MÉTHODES DE RÉÉCHANTILLONNAGE POUR L'ESTIMATION DE VARIANCE EN SONDAGE

A. C. DAVISON *, S. SARDY

RÉSUMÉ

Nous passons en revue des techniques de rééchantillonnage utilisées pour l'estimation de variance en sondage utilisant le calage et l'imputation. Les techniques de rééchantillonnage considérées sont basées sur la linéarisation, le jackknife, les répliques équilibrées répétées, et le bootstrap. Notre objectif est d'obtenir des conclusions pratiques basées sur des considérations théoriques et des comparaisons empiriques. Parmi les méthodes de rééchantillonage, le bootstrap semble être préferable par sa flexibilité d'utilisation et sa stabilité. Pour les sondages massifs, la linéarisation a l'avantage de diminuer le fardeau de calcul.

Mots clés : Bootstrap, calage, imputation, jackknife, linéarisation, non-réponse, valeurs manquantes.

ABSTRACT

We review resampling techniques used for variance estimation in sample surveys under calibration and imputation of missing values. The techniques considered are based on linearisation, the jackknife, balanced repeated replication, and the bootstrap. Our purpose is to give practical recommendations based on theoretical considerations and empirical comparisons. Among the resampling methods, the flexibility and stability of the bootstrap make it preferable, while for massive surveys linearisation methods can reduce the computational effort considerably.

Keywords : Bootstrap, calibration, imputation, jackknife, linearisation, missing values, non-response, sample survey.

1. Introduction

En sondage des formules exactes de variance ne sont disponibles que pour des classes d'estimateurs limitées, de sorte que des approximations sont couramment utilisées en pratique. Des approximations classiques sont dérivées à partir de variantes de la méthode delta, mais, lorsque l'échantillon est petit ou que l'estimateur est complexe, il est naturel de se poser la question de l'exactitude

* Institut de Mathématiques, Ecole Polytechnique Fédérale de Lausanne, IMA-FSB-EPFL, Station 8, 1015 Lausanne, Suisse.
Anthony.Davison@epfl.ch, Sylvain.Sardy@epfl.ch

du développement sur lequel repose la validité des formules résultantes, et de rechercher d'autres approches. Une classe importante d'alternatives est basée sur les procédures de rééchantillonnage, qui impliquent des calculs répétés sur l'échantillon de départ. Les plus importantes sont d'abord le bootstrap, introduit par Efron (1979), ensuite le jackknife (Quenouille, 1949a ; Tukey, 1958), introduit au départ pour l'estimation du biais dans l'analyse de séries temporelles, et enfin le semi-échantillonnage (McCarthy, 1969), qui puise ses racines dans des travaux entrepris durant la fin des années 1950 au US Census Bureau (Hartigan, 1969 ; Hall, 2003 ; Wolter, 1985, §3.1). Le bootstrap est la procédure la plus flexible et la plus puissante, mais reste encore trop peu utilisé en sondage.

Le but de cet article est de passer en revue des travaux publiés sur les applications de ces idées en enquêtes, avec un accent sur les possibilités d'implémentation pratique plutôt que sur la théorie. Le chapitre 2 décrit brièvement les types d'estimateurs considérés et les procédures d'ajustement et d'imputation qui peuvent être utilisées pour les modifier. Les chapitres 3 à 6 décrivent les approches de la linéarisation, du jackknife, des répliques équilibrées répétées, et du bootstrap. Des comparaisons numériques sont passées en revue dans le chapitre 7, et nous concluons sur une brève discussion.

2. Préliminaires

2.1. Paramètres et estimateurs

Afin d'établir la notation nous considérons initialement le cas d'une réponse complète pour une méthode d'échantillonage stratifié en H strates à un degré sans remise. Soit n_h le nombre d'unités échantillonnées à partir des N_h unités de population dans la strate h, et soit π_{hi} la probabilité d'inclusion pour l'unité i de cette strate. Dans des enquêtes sur les ménages, cette unité peut consister en une grappe d'individus, auquel cas nous supposons que la réponse de l'unité d'intérêt résume la grappe. Ainsi la taille de la population totale est $N = \sum_{h=1}^{H} N_h$ et le nombre d'unités échantillonnées est $n = \sum_{h=1}^{H} n_h$. Soit x_{hi} et y_{hi} les variables mesurées sur les unités, où y_{hi} est la réponse d'intérêt scalaire et x_{hi} est un vecteur $q \times 1$ de variables auxiliaires.

Les paramètres de la population finie peuvent être classifiés en deux grands groupes. Le premier contient des quantités qui sont des fonctions lisses des réponses de la population finie, comme le total $\tau = \sum_{h=1}^{H} \sum_{i=1}^{N_h} y_{hi}$, le rapport de deux totaux de populations pour deux variables différentes,

$$\phi = \tau_1/\tau_2, \tag{1}$$

ou le changement dans le rapport entre deux occasions d'échantillonage différents,

$$\delta = \phi_1 - \phi_2. \tag{2}$$

L'autre groupe comprend les fonctions non lisses, comme la médiane,

$$y_{0.5} = \text{médiane}\left\{y_{hi} : i = 1, \ldots, N_h, h = 1, \ldots, H\right\},$$

et d'autres quantiles. Un exemple est $F(ay_b)$, où F est la fonction de répartition du revenu des ménages ; $a = b = 0.5$ donne la proportion de ménages dont le revenu est en-dessous de la moitié de la médiane, ce qui est un indicateur de seuil de pauvreté.

