Composante du produit nº 12-001-X
Division des méthodes d'enquêtes auprès des entreprises

Article

Bootstrap rééchelonné pour l'échantillonnage stratifié à plusieurs degrés

par John Preston

Décembre 2009

Statistique Canada Statistics Canada

Canada

Techniques d'enquête, décembre 2009
Vol. 35, N° 2, pp. 247-254
Statistique Canada, N° 12-001-X au catalogue

Bootstrap rééchelonné pour l'échantillonnage stratifié à plusieurs degrés

John Preston [1]

Résumé

Dans les enquêtes par sondage de grande portée, il est fréquent d'employer des plans de sondage stratifiés à plusieurs degrés où les unités sont sélectionnées par échantillonnage aléatoire simple sans remise à chaque degré. L'exécution de l'estimation de la variance sous ce genre de plan peut être assez fastidieuse, particulièrement pour les estimateurs non linéaires. Diverses méthodes bootstrap d'estimation de la variance ont été proposées, mais la plupart sont limitées à des plans à un seul degré ou à des plans en grappes à deux degrés. Nous proposons une extension de la méthode du bootstrap rééchelonné (Rao et Wu 1988) aux plans stratifiés à plusieurs degrés qui peut être adaptée facilement à n'importe quel nombre de degrés. Cette méthode convient pour une grande gamme de méthodes de repondération, y compris la classe générale des estimateurs par calage. Nous avons réalisé une étude par simulation Monte Carlo pour examiner la performance de l'estimateur de variance bootstrap rééchelonné à plusieurs degrés.

Mots clés : Bootstrap ; calage ; échantillonnage à plusieurs degrés ; stratification ; estimation de variance.

1. Introduction

Les plans d'échantillonnage stratifié à plusieurs degrés conviennent particulièrement bien pour les enquêtes par sondage de grande portée, parce qu'ils ont l'avantage de permettre le regroupement des efforts de collecte. Diverses méthodes d'estimation de la variance existent pour ces plans d'enquête complexes. Les plus fréquentes sont la méthode de linéarisation (ou méthode de Taylor) et les méthodes de rééchantillonnage, telles que le jackknife, les répliques équilibrées répétées et le bootstrap. Dans le cas de plans de sondage complexes, l'exécution de la méthode de linéarisation peut être assez fastidieuse, car elle nécessite l'établissement de formules de variance distinctes pour chaque estimateur non linéaire. Certaines approximations sont habituellement requises pour la variance de fonctions non linéaires, telles que les ratios, les coefficients de corrélation et les coefficients de régression, ainsi que les fonctionnelles, telles que les quantiles.

En revanche, les diverses méthodes de rééchantillonnage s'appuient sur une seule formule de variance pour tous les estimateurs. Les méthodes de répétition de l'échantillon peuvent refléter les effets d'une vaste gamme de méthodes de repondération, y compris le calage, et des ajustements dus à la non-réponse et à la couverture insuffisante de population. Les méthodes du jackknife et des répliques répétées équilibrées ne sont applicables qu'aux plans stratifiés à plusieurs degrés dans lesquels les grappes sont échantillonnées avec remise ou les fractions d'échantillonnage au premier degré sont négligeables. Un certain nombre de méthodes bootstrap pour l'échantillonnage en population finie ont été proposées dans la littérature, y compris le bootstrap avec remise (McCarthy et Snowden 1985), le bootstrap rééchelonné (Rao et Wu 1988), le bootstrap à concordance-miroir (*mirror-match bootstrap*) (Sitter 1992a) et le bootstrap sans remise (Gross 1980 ; Bickel et Freedman 1984 ; Sitter 1992b). Un résumé de ces méthodes bootstrap peut être consulté dans Shao et Tu (1995).

La plupart de ces méthodes bootstrap sont limitées à des plans à un seul degré ou à des plans à plusieurs degrés dans lesquels les unités d'échantillonnage de premier degré sont sélectionnés avec remise ou les fractions d'échantillonnage de premier degré sont faibles dans la plupart des strates. Cependant, dans de nombreuses enquêtes par sondage de grande portée, il est courant d'employer des plans de sondage à plusieurs degrés fortement stratifiés où les unités sont sélectionnées par échantillonnage aléatoire simple sans remise à chaque degré. Les enquêtes employeurs-employés, telles que la Survey of Employee Earnings and Hours (ABS 2008) et les enquêtes écoles-élèves, telles que la National Survey on the Use of Tobacco by Australian Secondary School Students (White et Hayman 2006), sont des exemples typiques de ce genre d'enquête.

McCarthy et Snowden (1985) ont proposé une extension de leur bootstrap avec remise à l'échantillonnage à deux degrés dans le cas particulier où les tailles de grappes sont égales et les tailles d'échantillon dans les grappes sont égales, tandis que Rao et Wu (1988) et Sitter (1992a) ont donné des extensions de leurs méthodes respectives du bootstrap rééchelonné et du bootstrap à concordance-miroir à l'échantillonnage en grappes à deux degrés. Récemment, Funaoka, Saigo, Sitter et Toida (2006) ont proposé des méthodes bootstrap de type bernoullien, le bootstrap bernoullien général et le bootstrap bernoullien abrégé, qui permettent de traiter facilement les plans stratifiés à plusieurs degrés dans lesquels les unités sont sélectionnées par

1. John Preston, Australian Bureau of Statistics, 639 Wickham Street, Fortitude Valley QLD 4006, Australie. Courriel : john.preston@abs.gov.au.

échantillonnage aléatoire simple sans remise à chaque degré. Le bootstrap bernoullien général a l'avantage de permettre le traitement de toute combinaison de tailles d'échantillon, mais il requiert un beaucoup plus grand nombre de générations de nombres aléatoires que le bootstrap bernoullien abrégé.

Dans le présent article, nous proposons une extension de la méthode du bootstrap rééchelonné à l'échantillonnage stratifié à plusieurs degrés dans lequel les unités sont sélectionnées par échantillonnage aléatoire simple sans remise à chaque degré. À la section 2, nous présentons la notation pour l'échantillonnage stratifié à plusieurs degrés. À la section 3, nous décrivons l'extension de l'estimateur bootstrap rééchelonné à l'échantillonnage à plusieurs degrés. À la section 4, nous présentons les principaux résultats d'une étude par simulation. Enfin, à la section 5, nous tirons certaines conclusions.

