biais) des EQMP et de l'EPCMP (en se servant de formules comme celles du tableau 3) et déterminer l'estimateur composite optimal formé de l'estimateur général et de l'estimateur synthétique. Bien qu'en pratique, il soit presque impossible de connaître les  $U_i$  pour des quantités infra-provinciales, on peut envisager l'application de la méthode à des niveaux supérieurs. Par exemple, on pourrait assimiler les provinces aux quantités désignées par l'indice  $i$  et se servir de facteurs de redressement (général et synthétique) calculés pour l'ensemble du Canada. Une deuxième solution serait de créer une population artificielle (comme dans Shirm et Preston (1987) ou Wolter et Causey (1991) par exemple) où les  $U_i$  sont supposés connus.

## 5. SUJETS DE RECHERCHE

L'étude que nous venons d'exposer n'est qu'une première étape dans l'analyse et la comparaison du rendement de divers estimateurs d'une série de chiffres de population. Beaucoup de recherches restent encore à faire sur de nombreux sujets.

Premièrement, il faut étudier plus en profondeur l'EQMP de l'estimateur de test préliminaire et de l'estimateur composite pour les cas plus généraux décrits dans les sections 3 et 4. Bien que l'on ne soit pas encore parvenu à définir des expressions analytiques pour ces EQMP, il se pourrait que les résultats plus généraux relatifs aux estimateurs de test préliminaire et aux estimateurs de Stein-Rule dont font état Judge et Bock (1978) s'avèrent utiles. Si tel était le cas, nous pourrions nous servir de ces résultats pour répondre à des questions comme celles-ci: Peut-on trouver des valeurs critiques optimales pour les estimateurs de test préliminaire des sections 3.3 et 4.1 suggérés par Fellegi? Comment l'EQMP de l'estimateur de test préliminaire se compare-t-elle dans la pratique à celle des trois autres estimateurs?



En deuxième lieu, il faut aussi pousser plus loin l'analyse de la sensibilité des résultats par rapport à différents poids dans la fonction de perte. Les résultats de la section 3 avaient été obtenus à la suite de l'utilisation d'un poids égal à l'inverse de l'effectif recensé ou de l'effectif recensé corrigé pour chaque province. Si la pondération avait été différente, les résultats ne seraient pas les mêmes. Un poids plus général qui pourrait être intéressant est  $w_k = Y_k^\gamma$ , où  $\gamma$  est une sorte de paramètre de puissance. On pourrait alors étudier la sensibilité des résultats de la section 3 par rapport à ce paramètre.

Finalement, bien que les méthodes que nous venons de décrire dans cet article constituent un ensemble organisé pour l'élaboration et l'évaluation d'estimateurs, la manière exacte dont elles seront utilisées reste à définir. Voici les questions particulières qu'il faut résoudre:

1. Quelle est l'importance relative de diverses fonctions telles que les totaux, les proportions et les taux de croissance? Des fonctions différentes donneront des résultats différents mais en définitive, il faut choisir un seul estimateur afin d'assurer la cohérence des estimations.
2. À quels niveaux d'agrégation géographique et démographique ces méthodes devraient-elles être utilisées? Par exemple, devrait-on utiliser l'estimateur de test préliminaire ou l'estimateur composite décrits dans la section 3 au niveau provincial, au niveau provincial croisé avec les groupes d'âge-sexe ou à des niveaux encore plus fins? Les résultats obtenus dépendent du niveau d'analyse utilisé.
3. Pourrions-nous même envisager d'utiliser les estimateurs composites comme estimateurs de premier plan, par exemple comme estimateurs de la population totale des provinces? Il serait peut-être difficile d'expliquer aux utilisateurs pourquoi les redressements ne concordent pas avec les estimations officielles du sous-dénombrement.

Comme il faut faire preuve de jugement professionnel pour résoudre de telles questions, la décision d'opérer ou non un redressement ne peut être prise de façon automatique suivant des critères préétablis (il en va de même du choix de la méthode de redressement). Tandis que les méthodes que nous avons décrites dans cet article peuvent à coup sûr servir de jalons, la décision finale devra reposer sur un examen sérieux où l'on mettra en balance les possibilités de gain dans la précision des estimations et la facilité avec laquelle les méthodes choisies pourront être communiquées aux utilisateurs des données du PED et comprises par ceux-ci.

## REMERCIEMENTS

L'auteur tient à remercier le rédacteur en chef, les deux arbitres ainsi que Richard Carter pour les nombreux commentaires utiles qui ont contribué à rehausser la qualité de cet article.

## BIBLIOGRAPHIE

- ANDREWS, D. (1991). Discussion sollicitée lors de la réunion du comité consultatif des méthodes statistiques de Statistique Canada, octobre 1991.
- BROOK, R.J. (1976). On the use of a regret function to set significance points in prior tests of estimation. *Journal of the American Statistical Association*, 71, 353, 126-131.
- CITRO, C.F., et COHEN, M.L. (1985). *The Bicentennial Census: New Directions for Methodology in 1990*. Washington, D.C.: National Academy Press.
- COHEN, A. (1965). Estimates of the linear combination of parameters in the mean vector of a multivariate distribution. *Annals of Mathematical Statistics*, 36, 78-87.
- FELLEGI, I.P. (1980). Should the census count be adjusted for allocation purposes? Equity considerations. Dans *Proceedings of the 1980 Conference on Census Undercount*, U.S. Bureau of the Census.

- JUDGE, G.G., et BOCK, M.E. (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*. Amsterdam-New York-Oxford: North-Holland Publishing Company.
- ROYCE, D., et LUC, M. (1990). Recalculation of Fellegi's test statistics on census adjustment for the 1981 and 1986 censuses. Rapport interne, Statistique Canada.
- ROYCE, D. (1991). Technical criteria for adjusting the population estimates program for census coverage error. Rapport interne, Statistique Canada.
- ROYCE, D. (1992). Incorporating estimates of census coverage error into the Canadian population estimates program. *Proceedings of the Eighth Annual Research Conference*, Bureau of the Census, Washington, DC (à paraître).
- SAWA, T., et HIROMATSU, T. (1973). Minimax regret significance points for a preliminary test in regression analysis. *Econometrica*, 41, 1093-1101.
- SHIRM, A.L., et PRESTON, S.H. (1987). Census undercount adjustment and the quality of geographic population distributions. *Journal of the American Statistical Association*, 82, 400, 965-978.
- SPENCER, B. (1980). Implications of equity and accuracy for undercount adjustment: A decision-theoretic approach. Dans *Proceedings of the 1980 Conference on Census Undercount*, Bureau of the Census, Washington, DC.
- SPENCER, B. (1986). Conceptual issues in measuring improvement in population estimates. Dans *Proceedings of the Second Annual Research Conference*, Bureau of the Census, Washington, DC, 393-407.
- TORO-VIZCORRONDO, C., et WALLACE, T.D. (1968). A test of the mean square error criterion for restrictions in linear regression. *Journal of the American Statistical Association*, 63, 322, 558-572.
- WOLTER, K.M., et CAUSEY, B.D. (1991). Evaluation of procedures for improving population estimates for small areas. *Journal of the American Statistical Association*, 86, 414, 278-284.