L'estimation de paramètres de la population finie est basée sur les données de n unités échantillonnées et sur leur probabilité d'inclusion dans le plan d'échantillonnage. Le plus important estimateur d'un total est l'estimateur de Horvitz–Thompson

$$\hat{\tau} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \omega_{hi} y_{hi} = \omega^{\mathrm{T}} y, \tag{3}$$

où y est le vecteur $n \times 1$ de réponses échantillonnées et ω est le vecteur $n \times 1$ de leurs poids $\omega_{hi} = 1/\pi_{hi}$, les probabilités d'inclusion inverses. La variance exacte de $\hat{\tau}$ implique les probabilités d'inclusion conjointes pour les différentes unités. Elle peut être utilisée pour développer des formules de variance pour des estimateurs de paramètres tel que (1) et (2) ci-dessus, mais des complications surviennent lorsque les poids eux-mêmes sont aléatoires, ou lorsqu'une partie de la réponse n'est pas disponible.

Ci-dessous θ et $\hat{\theta}$ notent un paramètre quelconque de la population finie et son estimateur, respectivement.

2.2. Calage

Souvent les totaux de la population sont connus pour certaines des variables auxiliaires mesurées, telles que des indicateurs de facteur provenant par exemple de tableaux de contingence. Cette information peut être utilisée pour augmenter la précision de l'estimation par calage (calibration en anglais).

Soit X la matrice $n \times q$ contenant les q variables auxiliaires. En supposant que q_C marges des q variables auxiliaires sont connues, avec $q_C \leqslant q$, on appelle c le vecteur $q_C \times 1$ de marges connues, et X_C la sous-matrice $n \times q_C$ de variables auxiliaires correspondantes. Utilisant l'estimation d'un total pour illustrer l'idée du calage, la qualité de l'estimateur de Horvitz–Thompson peut être amélioré en calant les poids w_{hi} en (3) de façon à être proche des poids originaux ω selon une certaine métrique G, sous les contraintes que les variables auxiliaires pondérées concordent avec les marges (Deville et Särndal, 1992) :

$$\min_{w_{hi}} \sum_{h=1}^{H} \sum_{i=1}^{n_h} \omega_{hi} G(w_{hi}/\omega_{hi}), \text{ tel que } \sum_{h=1}^{H} \sum_{i=1}^{n_h} w_{hi} x_{Chi} = c.$$

Une mesure de distance communément utilisée dans la pratique est la métrique ℓ_2, $G(x) = (x-1)^2/2$; dans ce cas les poids calés valent

$$w = \omega + \Omega X_C (X_C^{\mathrm{T}} \Omega X_C)^{-1} (c - X_C^{\mathrm{T}} \omega), \tag{4}$$

où Ω désigne la matrice diagonale dont les éléments sont les ω_{hi}. Ainsi l'estimateur calé de Horvitz–Thompson pour le total s'écrit

$$\begin{aligned} \hat{\tau} &= w^{\mathrm{T}} y = \omega^{\mathrm{T}} y + (c - X_C^{\mathrm{T}}\omega)^{\mathrm{T}} (X_C^{\mathrm{T}} \Omega X_C)^{-1} X_C^{\mathrm{T}} \Omega y \\ &= \omega^{\mathrm{T}} y + (c - X_C^{\mathrm{T}}\omega)^{\mathrm{T}} \hat{\gamma}, \end{aligned} \tag{5}$$

$\hat{\gamma}$ étant l'estimateur de régression de y sur X_C avec poids Ω. Le deuxième terme de $\hat{\tau}$ est un ajustement aux poids ω qui tient compte de la différence entre $X_C^{\mathrm{T}}\omega$ et de la valeur de la population cible c des marges.

Un grand nombre d'autres mesures de distance entre w et ω peut être envisagé. Par exemple, la mesure multiplicative $G(x) = x \log x - x + 1$ (Deville *et al.*, 1993) garantit que les poids calés, comme les poids originaux, sont positifs. Cependant les résultats, en utilisant diverses métriques G, sont asymptotiquement équivalents (Deville et Särndal, 1992), et en pratique (Deville *et al.*, 1993) ont observé que, bien que les poids individuels peuvent varier beaucoup en fonction des méthodes, le choix de G a un impact relativement faible sur l'estimation ponctuelle et sur l'estimation de variance. Etant donné que le calcul de poids pour des mesures plus compliquées, telles que la mesure multiplicative, nécessite plus d'effort de calcul, nos recherches se concentrent surtout sur les poids calés (4) avec la métrique ℓ_2.

2.3. Imputation

Dans la pratique l'application de méthodes d'enquêtes amène rarement à des jeux de données complets. Il existe deux types principaux de non-réponses, la non-réponse totale ('unit non-response' en anglais), pour laquelle toutes les données pour une unité échantillonnée manquent, et la non-réponse partielle ('item non-response'), pour laquelle une partie des données de l'unité manque.

La non-réponse totale est souvent traitée par calage ou toute autre forme d'ajustement de poids, mais la non-réponse partielle peut être traitée par imputation des données manquantes en utilisant un modèle d'imputation ajusté sur les données complètes. L'imputation est dite propre (Rubin, 1996) si les estimateurs ponctuels et de variance sont approximativement non biaisés pour leurs analogues avec données complètes.

Les modèles d'imputation et les valeurs imputées sont de deux types : déterministe, ce qui signifie que le modèle impute une fonction déterministe des données observées ; et stochastique, ce qui signifie que le modèle impute une fonction aléatoire des données observées.