2. Échantillonnage stratifié à plusieurs degrés

Pour simplifier, nous présentons le cas de l'échantillonnage stratifié à trois degrés. Considérons une population finie U divisée en H strates non chevauchantes $U = \{U_1, ..., U_H\}$, où U_h est constituée de N_{1h} unités primaires d'échantillonnage (UPE). Au premier degré, un échantillon aléatoire simple sans remise (EASSR) de n_{1h} UPE est tiré avec probabilités de sélection $\pi_{1hi} = n_{1h} / N_{1h}$ dans chaque strate h. Supposons que l'UPE i sélectionnée dans la strate h est constituée de N_{2hi} unités secondaires d'échantillonnage (USE). Au deuxième degré, un EASSR de n_{2hi} USE est tiré avec probabilités de sélection $\pi_{2hij} = n_{2hi} / N_{2hi}$ dans chaque UPE sélectionnée. Supposons que l'USE j sélectionnée dans l'UPE i sélectionnée dans la strate h est constituée de N_{3hij} unités finales d'échantillonnage (UFE). Au troisième degré, un EASSR de n_{3hij} UFE est tiré avec probabilités de sélection $\pi_{3hijk} = n_{3hij} / N_{3hij}$ dans chaque USE sélectionnée.

L'objectif est d'estimer le total de population $Y = \sum_{h=1}^{H}\sum_{i=1}^{N_{1h}}\sum_{j=1}^{N_{2hi}}\sum_{k=1}^{N_{3hij}} y_{hijk}$, où y_{hijk} est la valeur de la variable d'intérêt y pour l'UFE k dans l'USE j dans l'UPE i dans la strate h. Une estimation sans biais de Y est donnée par :

$$\hat{Y} = \sum_{h=1}^{H}\hat{Y}_h = \sum_{h=1}^{H}\frac{N_{1h}}{n_{1h}}\sum_{i=1}^{n_{1h}}\frac{N_{2hi}}{n_{2hi}}\sum_{j=1}^{n_{2hi}}\frac{N_{3hij}}{n_{3hij}}\sum_{k=1}^{n_{3hij}} y_{hijk}$$

où $\hat{Y}_h = (N_{1h} / n_{1h})\sum_{i=1}^{n_{1h}}\hat{Y}_{hi}$, $\hat{Y}_{hi} = (N_{2hi} / n_{2hi})\sum_{j=1}^{n_{2hi}}\hat{Y}_{hij}$ et $\hat{Y}_{hij} = (N_{3hij}/n_{3hij})\sum_{k=1}^{n_{3hij}} y_{hijk}$. Cet estimateur peut également s'écrire sous la forme $\hat{Y} = \sum_{h=1}^{H}\sum_{i=1}^{n_{1h}}\sum_{j=1}^{n_{2hi}}\sum_{k=1}^{n_{3hij}} w_{hijk} y_{hijk}$, où $w_{hijk} = w_{1hi}w_{2hij}w_{3hijk} = (N_{1h}/n_{1h})(N_{2hi}/n_{2hi})(N_{3hij}/n_{3hij})$ est le poids d'échantillonnage pour l'UFE k dans l'USE j dans l'UPE i dans la strate h.

Un estimateur sans biais de $\mathrm{Var}(\hat{Y})$ est donné par (Särndal, Swensson et Wretman 1992) :

$$\begin{aligned}\hat{\mathrm{Var}}(\hat{Y}) = {} & \sum_{h=1}^{H}\frac{N_{1h}^2}{n_{1h}}(1-f_{1h})s_{1h}^2 \\ & + \sum_{h=1}^{H}\frac{N_{1h}}{n_{1h}}\sum_{i=1}^{n_{1h}}\frac{N_{2hi}^2}{n_{2hi}}(1-f_{2hi})s_{2hi}^2 \\ & + \sum_{h=1}^{H}\frac{N_{1h}}{n_{1h}}\sum_{i=1}^{n_{1h}}\frac{N_{2hi}}{n_{2hi}}\sum_{j=1}^{n_{2hi}}\frac{N_{3hij}^2}{n_{3hij}}(1-f_{3hij})s_{3hij}^2 \quad (2.1)\end{aligned}$$

où $f_{1h} = (n_{1h}/N_{1h})$, $f_{2hi} = (n_{2hi}/N_{2hi})$, $f_{3hij} = (n_{3hij}/N_{3hij})$, $\bar{\hat{Y}}_h = \sum_{i=1}^{n_{1h}}\hat{Y}_{hi}/n_{1h}$, $\bar{\hat{Y}}_{hi} = \sum_{j=1}^{n_{2hi}}\hat{Y}_{hij}/n_{2hi}$, $\bar{y}_{hij} = \sum_{k=1}^{n_{3hij}} y_{hijk}/n_{3hij}$, $s_{1h}^2 = \sum_{i=1}^{n_{1h}}(\hat{Y}_{hi} - \bar{\hat{Y}}_h)^2/(n_{1h}-1)$, $s_{2hi}^2 = \sum_{j=1}^{n_{2hi}}(\hat{Y}_{hij} - \bar{\hat{Y}}_{hi})^2/(n_{2hi}-1)$ et $s_{3hij}^2 = \sum_{k=1}^{n_{3hij}}(y_{hijk} - \bar{y}_{hij})^2/(n_{3hij}-1)$.

3. Bootstrap rééchelonné pour l'échantillonnage stratifié à plusieurs degrés

Rao et Wu (1988) ont proposé un changement d'échelle de la méthode du bootstrap standard pour divers plans d'échantillonnage, dont l'échantillonnage stratifié. Comme les facteurs d'échelle rajustent les valeurs des données d'enquête, cette méthode ne s'applique qu'à des statistiques lisses. Rao, Wu et Yue (1992) ont présenté une modification de cette méthode du bootstrap rééchelonné dans la laquelle les facteurs d'échelle sont appliqués aux poids de sondage plutôt qu'aux valeurs des données d'enquête. Cette méthode du bootstrap rééchelonné modifiée est équivalente à la méthode originale, mais a l'avantage supplémentaire d'être applicable à des statistiques non lisses ainsi qu'à des statistiques lisses. Kovar, Rao et Wu (1988) ont montré que si l'on utilise une taille d'échantillon bootstrap de $n_h^* = n_h - 1$, l'estimateur bootstrap rééchelonné donne de bons résultats pour les statistiques lisses.