Une approche déterministe répandue pour l'imputation de réponses manquantes basée sur les vecteurs correspondants de variables auxiliaires x_{hi} est d'utiliser un modèle linéaire ou l'une de ses généralisations, tel que la régression robuste ou logistique. Tous ces estimateurs sont des M-estimateurs, et les équations normales correspondantes pour une estimation des paramètres β du modèle d'imputation choisi à travers les strates peuvent s'écrire comme

une équation vectorielle

$$\sum_{h=1}^{H}\sum_{i=1}^{n_h} x_{hi}\psi(y_{hi}, x_{hi}; \beta) = 0, \tag{6}$$

où ψ est la dérivée de la fonction de perte, dont la forme dépend de ce modèle ; pour la régression logistique les y_{hi} sont des variables indicatrices binaires et

$$\psi(y, x; \beta) = y - \frac{\exp(x^{\mathrm{T}}\beta)}{1 + \exp(x^{\mathrm{T}}\beta)}.$$

Posant $\psi(y, x; \beta) = y - x^{\mathrm{T}}\beta$ revient à considérer l'estimation par moindres carrés. Un autre cas particulier est l'imputation de rapport utilisant un scalaire x, obtenu en posant $\psi(y, x; \beta) = y - \beta x$. Un modèle d'imputation plus robuste est obtenu en utilisant la fonction robustifiée $\psi(y, x; \beta) = (y - x^{\mathrm{T}}\beta)\min(|y - x^{\mathrm{T}}\beta|, \xi)$ (Huber, 1981). La quantité ξ contrôle la robustesse de l'ajustement, avec $\xi \to \infty$ recouvrant le fragile estimateur des moindres carrés, et $\xi \to 0$ donnant une plus grande robustesse. Une fois que le M-estimateur du modèle linéaire a été calculé, une réponse manquante peut être prédite à l'aide du modèle et des variables auxiliaires correspondantes. Des explications similaires s'appliquent à la régression logistique dans le cas de réponses binaires.

Pour les calculs numériques illustratifs ci-dessous nous supposons soit un modèle linéaire commun à travers les strates,

$$y_{hi} = x_{hi}^{\mathrm{T}}\beta + \epsilon_{hi}, \tag{7}$$

soit un modèle linéaire différent pour chaque strate,

$$y_{hi} = x_{hi}^{\mathrm{T}}\beta_h + \epsilon_{hi}, \quad h = 1, \ldots, H, \tag{8}$$

où β et β_h sont les coefficients du modèle linéaire, et x_{hi}^{T} sont les variables auxiliaires de l'unité i dans la strate h. Nous ajustons alors ce modèle à ces individus pour lesquels à la fois x_{hi} et y_{hi} sont disponibles, et utilisons ensuite le modèle linéaire ajusté pour imputer les réponses prédites pour les individus pour lesquels y_{hi} manque. Soit $z_{hi} = I(y_{hi} \neq \mathtt{NA})$ la variable aléatoire indicatrice correspondant à la réponse observée (`NA` est l'acronyme de 'Not Available' utilisé par les logiciels pour indiquer une valeur manquante), soit $Z = \mathrm{diag}(z)$ la matrice diagonale $n \times n$ de ces indicatrices, soit X la matrice $n \times q$ des variables auxiliaires correspondantes à la fois aux répondants et aux non-répondants, et soit $\hat{y} = X\hat{\beta}$ le vecteur $n \times 1$ des valeurs ajustées par le modèle de régression utilisé pour l'imputation. Alors (4) implique que l'estimateur calé et imputé de Horvitz–Thompson s'écrit comme

$$\begin{aligned}\hat{\tau} &= w^{\mathrm{T}}\{Zy + (I - Z)\hat{y}\} = \omega^{\mathrm{T}}Zy + (c - X_C^{\mathrm{T}}\omega)^{\mathrm{T}}(X_C^{\mathrm{T}}\Omega X_C)^{-1}X_C^{\mathrm{T}}\Omega Zy \\ &\quad + \omega^{\mathrm{T}}(I - Z)X\hat{\beta} + (c - X_C^{\mathrm{T}}\omega)^{\mathrm{T}}(X_C^{\mathrm{T}}\Omega X_C)^{-1}X_C^{\mathrm{T}}\Omega(I - Z)X\hat{\beta}.\end{aligned} \tag{9}$$

D'autres M-estimateurs peuvent être implémentés de façon similaire.

Si le calcul de la variance de l'estimateur calé et imputé de Horvitz–Thompson est réalisé en considérant les réponses imputées $\hat{y}$ comme vraies, alors une sous-estimation de la variance en résulte. Aussi, la technique d'estimation de variance employée doit refléter l'inflation de variance due à l'imputation. Une technique simple mais intensive est d'utiliser le rééchantillonnage en traitant l'estimateur (calé et imputé) comme n'importe quel estimateur. Une autre puissante technique est l'imputation multiple (Rubin, 1987), où les formules standards sont calculées pour différents jeux de données pour lesquels les données manquantes sont imputées stochastiquement et combinées de manière à refléter l'effet de l'imputation. Cette approche a été sujette à controverse pour des enquêtes par certains auteurs ; voir par exemple Fay (1996). D'autres discusssions récentes peuvent être trouvées dans Zhang (2003) et Nielsen (2003).

Avec une imputation propre, les estimateurs ponctuels et de variance sont approximativement non biaisés. En pratique, il est impossible de savoir si une approche d'imputation est propre ou pas, parce que le mécanisme des données manquantes est inconnu. Puisque les modèles décrits sont simplement des approximations brutes d'une réalité complexe, Rubin (1996) recommande d'inclure autant de variables auxiliaires que possible, l'idée étant d'approximer cette réalité au mieux avec les données disponibles, même au prix d'une augmentation de la variance des estimations éventuelles.

2.4. Discussion

Pour des sondages simples, des livres tels que Cochran (1977) contiennent des formules donnant des estimations de variance d'une grande variété d'estimateurs. Cependant, la complexité des procédures modernes de sondage ne permet souvent pas de dériver des formules simples. Des méthodes de rééchantillonnage telles que le jackknife ou le bootstrap offrent d'excellentes alternatives (Droesbeke *et al.*, 1987). L'adaptation aux enquêtes des méthodes d'estimation de variance basées sur le rééchantillonnage nécessite cependant un soin particulier parce qu'il doit prendre en compte les structures de dépendances complexes induites par la méthode d'échantillonnage ainsi que par le calage et l'imputation.