Bien que les échantillons bootstrap soient habituellement sélectionnés avec remise, Chipperfield et Preston (2007) ont modifié la méthode du bootstrap rééchelonné pour l'appliquer à la situation où les échantillons bootstrap sont sélectionnés sans remise. Sous cette méthode du bootstrap rééchelonné sans remise, nous pouvons montrer que le choix de $n_h^* = \lfloor n_h/2 \rfloor$ ou $n_h^* = \lceil n_h/2 \rceil$ est optimal, où les opérateurs $\lfloor x \rfloor$ et $\lceil x \rceil$ arrondissent l'argument x vers le bas et vers le haut, respectivement, à l'entier le plus proche. Le choix de $n_h^* = \lfloor n_h/2 \rfloor$ à la propriété désirable de donner lieu à des poids bootstrap qui ne sont jamais négatifs.

Pour simplifier l'exposé, nous présentons le cas de l'échantillonnage stratifié à trois degrés, mais la méthode proposée peut être étendue facilement à n'importe quel nombre de degrés. La procédure du bootstrap rééchelonné sans remise pour l'échantillonnage stratifié à trois degrés est la suivante :

a) Tirer un échantillon aléatoire simple de n_{1h}^* UPE sans remise parmi les n_{1h} UPE dans l'échantillon. Soit δ_{1hi} égal à 1 si l'UPE i dans la strate h est sélectionnée et 0 autrement. Calculer les poids bootstrap des UPE :

$$w_{1hi}^* = w_{1hi}\left(1 - \lambda_{1h} + \lambda_{1h}\frac{n_{1h}}{n_{1h}^*}\delta_{1hi}\right)$$

où $\lambda_{1h} = \sqrt{n_{1h}^*(1 - f_{1h})/(n_{1h} - n_{1h}^*)}$.

b) Dans chaque UPE figurant dans l'échantillon, tirer un échantillon aléatoire simple de n_{2hi}^* USE sans remise parmi les n_{2hi} USE qui figurent dans l'échantillon. Soit δ_{2hi} égal à 1 si l'USE j dans l'UPE i dans la strate h est sélectionnée et 0 autrement. Calculer les poids bootstrap conditionnels des USE :

$$w_{2hij}^* = w_{2hij}\frac{w_{1hi}}{w_{1hi}^*}\left(1 - \lambda_{1h} + \lambda_{1h}\frac{n_{1h}}{n_{1h}^*}\delta_{1hi} - \lambda_{2hi}\sqrt{\frac{n_{1h}}{n_{1h}^*}}\delta_{1hi} + \lambda_{2hi}\sqrt{\frac{n_{1h}}{n_{1h}^*}}\delta_{1hi}\frac{n_{2hi}}{n_{2hi}^*}\delta_{2hij}\right)$$

où $\lambda_{2hi} = \sqrt{n_{2hi}^* f_{1h}(1 - f_{2hi})/(n_{2hi} - n_{2hi}^*)}$.

c) Dans chaque USE figurant dans l'échantillon, tirer un échantillon aléatoire simple de n_{3hij}^* UFE sans remise parmi les n_{3hij} UFE qui figurent dans l'échantillon. Soit δ_{3hijk} égal à 1 si l'UFE k dans l'USE j dans l'UPE i dans la strate h est sélectionnée et 0 autrement. Calculer les poids bootstrap conditionnels des UFE :

$$w_{3hijk}^* = w_{3hijk}\frac{w_{1hi}}{w_{1hi}^*}\frac{w_{2hij}}{w_{2hij}^*}\left(1 - \lambda_{1h} + \lambda_{1h}\frac{n_{1h}}{n_{1h}^*}\delta_{1hi} - \lambda_{2hi}\sqrt{\frac{n_{1h}}{n_{1h}^*}}\delta_{1hi} + \lambda_{2hi}\sqrt{\frac{n_{1h}}{n_{1h}^*}}\delta_{1hi}\frac{n_{2hi}}{n_{2hi}^*}\delta_{2hij} - \lambda_{3hij}\sqrt{\frac{n_{1h}}{n_{1h}^*}}\delta_{1hi}\sqrt{\frac{n_{2hi}}{n_{2hi}^*}}\delta_{2hij} + \lambda_{3hij}\sqrt{\frac{n_{1h}}{n_{1h}^*}}\delta_{1hi}\sqrt{\frac{n_{2hi}}{n_{2hi}^*}}\delta_{2hij}\frac{n_{3hij}}{n_{3hij}^*}\delta_{3hijk}\right)$$

où $\lambda_{3hij} = \sqrt{n_{3hij}^* f_{1h} f_{2hi}(1 - f_{3hij})/(n_{3hij} - n_{3hij}^*)}$.

d) Calculer les estimations bootstrap :

$$\hat{Y}^* = \sum_{h=1}^{H}\sum_{i=1}^{n_{1h}}\sum_{j=1}^{n_{2hi}}\sum_{k=1}^{n_{3hij}} w_{hijk}^* y_{hijk}, \quad \hat{\theta} = g(\hat{Y}^*)$$

où $w_{hijk}^* = w_{1hi}^* w_{2hij}^* w_{3hijk}^*$.

e) Répéter indépendamment les étapes a) à d) un grand nombre de fois, B, et calculer les estimations bootstrap, $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \ldots, \hat{\theta}^{(B)}$.

f) L'estimateur de variance bootstrap de $\hat{\theta}$ est donné par :

$$\mathrm{Var}(\hat{\theta}) = E_*(\hat{\theta} - E_*(\hat{\theta}))^2 \qquad (3.1)$$

ou par l'approximation Monte Carlo :

$$\mathrm{Var}(\hat{\theta}) = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^{(b)} - \bar{\hat{\theta}})^2$$

où $\bar{\hat{\theta}} = \sum_{b=1}^{B}\hat{\theta}^{(b)}/B$.

Nous montrons en annexe que l'estimateur de variance bootstrap rééchelonné à plusieurs degrés pour l'échantillonnage stratifié à trois degrés défini par (3.1) se réduit à l'estimateur de variance à trois degrés sans biais standard (2.1) dans le cas où $\hat{\theta}$ est un estimateur linéaire. Le choix de $n_{1h}^* = \lfloor n_{1h}/2 \rfloor$, $n_{2hi}^* = \lfloor n_{2hi}/2 \rfloor$ et $n_{3hij}^* = \lfloor n_{3hij}/2 \rfloor$ sera optimal et aura la propriété désirable de donner lieu à des poids bootstrap qui ne sont jamais négatifs.

La procédure proposée peut facilement être étendue à n'importe quel nombre de degrés en ajoutant des termes de la forme $-\lambda_R(\prod_{r=1}^{R-1}\sqrt{(n_r/n_r^*)}\delta_r) + \lambda_R(\prod_{r=1}^{R-1}\sqrt{(n_r/n_r^*)}\delta_r)(n_R/n_R^*)\delta_R$ à chaque degré, R, aux ajustements des poids bootstrap, où $\lambda_R = \sqrt{n_R^*(\prod_{r=1}^{R-1} f_r)(1 - f_R)/(n_R - n_R^*)}$.

Yeo, Mantel et Liu (1999) ont présenté une amélioration du bootstrap rééchelonné qui tient compte des ajustements apportés aux poids de sondage, tels que la poststratification. Par exemple, considérons un cas simple de calage non intégré au moyen d'information auxiliaire pour l'échantillonnage stratifié à deux degrés (Estevao et Särndal 2006) dont le double objectif est de produire des estimations pour une variable d'intérêt de premier degré $Y_1 = \sum_{(hi)\in U} y_{1hi}$, ainsi qu'une variable d'intérêt de deuxième degré, $Y_2 = \sum_{(hij)\in U} y_{2hij}$. Supposons qu'il existe :

i) un ensemble de p variables auxiliaires de premier degré $\boldsymbol{x}_{1hi}$ pour lesquelles les totaux de population $\boldsymbol{X}_1 = \sum_{(hi)\in U}\boldsymbol{x}_{1hi}$ sont connus, et où les totaux de population sont produits à partir d'une liste d'UPE pour lesquelles les $\boldsymbol{x}_{1hi}$ sont connus pour chaque UPE dans la population ; et

ii) un ensemble de q variables auxiliaires de deuxième degré $\boldsymbol{x}_{2hij}$ pour lesquelles les totaux de population $\boldsymbol{X}_2 = \sum_{(hij)\in U}\boldsymbol{x}_{2hij}$ sont connus, où les totaux de population sont obtenus auprès d'une source extérieure.

Les variables auxiliaires peuvent être utilisées pour former les estimateurs par calage de premier et de deuxième degrés :

$$\hat{Y}_{\mathrm{CAL1}} = \sum_{(hi)\in s_1} \tilde{w}_{1hi} y_{1hi}$$

$$\hat{Y}_{\mathrm{CAL2}} = \sum_{(hij)\in s_2} \tilde{w}_{12hij} y_{2hij}$$

où les poids de calage de premier degré, $\tilde{w}_{1hi}$, et les poids de calage de premier et de deuxième degrés combinés, $\tilde{w}_{12hij}$, sont donnés par :

$$\tilde{w}_{1hi} = w_{1hi}\left(1 + \left(\boldsymbol{X}_1 - \sum_{(hi)\in s_1} w_{1hi}\boldsymbol{x}_{1hi}\right)^T \left(\sum_{(hi)\in s_1} w_{1hi}\boldsymbol{x}_{1hi}\boldsymbol{x}_{1hi}^T\right)^{-1}\boldsymbol{x}_{1hi}\right)$$

$$\tilde{w}_{12hij} = w_{1hi}w_{2hij}\left(1 + \left(\boldsymbol{X}_2 - \sum_{(hij)\in s_2} w_{1hi}w_{2hij}\boldsymbol{x}_{2hij}\right)^T \left(\sum_{(hij)\in s_2} w_{1hi}w_{2hij}\boldsymbol{x}_{2hij}\boldsymbol{x}_{2hij}^T\right)^{-1}\boldsymbol{x}_{2hij}\right).$$

La méthode du bootstrap rééchelonné à plusieurs degrés peut alors être modifiée facilement de manière semblable pour traiter ces estimateurs par calage en remplaçant l'étape d) de la procédure de la façon suivante :

d) Calculer les poids bootstrap calés de premier et de deuxième degrés de la même manière que les poids calés de premier et de deuxième degrés :

$$\tilde{w}^*_{1hi} = w^*_{1hi}\left(1 + \left(\boldsymbol{X}_1 - \sum_{(hi)\in s_1} w^*_{1hi}\boldsymbol{x}_{1hi}\right)^T \left(\sum_{(hi)\in s_1} w^*_{1hi}\boldsymbol{x}_{1hi}\boldsymbol{x}_{1hi}^T\right)^{-1}\boldsymbol{x}_{1hi}\right)$$

$$\tilde{w}^*_{12hij} = w^*_{1hi}w^*_{2hij}\left(1 + \left(\boldsymbol{X}_2 - \sum_{(hij)\in s_2} w^*_{1hi}w^*_{2hij}\boldsymbol{x}_{2hij}\right)^T \left(\sum_{(hij)\in s_2} w^*_{1hi}w^*_{2hij}\boldsymbol{x}_{2hij}\boldsymbol{x}_{2hij}^T\right)^{-1}\boldsymbol{x}_{2hij}\right).$$

Les estimations bootstrap calées de premier et de deuxième degrés sont calculées de la façon suivante :

$$\hat{Y}^*_{\mathrm{CAL1}} = \sum_{(hi)\in s_1} \tilde{w}^*_{1hi} y_{1hi}$$

$$\hat{Y}^*_{\mathrm{CAL2}} = \sum_{(hij)\in s_2} \tilde{w}^*_{12hij} y_{2hij}.$$

Cette procédure peut être adaptée facilement à n'importe quel type de calage et étendue à n'importe quel nombre de degrés. Cette modification du bootstrap rééchelonné tient compte des ajustements apportés aux poids de sondage à cause du calage. Idéalement, tous les ajustements des poids de sondage, y compris ceux dus à la non-réponse et à la couverture insuffisante de la population devraient également être appliqués aux poids bootstrap.

4. Étude par simulation

Pour examiner la performance de l'estimateur de variance bootstrap rééchelonné à plusieurs degrés, nous avons procédé à une étude par simulation Monte Carlo. La simulation était limitée à l'échantillonnage stratifié à deux degrés et était fondée sur dix populations artificielles, qui étaient chacune subdivisée en $H = 5$ strates, avec $N_{1h} = 50$ unités de premier degré dans chaque strate et $N_{2hi} = 40$ unités de deuxième degré dans chaque unité de premier degré.