3. Linéarisation

3.1. Formules de variance

L'approche générale de construction d'estimateurs de variance par linéarisation infinitésimale (Efron, 1982) permet l'obtention de formules de variance théorique pour des statistiques dérivables $\hat{\theta}$. Elle utilise les composantes de la fonction d'influence (Hampel *et al.*, 1986), qui peuvent être interprétées comme les dérivées de $\hat{\theta}$ par rapport aux poids associés aux observations individuelles. Ainsi pour l'échantillonnage stratifié aléatoire sans remise, si l_{hi} est la valeur d'influence empirique correspondant à la i-ème observation dans

la strate h, alors l'estimateur de la variance peut être exprimé comme

$$\sum_{h=1}^{H}(1-f_h)\frac{1}{(n_h-1)n_h}\sum_{i=1}^{n_h} l_{hi}^2, \tag{10}$$

où, pour tenir compte du plan d'échantillonnage, $f_h = n_h/N_h$ est la fraction d'échantillonnage dans la h-ème strate. Par exemple, Berger et Skinner (2003) dérivent les fonctions d'influence pour estimer la variance de la fonction $F(ay_b)$ de §2.1. Davison et Hinkley (1997, §2.11). donnent des détails sur ces développements et d'autres exemples de calculs, et Campbell (1980) décrit leur extension dans les situations d'une population finie. Dans la section suivante, nous donnons les fonctions d'influence empirique pour l'estimateur de Horvitz–Thompson calé et imputé.

La méthode de linéarisation infinitésimale peut en outre être combinée à la linéarisation en chaîne et à la méthode delta (Rao, 1988 ; Binder, 1996 ; Deveille, 1999). Elle genéralise la méthode de linéarisation de Taylor classique qui s'applique aux estimateurs s'exprimant comme fonction dérivable $\hat{\theta} = g(\hat{\varphi})$ d'un vecteur d'estimateurs linéaires $\hat{\varphi}$. Le développement en séries de Taylor de g à l'ordre un donne

$$\hat{\theta} = g(\hat{\varphi}) \doteq g(\varphi) + \nabla g(\varphi)^{\mathrm{T}}(\hat{\varphi}-\varphi),$$

où ∇g est le vecteur qui contient les dérivées de g par rapport aux éléments de φ. La variance de $\hat{\theta}$ peut ainsi être approximée par $\nabla g(\varphi)^{\mathrm{T}}\mathrm{var}(\hat{\varphi})\nabla g(\varphi)$, et l'estimateur de variance est obtenu en remplaçant φ par $\hat{\varphi}$ et en utilisant la matrice de covariance empirique pour $\mathrm{var}(\hat{\varphi})$. Pour des échantillons stratifiés, l'estimateur de variance s'écrit

$$v_L = \nabla g(\hat{\varphi})^{\mathrm{T}}\hat{\mathrm{var}}(\hat{\varphi})\nabla g(\hat{\varphi}) = \nabla g(\hat{\varphi})^{\mathrm{T}}\left(\sum_{h=1}^{H}\hat{V}_h\right)\nabla g(\hat{\varphi}), \tag{11}$$

où $\hat{V}_h$ est la contribution à la matrice de variance estimée pour $\hat{\varphi}$ de la h-ème strate. Sous des conditions faibles, Krewski et Rao (1981) montre que (11) est un estimateur convergent de la variance asymptotique de $\hat{\theta}$.

3.2. Estimateur de Horvitz–Thompson

Nous considérons maintenant l'estimateur de Horvitz–Thompson et donnons sa fonction d'influence empirique pour l'échantillonnage stratifié dans trois situations de complexité croissante :

– L'estimateur standard donné dans (3), pour lequel

$$l_{hi} = n_h \omega_{hi} y_{hi} - \omega_h^{\mathrm{T}} y_h; \tag{12}$$

– L'estimateur calé (5), pour lequel

$$\begin{aligned} l_{hi} &= (n_h \omega_{hi} y_{hi} - \omega_h^{\mathrm{T}} y_h) + (X_{Ch}^{\mathrm{T}} \omega_h - n_h \omega_{hi} x_{Chi})^{\mathrm{T}} \hat{\gamma} \\ &\quad + n_h \omega_{hi} (c - X_C^{\mathrm{T}} \omega)^{\mathrm{T}} (X_C^{\mathrm{T}} \Omega X_C)^{-1} x_{Chi} (y_{hi} - x_{Chi}^{\mathrm{T}} \hat{\gamma}), \end{aligned} \tag{13}$$

où ω_h et y_h sont des vecteurs de poids et de réponses $n_h \times 1$ pour la h-ème strate, X_{Ch} est la matrice de calage des variables auxiliaires $n_h \times q_C$ pour la h-ème strate, et $\hat{\gamma} = (X_C^{\mathrm{T}} \Omega X_C)^{-1} X_C^{\mathrm{T}} \Omega y$; et

– l'estimateur calé avec imputation de valeurs manquantes (9). Soit

$$\hat{\gamma}_M = (X_C^{\mathrm{T}} \Omega X_C)^{-1} X_C^{\mathrm{T}} \Omega (I - Z) \hat{y}$$

correspondant à $\hat{\gamma}$ pour les individus avec réponse manquante, et soit $l_i(\hat{\beta})$ les éléments du vecteur $q \times 1$ de fonctions d'influence pour les coefficients de régression d'imputation, correspondant à la différentiation par rapport à la i-ème unité dans la strate h. Alors pour $i = 1, \ldots, n_h$ and $h = 1, \ldots, H$,

$$\begin{aligned} l_{hi} &= (n_h \omega_{hi} z_{hi} y_{hi} - \omega_h^{\mathrm{T}} Z_h y_h) + (X_{Ch}^{\mathrm{T}} \omega_h - n_h \omega_{hi} x_{Chi})^{\mathrm{T}} \hat{\gamma} \\ &\quad + n_h \omega_{hi} (c - X_C^{\mathrm{T}} \omega)^{\mathrm{T}} (X_C^{\mathrm{T}} \Omega X_C)^{-1} x_{Chi} (z_{hi} y_{hi} - x_{Chi}^{\mathrm{T}} \hat{\gamma}) \\ &\quad + \left\{ n_h \omega_{hi} (1 - z_{hi}) \hat{y}_{hi} - \omega_h^{\mathrm{T}} (I_h - Z_h) \hat{y}_h \right\} + \omega^{\mathrm{T}} (I - Z) X l_i(\hat{\beta}) \\ &\quad + (X_{Ch}^{\mathrm{T}} \omega_h - n_h \omega_{hi} x_{Chi})^{\mathrm{T}} \hat{\gamma}_M \\ &\quad + n_h \omega_{hi} (c - X_C^{\mathrm{T}} \omega)^{\mathrm{T}} (X_C^{\mathrm{T}} \Omega X_C)^{-1} x_{Chi} \left\{ (1 - z_{hi}) \hat{y}_{hi} - x_{Chi}^{\mathrm{T}} \hat{\gamma}_M \right\} \\ &\quad + (c - X_C^{\mathrm{T}} \omega)^{\mathrm{T}} (X_C^{\mathrm{T}} \Omega X_C)^{-1} X_{Ch}^{\mathrm{T}} \Omega_h (I_h - Z_h) X_h l_i(\hat{\beta}). \end{aligned} \tag{14}$$

Le modèle d'imputation déterministe (7) pour les moindres carrés donne

$$l_i(\hat{\beta}) = n_h z_i (X^{\mathrm{T}} Z X)^{-1} x_i \left\{ y_i - x_i^{\mathrm{T}} (X^{\mathrm{T}} Z X)^{-1} X^{\mathrm{T}} Z y \right\}, \; i = 1, \ldots, \sum_{h=1}^{H} n_h,$$

où X est la matrice de régression. Pour le modèle d'imputation (8) pour lequel les coefficients de régression varient par strate, les $l_i(\hat{\beta})$ dans (14) sont

$$l_i(\hat{\beta}_h) = n_h z_i (X_h^{\mathrm{T}} Z_h X_h)^{-1} x_i \left\{ y_i - x_i^{\mathrm{T}} (X_h^{\mathrm{T}} Z_h X_h)^{-1} X_h^{\mathrm{T}} Z_h y_h \right\},$$

où X_h, Z_h, et y_h sont la matrice de variables auxiliaires, la matrice indicatrice pour les réponses observées, et le vecteur de réponses pour la strate h.

On notera que (12), (13) et (14) correspondent dans l'ordre à des résidus simples, calés, puis avec une composante pour la réponse et la non-réponse des rédidus calés. Ces formules sont utilisées par la suite dans nos comparaisons numériques.

4. Jackknife

4.1. Idées de base

Le jackknife, introduit comme méthode pour l'estimation du biais par Quenouille (1949a,b) et ensuite proposé pour l'estimation de variance par Tukey (1958), implique la suppression systématique de groupes d'unités, le recalcul de la statistique avec chaque groupe supprimé tour à tour, et la combinaison de toutes ces statistiques recalculées.

Le jackknife le plus simple consiste à écarter chaque observation d'un échantillon tour à tour, recalculer la statistique pour obtenir des valeurs $\hat{\theta}_{-1}, \ldots, \hat{\theta}_{-n}$ en plus de sa valeur originale, $\hat{\theta}$, et ensuite à estimer le biais et la variance de l'estimateur par

$$\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{-i}-\hat{\theta}) = (n-1)(\bar{\hat{\theta}}-\hat{\theta}), \quad \frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{-i}-\bar{\hat{\theta}})^2.$$

Pour que ces estimateurs soient convergents, la statistique $\hat{\theta}$ doit être suffisamment lisse comme fonction des observations, par exemple posséder un développement de Taylor. Certains types d'estimateurs non-lisses, comme ceux basés sur les quantiles, ne possédent pas un développement convenable, de sorte que les estimateurs de leur variance de type jackknife simple ne sont pas convergents (Efron, 1982). Bien que ceci est un désavantage majeur du point de vue théorique, une considération pratique montre qu'en dépit de son inconsistence, l'estimateur de variance jackknife, pour des quantiles d'échantillons basés sur des échantillons d'enquêtes, peut être compétitif par rapport à d'autres estimateurs en terme de moindres carrés moyens (Rao *et al.*, 1982). Shao et Wu (1989) montrent que l'inconsistence peut être réparée en supprimant des groupes de d observations, où $d = o(n) \to \infty$ lorsque $n \to \infty$.

Rao et Wu (1985) ont montré que la linéarisation et le jackknife sont asymptotiquement équivalents au premier ordre, et que dans certains cas ils produisent exactement les mêmes estimations de variance. Quelques simulations tendent aussi à montrer que le jackknife est moins biaisé mais plus variable que l'estimateur de variance de type linéarisation. Berger et Skinner (2005) développent une théorie pour l'estimation de variance jackknife pour une grande classe d'estimateurs rééchantillonnage à probabilités inégales, et Berger et Rao (2006) développent ceci pour l'imputation.

On peut considérer le jackknife comme un moyen numérique pour obtenir les dérivées de $\hat{\theta}$ par rapport aux observations. Ecarter une observation entraîne une perturbation de l'ordre de $O(n^{-1})$ de l'échantillon, et comme la variance de l'échantillonnage est typiquement de $O(n^{-1/2})$, on peut soupçonner que les perturbations dues au jackknife sont trop locales pour bien estimer la variance de l'estimateur. Des travaux théoriques dans le cas des données indépendentes suggèrent que les estimateurs de variance jackknife tendent à sous-estimer légèrement la vraie variance, et que cette sous-estimation peut être attribuée à la localité du développement de Taylor.