En premier lieu, nous avons produit la variable auxiliaire de premier degré x_{1hi} pour chaque unité de premier degré i dans la strate h à partir de la loi normale $N(\mu_{x1h}, (1-\rho_{x1b})\,\sigma^2_{x1b}/\rho_{x1b})$. En deuxième lieu, nous avons produit la variable auxiliaire de deuxième degré, x_{2hij}, et les variables cibles de deuxième degré, y_{2hij} et z_{2hij}, pour chaque unité de deuxième degré j à l'intérieur de l'unité de premier degré i dans la strate h à partir de la loi normale multivariée $N_3(\boldsymbol{\mu}_{2hi}, \boldsymbol{\Sigma}_{2hi})$, où $\boldsymbol{\mu}_{2hi}$ est le vecteur des moyennes :

$$\boldsymbol{\mu}_{2hi} = \begin{bmatrix} \mu_{x2hi} \\ \mu_{y2hi} \\ \mu_{z2hi} \end{bmatrix}$$

avec $\mu_{x2hi} = \mu_{y2hi} = \mu_{z2hi} = x_{1hi}$, et $\boldsymbol{\Sigma}_{2hi}$ est la matrice de variance-covariance :

$$\boldsymbol{\Sigma}_{2hi} = \begin{bmatrix} \sigma^2_{x2hi} & \rho_{xy2hi}\sigma_{x2hi}\sigma_{y2hi} & \rho_{xz2hi}\sigma_{x2hi}\sigma_{z2hi} \\ \rho_{xy2hi}\sigma_{x2hi}\sigma_{y2hi} & \sigma^2_{y2hi} & \rho_{yz2hi}\sigma_{y2hi}\sigma_{z2hi} \\ \rho_{xz2hi}\sigma_{x2hi}\sigma_{z2hi} & \rho_{yz2hi}\sigma_{y2hi}\sigma_{z2hi} & \sigma^2_{z2hi} \end{bmatrix}$$

avec $\sigma^2_{x2hi} = \sigma^2_{y2hi} = \sigma^2_{z2hi} = (1-\rho_{w2hi})\,\sigma^2_{w2hi}/\rho_{w2hi}$.

Les valeurs des paramètres que nous avons maintenues constantes dans les dix populations étaient $\mu_{x1h} = 25 \times (h+1)$, $\sigma^2_{b1h} = 10$, $\sigma^2_{w2hi} = 100$, $\rho_{xy2hi} = \rho_{xz2hi} = 0{,}75$ et $\rho_{yz2hi} = 0{,}50$. Les valeurs des paramètres que nous avons fait varier dans les dix populations étaient f_{1h}, les fractions d'échantillonnage de premier degré, f_{2hi}, les fractions d'échantillonnage de deuxième degré, ρ_{b1h} et ρ_{w2hi}. Les valeurs de ces paramètres sont présentées au tableau 1.

Tableau 1
Caractéristiques des populations simulées

	f_{1h}	f_{2hi}	ρ_b	ρ_w
Pop. I	0,1	0,1	0,75	0,75
Pop. II	0,1	0,1	0,25	0,75
Pop. III	0,1	0,5	0,75	0,75
Pop. IV	0,1	0,5	0,25	0,75
Pop. V	0,1	0,5	0,25	0,25
Pop. VI	0,5	0,1	0,75	0,75
Pop. VII	0,5	0,1	0,75	0,25
Pop. VIII	0,5	0,1	0,25	0,25
Pop. IX	0,3	0,3	0,75	0,25
Pop. X	0,3	0,3	0,25	0,25

Les paramètres d'intérêt dans l'étude par simulation étaient la moyenne de population, μ_y, le ratio de population, $R_{yz} = \mu_y/\mu_z$, le coefficient de corrélation de population, $\rho_{yz} = \sigma_{yz}/\sigma_y\sigma_z$, le coefficient de régression de population, $\beta_{yz} = \sigma_{yz}/\sigma^2_y$ et la médiane de population, M_y.

Afin d'estimer ces paramètres d'intérêt en utilisant les estimateurs de variance bootstrap à plusieurs degrés, nous avons tiré un total de $S = 20\,000$ échantillons aléatoires simples à deux degrés indépendants sans remise dans chacune des dix populations artificielles. En outre, nous avons tiré un total général de $T = 100\,000$ échantillons aléatoires simples à deux degrés indépendants sans remise dans chacune des dix populations artificielles afin d'estimer les variances de population réelles pour les paramètres d'intérêt. Pour calculer les estimateurs de variance bootstrap à plusieurs degrés, nous avons utilisé $B = 100$ échantillons bootstrap.

Pour comparer l'exactitude des estimateurs de variance bootstrap à plusieurs degrés, nous nous sommes servis du biais relatif (BR) et de la racine relative de l'erreur quadratique moyenne (RREQM). Nous avons calculé ces mesures comme il suit :

$$\mathrm{BR} = \frac{1}{\hat{\mathrm{Var}}(\hat{Y})}\left[\frac{1}{S}\sum_{s=1}^{S}(\mathrm{Var}_*(\hat{Y}_s) - \hat{\mathrm{Var}}(\hat{Y}))\right]$$

$$\mathrm{RREQM} = \frac{1}{\hat{\mathrm{Var}}(\hat{Y})}\sqrt{\frac{1}{S}\sum_{s=1}^{S}(\mathrm{Var}_*(\hat{Y}_s) - \hat{\mathrm{Var}}(\hat{Y}))^2}$$

où $\hat{\mathrm{Var}}(\hat{Y}) = T^{-1}\sum_{t=1}^{T}(\hat{Y}_t - Y)^2$ est la variance de population réelle estimée, et $\mathrm{Var}_*(\hat{Y}_s)$ est l'estimateur de variance bootstrap à plusieurs degrés pour le s^{e} échantillon de simulation.

Nous avons comparé l'estimateur de variance bootstrap rééchelonné à plusieurs degrés (EBRP) à l'estimateur de variance bootstrap rééchelonné à un degré (EBRU) et à l'estimateur de variance bootstrap bernoullien général (EBB) proposé par Funaoka et coll. (2006), pour des échantillons bootstrap utilisant les poids d'estimation non calés, $w_{hij} = w_{1hi}w_{2hij}$. Les tableaux 2 et 3 donnent le biais relatif et la racine relative de l'erreur quadratique moyenne des estimateurs EBRP, EBRU et EBB en utilisant des poids d'estimation non calés pour les dix populations artificielles.