Miller (1974) propose une excellente revue des premiers travaux sur le jackknife, et une étude récente plus détaillée est donnée par Shao et Tu (1995). Des estimateurs plus généraux de sous-échantillonnage, utiles surtout pour les séries temporelles et d'autres formes de données corrélées, ont été proposés par Politis *et al.* (1999).

4.2 Jackknife modifié

Des complications apparaissent dans le contexte des enquêtes car les plans d'échantillonnage sans remise mènent à des observations corrélées et les données sont stratifiées. Avec des données d'enquêtes stratifiées, le jackknife simple entraîne la suppression tour à tour de chaque unité primaire d'échantillonnage. La suppression de la i-ème unité dans la h-ème strate change le poid pour la j-ème unité dans la k-ème strate en

$$\omega_{kj}^{(hi)} = \begin{cases} \omega_{kj}, & k \neq h, \\ \omega_{hj} n_h/(n_h - 1), & k = h, j \neq i, \\ 0, & k = h, j = i. \end{cases}$$

Très souvent le rapport $n_h/(n_h - 1) \doteq 1$, et alors ces poids actualisés donnent les statistiques $\hat{\theta}_{h,-i}$, c'est-à-dire $\hat{\theta}$ calculé sans la i-ème unité de la strate h, pour $i = 1, \ldots, n_h$ et $h = 1 \ldots, H$. L'estimation de variance par jackknife simple est alors donnée par

$$v_{\mathrm{J}} = \sum_{h=1}^{H} \frac{(1 - f_h)(n_h - 1)}{n_h} \sum_{i=1}^{n_h} (\hat{\theta}_{h,-i} - \bar{\hat{\theta}}_h)^2, \qquad \bar{\hat{\theta}}_h = n_h^{-1} \sum_{i=1}^{n_h} \hat{\theta}_{h,-i}. \tag{15}$$

Rao et Wu (1985) montrent que remplacer $\bar{\hat{\theta}}_h$ dans (15) par $n^{-1} \sum_{h=1}^{H} \sum_{i=1}^{n_h} \hat{\theta}_{h,-i}$ ou par $H^{-1} \sum_{h=1}^{H} \bar{\hat{\theta}}_h$ a peu d'effet sur la performance de v_{J} dans le sens que les trois versions sont asymptotiquement équivalentes au deuxième ordre. Kovar *et al.* (1988) observent que les trois estimateurs de variance résultants sont également comparables pour des échantillons finis.

Une autre complication du jackknife est qu'il nécessite un gros effort de calcul parce qu'un total de n suppressions est nécessaire ; le fardeau de calcul résultant est souvent trop grand pour que ce simple jackknife se révèle utile. La suppression de d éléments d'échantillon à la fois (Shao et Wu, 1989) nécessiterait $\binom{n}{d}$ recalculs de la statistique, ce qui est encore moins faisable, mais il est courant de réduire le nombre de réplications nécessaires en séparant la strate h en m_h blocs disjoints de d observations, et d'ensuite recalculer la statistique avec chacun de ces blocs supprimés. Le nombre total de recalculs est ainsi $m = \sum_{h=1}^{H} m_h$ plutôt que n ou $\binom{n}{d}$. La suppression des unités de la partie de la h-ème strate dont les indices appartiennent à l'ensemble $\mathcal{S}$ change le poids pour la j-ème unité dans la k-ème strate en

$$\omega_{kj}^{(h\mathcal{S})} = \begin{cases} \omega_{kj}, & k \neq h, \\ \omega_{hj} n_h/(n_h - |\mathcal{S}|), & k = h, j \notin \mathcal{S}, \\ 0, & k = h, j \in \mathcal{S}, \end{cases}$$

où $|\mathcal{S}|$ représente le nombre d'éléments d dans $\mathcal{S}$. Des variantes de ces procédures jackknife groupées choisissent les membres du bloc aléatoirement lors de chaque recalcul, avec ou sans remise (Krewski et Rao, 1981 ; Shao et Tu, 1995, §5.2.2). Etant donné que l'échantillonnage avec remise est simple et produit des résultats comparables à l'échantillonnage sans remise, l'échantillonnage avec remise jackknife à d-suppressions semble préférable. Pour chacun de ceux-ci l'estimation de variance est

$$v_{J,-d} = \sum_{h=1}^{H} \frac{(1-f_h)(n_h-d)}{n_h} \sum_{i=1}^{m_h} (\hat{\theta}^d_{h,-i} - \bar{\hat{\theta}}^d_h)^2,$$

où $\hat{\theta}^d_{h,-i}$ est l'estimateur calculé à partir des données avec le i-ème bloc de taille d exclu de la h-ème strate, et $\bar{\hat{\theta}}^d_h$ est la moyenne pour cette strate. De légères variations des formules ci-dessus sont nécessaires lorsque n_h n'est pas exactement divisible par d. Enfin Kim et Sitter (2003) décrivent une approche pour réduire les calculs dans le cas d'un échantillonnage à deux phases, et donnent d'autres références dans ce domaine.

Le jackknife à d-suppressions a l'avantage théorique supplémentaire d'offrir la convergence à des statistiques non lisses telles que les quantiles. Ainsi si $d, m_h \to \infty$ en même temps que $n_h \to \infty$, alors le jackknife de bloc produit des estimateurs de variance convergents (Shao et Wu, 1989). Et moins la statistique d'intérêt est lisse, plus d doit être grand. Cependant choisir d et m_h n'est pas un problème résolu lorsque n_h est fini.

4.3. Imputation

Pour prendre en compte la non-réponse, Rao et Shao (1992) décrivent une version convergente de l'estimateur jackknife simple en utilisant un mécanisme d'imputation de type 'hot deck' particulier. Chen et Shao (2001) montrent que cette approche échoue pour l'imputation de plus proche voisin ; voir aussi Fay (1996). En procédant à une ré-imputation chaque fois qu'une unité ou qu'un bloc de d unités est supprimé, le jackknife peut être utilisé simplement mais avec un coût de calcul encore plus élevé.

5. Répliques équilibrées répétées

5.1. Idées de base

Le demi-échantillonnage équilibré (McCarthy, 1969) est la forme la plus simple de réplication équilibrée répétée ('balanced repeated replication' en anglais). Il fut au départ développé pour les plans stratifiés à plusieurs degrés avec deux unités d'échantillonnages primaires effectuées avec remise lors du premier degré. Lorsqu'un échantillon consiste en deux observations dans chacune des H strates, un demi-échantillon est formé en prenant une observation pour chaque strate. Le poids ω_{hi} attaché à la i-ème unité dans la strate h est changé en $2\omega_{hi}$ si cette unité est incluse dans le demi-échantillon, et en 0

sinon, et la statistique est recalculée avec le demi-échantillon et ces nouveaux poids.

Le calcul des 2^H telles statistiques répliquées est infaisable à moins que H ne soit petit, mais des idées de plan d'expérience permettent d'obtenir avec moins de réplications la même estimation de variance pour des statistiques linéaires, à condition que les réplications soient équilibrées (Wolter, 1985, appendice A). Soit $\alpha_{hr} = \pm 1$ selon que la première ou la deuxième des deux unités dans la h-ème strate est retenue dans la r-ème réplication, et soit $\hat{\theta}_r$ la valeur de l'estimateur $\hat{\theta}$ calculé en utilisant le r-ème demi-échantillon et ses poids ajustés. Alors l'ensemble des R demi-échantillons est équilibré si

$$\sum_{r=1}^{R} \alpha_{hr}\alpha_{kr} = 0, \quad \text{pour tous} \quad h \neq k, k, h = 1, \ldots, H,$$

où on prend $R \in \{H, H+1, H+2, H+3\}$, et dans ce cas l'estimateur de variance de réplication équilibrée répétée

$$v_{\mathrm{RER}} = R^{-1} \sum_{r=1}^{R} (\hat{\theta}_r - \hat{\theta})^2 \tag{16}$$

produit la même estimation de variance que par calcul utilisant l'ensemble entier de 2^H demi-échantillons, pour des statistiques linéaires. Pour beaucoup de statistiques, v_{RER} est égal à la fois à v_{J} et à l'estimateur de variance de linéarisation. On pourrait aussi remplacer $\hat{\theta}$ en (16) par $R^{-1}\sum_{r=1}^{R}\hat{\theta}_r$ sans changer ses propriétés asymptotiques.

Le calage est appliqué à un demi-échantillon en calant les poids ré-échellonnés en utilisant les marges de la population. Lorsqu'il n'y a pas de valeurs manquantes, v_{RER} est convergent pour les variances de fonctions non-linéaires (Krewski et Rao, 1981). Il s'avère que v_{RER} est stochastiquement proche de l'estimateur de variance de linéarisation, mais pas aussi proche que ne l'est v_{J}. La convergence de v_{RER} pour des estimateurs de quantiles a été établie par Shao et Wu (1992) et Shao et Rao (1994).

5.2. Strate de taille $n_h > 2$

Deux généralisations principales ont été proposées quand il y a plus de $n_h = 2$ observations par strate. La première (Gurney et Jewett, 1975 ; Gupta et Nigam, 1987 ; Wu, 1991 ; Sitter, 1993) utilise des tableaux orthogonaux, ce qui nécessite un grand nombre de répliques, la rendant peu pratique pour beaucoup d'applications. Une deuxième généralisation plus pragmatique est de regrouper les unités d'échantillonnages primaires dans chaque strate en deux groupes de tailles $m_h = \lfloor n_h/2 \rfloor$ et $n_h - m_h$, et d'appliquer la réplication équilibrée répétée utilisant les groupes ('grouped balanced repeated replication') plutôt que des unités individuelles (Rao et Shao, 1996 ; Wolter, 1985, §3.7). Les poids pour la r-ème réplique sont ajustés en utilisant la

formule

$$\omega_{hi}^r = \begin{cases} \omega_{hi}\left[1+\left\{\dfrac{(n_h-m_h)(1-f_h)}{m_h}\right\}^{1/2}\right], & \alpha_{hr}=1, \\ \omega_{hi}\left[1-\left\{\dfrac{m_h(1-f_h)}{n_h-m_h}\right\}^{1/2}\right], & \alpha_{hr}=-1, \end{cases} \tag{17}$$

où le facteur $(1-f_h)$ est un ajustement pour l'effet de l'échantillonnage sans remise et la condition $m_h \leqslant n_h/2$ assure que les poids ajustés sont positifs. Le choix $m_h = n_h/2$ conduit à l'ajustement du poids original à $2\omega_{hi}$ ou 0 lorsque f_h peut être négligé. L'estimateur de variance reste défini par (16), mais l'effet du regroupement n'induit plus l'équivalence avec l'estimateur de variance standard dans le cas linéaire.

5.3. Estimateurs améliorés

L'estimateur de variance v_{RER} peut s'avérer très variable, pour deux raisons principales. La première est que la loi de (16) est approximativement proportionelle à la loi χ^2_{R-1}. Ainsi le coefficient de variation de (16) est à peu près $(2/R)^{1/2}$; ceci est appréciable lorsque H est petit. Le coefficient de variation peut être réduit en augmentant les degrés de libertés, par exemple en séparant les H strates en sous-strates artificielles, auxquelles on applique alors le demi-échantillonnage.