Dans le cas de fonctions linéaires, telles que les moyennes, et de fonctions non linéaires, telles que les ratios, les coefficients de corrélation et les coefficients de régression, l'EBRP donne de meilleurs résultats que l'EBRU et l'EBB en ce qui concerne le biais relatif et la racine relative de l'erreur quadratique moyenne. Alors que l'EBRP donne systématiquement de bons résultats dans les dix populations artificielles, l'EBRU ne le fait que pour les populations artificielles III, IV et V, dans lesquelles les fractions d'échantillonnage de premier degré sont faibles ($f_{1h} = 0{,}1$) et celles de deuxième degré, grandes ($f_{2hi} = 0{,}5$), et l'EBB ne donne de bons résultats que dans les populations artificielles VI, VII et VIII, dans lesquelles les fractions d'échantillonnage de premier degré sont grandes ($f_{1h} = 0{,}5$) et celles de deuxième degré, faibles ($f_{2hi} = 0{,}1$). Ces fractions d'échantillonnage sont semblables aux fractions d'échantillonnage de premier degré et de deuxième degré utilisées dans l'étude par simulation présentée dans Funaoka et coll. (2006). Les divers niveaux de corrélation entre les unités de premier degré, et entre les unités de deuxième degré à l'intérieur des unités de premier degré, dont il a été tenu compte en faisant varier les paramètres ρ_b et ρ_w, ont peu d'effet sur la performance des estimateurs de variance.

Tableau 2
Biais relatif (%) des estimateurs de variance

	Moyenne (μ_y)			Moyenne (μ_z)			Ratio (R_{yz})		
	EBRP	**EBRU**	**EBB**	**EBRP**	**EBRU**	**EBB**	**EBRP**	**EBRU**	**EBB**
Pop. I	-0,28	-6,73	27,10	0,42	-6,63	27,32	0,00	-9,07	36,22
Pop. II	-0,05	-2,21	11,83	0,59	-1,64	12,54	-0,43	-9,26	36,40
Pop. III	-0,79	-2,63	3,62	-0,93	-2,66	3,40	-0,17	-5,30	5,19
Pop. IV	-0,23	-0,52	3,60	-0,18	-0,46	3,61	0,53	-4,65	5,98
Pop. V	0,15	-1,60	4,55	0,15	-1,64	4,54	0,52	-4,85	6,41
Pop. VI	0,70	-39,18	-0,34	0,65	-39,36	-0,28	1,57	-46,40	1,30
Pop. VII	0,19	-46,19	-0,26	-0,06	-46,48	-0,57	-0,27	-48,19	-0,73
Pop. VIII	0,37	-38,62	-0,41	0,23	-39,36	-0,46	-0,26	-47,93	-0,62
Pop. IX	0,42	-20,85	-7,76	-0,51	-20,03	-8,41	0,13	-23,13	-8,87
Pop. X	-0,56	-12,35	-6,08	0,70	-10,87	-6,93	-0,72	-23,70	-9,51
	Coefficient de corrélation (ρ_{yz})			**Coefficient de régression (β_{yz})**			**Médiane (M_y)**		
	EBRP	**EBRU**	**EBB**	**EBRP**	**EBRU**	**EBB**	**EBRP**	**EBRU**	**EBB**
Pop. I	-2,31	-10,23	32,17	-0,08	-9,05	36,41	19,04	-19,86	33,21
Pop. II	-1,51	-8,41	29,65	0,05	-8,74	36,41	19,29	2,42	40,85
Pop. III	0,36	-4,37	5,69	0,05	-5,12	5,42	7,50	4,28	9,72
Pop. IV	2,18	-0,60	7,17	0,28	-5,05	5,70	17,40	16,17	34,37
Pop. V	0,79	-2,71	5,95	0,26	-5,40	6,34	8,29	4,78	11,49
Pop. VI	0,32	-46,67	0,14	0,89	-46,59	0,69	13,57	-33,56	9,15
Pop. VII	-0,07	-46,78	-0,39	-0,21	-47,85	-0,60	14,68	-38,16	11,86
Pop. VIII	0,31	-44,25	-0,27	-0,09	-47,54	-0,55	2,09	-38,90	-0,64
Pop. IX	-0,93	-23,02	-9,30	-0,20	-23,48	-9,20	8,08	-17,23	-1,97
Pop. X	-0,82	-19,35	-8,24	-1,02	-23,89	-9,75	2,10	-13,84	-5,46

Nota : L'erreur de simulation la plus importante sur les biais relatifs était inférieure à 0,7 %.

Dans le cas de statistiques non lisses, telles que les médianes, l'EBRP et l'EBB ont tous deux tendance à surestimer les variances de population réelles, tandis que l'EBRU a tendance à les sous-estimer. En outre, la racine relative de l'erreur quadratique moyenne est jusqu'à trois fois plus élevée pour les médianes que pour les moyennes. L'EBRP donne de meilleurs résultats que l'EBB pour les populations artificielles I à V, dans lesquelles les fractions d'échantillonnage de premier degré sont petites (f_{1h} = 0,1), tandis que l'EBB donne d'un peu meilleurs résultats que l'EBRP pour les populations artificielles VI à X, dans lesquelles les fractions d'échantillonnage de premier degré sont plus grandes (f_{1h} = 0,3 ou 0,5).

Cette surestimation du bootstrap rééchelonné à plusieurs degrés pour les médianes concorde avec les résultats présentés dans les études de Kovar et coll. (1988) et de Rao et coll. (1992) pour le bootstrap rééchelonné à un degré. Il convient de souligner que le bootstrap rééchelonné original, introduit par Rao et Wu (1988), a été élaboré uniquement pour des statistiques lisses, telles que les moyennes, les ratios, ainsi que les coefficients de corrélation et de régression.

Nous avons examiné l'EBRP en utilisant les poids d'estimation calés, $\tilde{w}_{hij} = w_{1hi}\tilde{w}_{2hij}$, qui satisfont la contrainte de calage $\sum_{(hij)\in s_2} w_{1hi}\tilde{w}_{2hij}x_{2hij} = X_2$, où $X_2 = \sum_{(hij)\in U} x_{2hij}$ est le total de population pour la variable auxiliaire de deuxième degré. Le tableau 4 donne le biais relatif et la racine relative de l'erreur quadratique moyenne de l'EBRP avec utilisation des poids d'estimation calés pour les quatre populations artificielles II, IV, VII et IX.

Les biais relatifs et les racines relatives de l'erreur quadratique moyenne de l'EBRP avec des poids d'estimation calés sont semblables à ceux obtenus en utilisant les poids d'estimation non calés.

5. Conclusion

Le présent article décrit l'extension de la méthode du bootstrap rééchelonné à l'échantillonnage à plusieurs degrés où les unités sont sélectionnées par échantillonnage aléatoire simple sans remise à chaque degré. Sous la méthode du bootstrap rééchelonné à plusieurs degrés proposée, les échantillons bootstrap sont sélectionnés sans remise et des facteurs d'échelle sont appliqués aux poids de sondage. Cette méthode proposée est relativement simple à mettre en œuvre et requiert un nombre considérablement plus faible de générations de nombres aléatoires que la méthode du bootstrap bernoullien général à plusieurs degrés. La méthode proposée convient aussi pour une vaste gamme de méthodes de repondération, y compris le calage, et d'ajustements pour tenir compte de la non-réponse et de la couverture insuffisante de la population. En outre, les résultats de l'étude par simulation Monte Carlo indiquent que le bootstrap

rééchelonné à plusieurs degrés donne de nettement meilleurs résultats que le bootstrap rééchelonné à un degré et que le bootstrap bernoullien à plusieurs degrés pour les statistiques lisses, telles que les moyennes, les ratios et les coefficients de corrélation et de régression.

Annexe

À la présente annexe, nous montrons que l'estimateur de variance bootstrap rééchelonné à plusieurs degrés pour l'échantillonnage stratifié à trois degrés se réduit à l'estimateur de variance à trois degrés standard sans biais (2.1) dans le cas où $\hat{\theta}$ est l'estimateur linéaire, $\hat{Y}^* = \sum_{h=1}^{H}\sum_{i=1}^{n_{1h}}\sum_{j=1}^{n_{2hi}}\sum_{k=1}^{n_{3hij}} w_{hijk}^* y_{hijk}$. L'estimateur de variance bootstrap de $\hat{Y}^*$ est donné par :

$$\mathrm{Var}(\hat{Y}^*) = \mathrm{Var}_{1*}(E_{2*}(E_{3*}(\hat{Y}^*)))$$

$$+ E_{1*}(\mathrm{Var}_{2*}(E_{3*}(\hat{Y}^*))) + E_{1*}(E_{2*}(\mathrm{Var}_{3*}(\hat{Y}^*))).$$

Tableau 3
Racine relative de l'erreur quadratique moyenne (%) des estimateurs de variance

	Moyenne (μ_y)			Moyenne (μ_z)			Ratio (R_{yz})		
	EBRP	**EBRU**	**EBB**	**EBRP**	**EBRU**	**EBB**	**EBRP**	**EBRU**	**EBB**
Pop. I	31,9	32,1	44,6	31,7	31,8	44,4	31,8	32,3	51,4
Pop. II	33,9	33,8	38,2	33,4	33,3	38,1	32,2	32,9	51,7
Pop. III	33,8	33,8	35,9	33,0	33,0	35,0	33,0	33,1	35,1
Pop. IV	35,3	35,3	37,4	35,2	35,2	37,3	32,8	32,8	35,0
Pop. V	32,0	31,9	34,2	34,3	34,2	36,5	33,0	33,1	35,6
Pop. VI	16,4	40,7	16,5	16,4	40,9	16,5	16,5	47,5	16,5
Pop. VII	16,1	47,4	16,4	16,1	47,8	16,4	16,1	49,0	16,1
Pop. VIII	16,3	40,3	16,5	16,7	40,9	16,3	16,2	48,8	16,1
Pop. IX	19,2	26,7	20,0	19,3	26,3	20,0	19,2	28,6	20,2
Pop. X	19,8	22,4	20,2	19,9	21,6	20,3	19,1	29,0	20,6
	Coefficient de corrélation (ρ_{yz})			**Coefficient de régression (β_{yz})**			**Médiane (M_y)**		
	EBRP	**EBRU**	**EBB**	**EBRP**	**EBRU**	**EBB**	**EBRP**	**EBRU**	**EBB**
Pop. I	47,8	46,3	68,7	36,6	37,2	55,3	88,7	80,1	89,8
Pop. II	48,4	47,1	66,6	37,4	37,9	55,6	93,4	91,0	115,9
Pop. III	35,9	35,6	38,4	37,5	37,6	39,9	80,4	80,3	81,1
Pop. IV	42,6	42,2	45,4	38,0	38,0	40,3	97,5	96,6	127,3
Pop. V	40,3	40,0	43,3	37,3	37,5	40,1	31,5	30,7	63,3
Pop. VI	21,6	48,4	21,7	16,9	47,8	17,0	55,3	51,4	52,0
Pop. VII	21,4	48,4	21,3	16,9	49,0	16,8	53,5	51,4	51,4
Pop. VIII	21,6	46,3	21,5	17,0	48,6	16,9	41,8	49,7	40,3
Pop. IX	21,5	29,4	22,5	20,5	29,9	21,6	46,1	42,7	42,7
Pop. X	22,7	27,8	23,4	20,6	30,2	21,9	39,7	38,9	37,9

Tableau 4
Biais relatif (%) et racine relative de l'erreur quadratique moyenne (%) de l'estimateur de variance bootstrap rééchelonné

	μ_y	R_{yz}	ρ_{yz}	β_{yz}	M_y
			Biais relatif (%)		
Pop. II	-0,42	-0,29	-1,51	-0,08	20,98
Pop. IV	0,40	0,49	1,83	0,08	18,28
Pop. VII	-0,22	-0,24	-0,03	-0,28	12,24
Pop. IX	0,62	0,19	-1,00	-0,20	7,24
			Racine relative de l'erreur quadratique moyenne (%)		
Pop. II	32,6	32,4	48,4	37,3	97,8
Pop. IV	32,8	32,8	44,6	37,9	99,4
Pop. VII	16,2	16,1	21,5	16,9	50,0
Pop. IX	19,1	19,2	21,6	20,5	43,8

Nota : L'erreur de simulation la plus importante sur les biais relatifs était inférieure à 0,6 %.

Au moyen de résultats standard sur l'espérance et la variance en ce qui concerne l'échantillonnage bootstrap EASSR et de certaines opérations algébriques fastidieuses mais simples, les composantes de l'estimateur de variance bootstrap sont données ci-après. L'espérance conditionnelle de $\hat{Y}^*$ sachant s_3 est

$$E_{3*}(\hat{Y}^*) =$$

$$\sum_{h=1}^{H}\sum_{i=1}^{n_{1h}}\sum_{j=1}^{n_{2hi}} w_{1i}w_{2ij}\left(1-\lambda_{1h}+\lambda_{1h}\frac{n_{1h}}{n^*_{1h}}\delta_{1hi}\right.$$

$$\left.-\lambda_{2hi}\sqrt{\frac{n_{1h}}{n^*_{1h}}}\delta_{1hi}+\lambda_{2hi}\sqrt{\frac{n_{1h}}{n^*_{1h}}}\delta_{1hi}\frac{n_{2hi}}{n^*_{2hi}}\delta_{2hij}\right)\hat{Y}_{ij}$$

et la variance conditionnelle de $\hat{Y}^*$ sachant s_3 est

$$\mathrm{Var}_{3*}(\hat{Y}^*) =$$

$$\sum_{h=1}^{H}\frac{N_{1h}}{n^*_{1h}}\sum_{i=1}^{n_{1h}}\frac{N_{2hi}}{n^*_{2hi}}\sum_{j=1}^{n_{2hi}}\delta_{1hi}\,\delta_{2hij}\,\frac{N^2_{3hij}}{n_{3hij}}\,(1-f_{3hij})\,s^2_{3hij}.$$

Les espérances conditionnelles de $E_{3*}(\hat{Y}^*)$ et de $\mathrm{Var}_{3*}(\hat{Y}^*)$ sachant s_2 sont

$$E_{2*}(E_{3*}(\hat{Y}^*)) = \sum_{h=1}^{H}\sum_{i=1}^{n_{1h}} w_{1i}(1-\lambda_{1h}+\lambda_{1h}\frac{n_{1h}}{n^*_{1h}}\delta_{1hi})\,\hat{Y}_{hi}$$

$$E_{2*}(\mathrm{Var}_{3*}(\hat{Y}^*)) =$$

$$\sum_{h=1}^{H}\frac{N_{1h}}{n^*_{1h}}\sum_{i=1}^{n_{1h}}\frac{N_{2hi}}{m_{2hi}}\sum_{j=1}^{n_{2hi}}\delta_{1hi}\,\frac{N^2_{3hij}}{n_{3hij}}\,(1-f_{3hij})\,s^2_{3hij}$$

et la variance conditionnelle de $E_{3*}(\hat{Y}^*)$ sachant s_2 est

$$\mathrm{Var}_{2*}(E_{3*}(\hat{Y}^*)) = \sum_{h=1}^{H}\frac{N_{1h}}{n^*_{1h}}\sum_{i=1}^{n_{1h}}\delta_{1hi}\,\frac{N^2_{2hi}}{n_{2hi}}\,(1-f_{2hi})\,s^2_{2hi}.$$

Enfin, les espérances conditionnelles de $E_{2*}(\mathrm{Var}_{3*}(\hat{Y}^*))$ et de $\mathrm{Var}_{2*}(E_{3*}(\hat{Y}^*))$ sachant s_1 sont

$$E_{1*}(E_{2*}(\mathrm{Var}_{3*}(\hat{Y}^*))) =$$

$$\sum_{h=1}^{H}\frac{N_{1h}}{n_{1h}}\sum_{i=1}^{n_{1h}}\frac{N_{2hi}}{n_{2hi}}\sum_{j=1}^{n_{2hi}}\frac{N^2_{3hij}}{n_{3hij}}\,(1-f_{3hij})\,s^2_{3hij}$$

$$E_{1*}(\mathrm{Var}_{2*}(E_{3*}(\hat{Y}^*))) = \sum_{h=1}^{H}\frac{N_{1h}}{n_{1h}}\sum_{i=1}^{n_{1h}}\frac{N^2_{2hi}}{n_{2hi}}\,(1-f_{2hi})\,s^2_{2hi}$$

qui sont égales aux troisième et deuxième termes de (2.1), respectivement, et la variance conditionnelle de $E_{2*}(E_{3*}(\hat{Y}^*))$ sachant s_1 est

$$\mathrm{Var}_{1*}(E_{2*}(E_{3*}(\hat{Y}^*))) = \sum_{h=1}^{H}\frac{N^2_{1h}}{n_{1h}}\,(1-f_{1h})\,s^2_{1h}$$

qui est égale au premier terme de (2.1).

Bibliographie

Australian Bureau of Statistics (ABS) (2008). Employee Earnings and Hours, Numéro de catalogue 6306.0.

Bickel, P.J., et Freedman, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Annals of Statistics*, 12, 470-482.

Chipperfield, J., et Preston, J. (2007). Bootstrap efficace pour les enquêtes-entreprises. *Techniques d'enquête*, 33, 187-193.

Estevao, V., et Särndal, C.-E. (2006). Survey estimates by calibration on complex auxiliary information. *Revue Internationale de Statistique*, 74, 127-147.

Funaoka, F., Saigo, H., Sitter, R.R. et Toida, T. (2006). Bootstrap de type Bernoulli pour l'échantillonnage stratifié à plusieurs degrés. *Techniques d'enquête*, 32, 169-175.

Gross, S. (1980). Median estimation in sample surveys. *Proceedings of the Section on Survey Research Methods*, 181-184.

Kovar, J.G., Rao, J.N.K. et Wu, C.F.J. (1988). Bootstrap and other methods to measure errors in survey estimates. *Canadian Journal of Statistics*, 16, 25-45.

McCarthy, P.J., et Snowden, C.B. (1985). The bootstrap and finite population sampling. Vital and Health Statistics (Séries 2 No 95), Public Health Service Publication 85-1369, Washington, DC : U.S. Government Printing Office.

Rao, J.N.K., et Wu, C.F.J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83, 231-241.

Rao, J.N.K., Wu, C.F.J. et Yue, K. (1992). Quelques travaux récents sur les méthodes de rééchantillonnage applicables aux enquêtes complexes. *Techniques d'enquête*, 18, 225-234.

Särndal, C.-E., Swenson, B. et Wretman, J. (1992). *Model Assisted Survey Sampling*. New York : Springer-Verlag.

Shao, J., et Tu, D. (1995). *The Jackknife and Bootstrap*. New York : Springer-Verlag.

Sitter, R.R. (1992a). A resampling procedure for complex survey data. *Journal of the American Statistical Association*, 87, 755-765.

Sitter, R.R. (1992b). Comparing three bootstrap methods for survey data. *Canadian Journal of Statistics*, 20, 135-154.

White, V., et Hayman J. (2006). Smoking behaviours of Australian secondary students in 2005**.** National Drug Strategy Monograph Séries No. 59. Canberra: Australian Government Department of Health and Ageing.

Yeo, D., Mantel, H. et Liu T.-P. (1999). Bootstrap variance estimation for the National Population Health Survey. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 778-783.