La deuxième difficulté est que certaines statistiques peuvent être fortement sensibles aux perturbations des poids, et il peut même s'avérer impossible de calculer tous les $\hat{\theta}_r$, par exemple, lorsque le dénominateur d'un rapport s'annule pour certaines repondérations, et lorsque la suppression de la moitié de l'échantillon résulte en des tailles d'échantillons trop petites.

Une solution suggérée par Robert Fay du US Census Bureau (Dippo *et al.*, 1984 ; Fay, 1989) est d'utiliser une méthode de repondération plus douce : un facteur ε est choisi, avec $0 < \varepsilon \leqslant 1$, et les poids sont perturbés en

$$\omega_{hi}^r = \begin{cases} \omega_{hi}\left[1+\varepsilon\left\{\dfrac{(n_h-m_h)(1-f_h)}{m_h}\right\}^{1/2}\right], & \alpha_{hr}=1, \\ \omega_{hi}\left[1-\varepsilon\left\{\dfrac{m_h(1-f_h)}{n_h-m_h}\right\}^{1/2}\right], & \alpha_{hr}=-1, \end{cases}$$

qui produit des estimations répliquées $\hat{\theta}_1(\varepsilon), \ldots, \hat{\theta}_R(\varepsilon)$ et un estimateur de variance

$$v_{\mathrm{RER}}(\varepsilon) = \frac{1}{R\varepsilon^2}\sum_{r=1}^{R}\left\{\hat{\theta}_r(\varepsilon)-\hat{\theta}\right\}^2.$$

En posant $\varepsilon = 1$ on retrouve les poids (17) et l'estimateur de variance (16), mais d'autres valeurs de ε reviennent à sous-pondérer plutôt que supprimer des groupes d'unités, et réduisent ainsi fortement l'instabilité des estimations

répliquées. Rao et Shao (1999) ont étudié les aspects théoriques de ces modifications de poids. Des travaux numériques (Judkins, 1990 ; Rao et Shao, 1999) suggèrent le choix de $\varepsilon = 0.5$.

5.4. Réplication équilibére répétée groupée de manière répétitive

Lorsque H est grand, la réplication équilibrée répétée groupée engendre un estimateur de variance convergent, mais ceci peut être très variable si $H \ll n$. De plus elle est non convergente lorsque H est fixé et que $n_h \to \infty$ (Rao et Shao, 1996), suggérant que cela peut mal fonctionner pour H petit. Ceci a été observé empiriquement, et comme mentionné ci-dessus, est due en partie au fait que lorsque H est fixé l'estimateur de variance v_{RER} a une loi proche de $c\chi^2_{R-1}$.

Pour réduire la variabilité de v_{RER} et pour enlever l'effet de la séparation aléatoire particulière des unités de la strate en groupes, Rao et Shao (1996) proposent de répéter la méthode sur des groupes sélectionnés de manière aléatoire pour fournir m estimations de variance, dont la moyenne produit une estimation de variance globalement plus stable. Cet estimateur de variance basé sur la réplication équilibrée répétée groupée de manière répétitive ('repeatedly grouped balanced repeated replication') est convergent lorsque $mR \to \infty$. En particulier, il est suggéré de choisir m tel que $mR \doteq n$, ce qui implique que m devrait être grand lorsque le nombre de strates H est petit. Lorsque H est grand, la règle suggère de choisir $m = 1$, ce qui necessite $O(n)$ répliques.

5.5. Imputation

Rao et Shao (1999) ont considéré l'extension de $v_{\mathrm{RER}}(\varepsilon)$ à l'imputation pour la non-réponse partielle. Shao *et al.* (1998) ajustent la réplication équilibrée répétée à la présence de non-réponses, en tenant compte d'un mécanisme d'imputation déterministe ou stochastique. Dans un plan d'échantillonnage stratifié à plusieurs degrés, ils établissent la convergence de l'estimateur de variance ajusté basé sur la réplication équilibrée répétée.

6. Bootstrap

6.1. Idées de base

Le bootstrap a été intensément étudié pour des données indépendantes et identiquement distribuées, où l'échantillonnage est fait avec remise, mais son adaptation aux enquêtes complexes n'est pas directe, parce qu'en général la méthode d'échantillonnage induit une dépendance. On doit également tenir compte du calage et de l'imputation lorsqu'on effectue le bootstrap sur un échantillon d'enquête.

L'idée du bootstrap est d'imiter la manière dont ont été engendrées les données. Comme les méthodes de réplication équilibrée répétée et du jackknife, le bootstrap re-caclule la statistique avec des échantillons tirés aléatoirement d'une pseudo-population, donnant les statistiques correspondantes $\hat{\theta}^*$. En

répétant ce processus R fois indépendemment, l'estimation de variance par bootstrap est donnée par

$$v_B = (R-1)^{-1} \sum_{r=1}^{R} (\hat{\theta}_r^* - \bar{\hat{\theta}}^*)^2, \quad \bar{\hat{\theta}}^* = R^{-1} \sum_{r=1}^{R} \hat{\theta}_r^*.$$

Le bootstrap standard utilise l'échantillonnage avec remise, mais ceci ne concorde pas avec l'échantillonnage sans remise généralement utilisé dans le contexte d'enquête. Ne pas intégrer une correction d'échantillon fini résulterait en un estimateur de variance biaisé. Pour des données stratifiées, le rééchantillonnage s'effecte indépendemment dans chaque strate.

L'exemple simple de la moyenne d'une population de taille N par échantillonnage de probabilités égales sans remise illustre ce problème de biais. En effet, pour la moyenne d'échantillon $\bar{y} = n^{-1} \sum_{i \in \mathcal{S}} y_i$, où $\mathcal{S} \subset \{1, \ldots, N\}$ contient les n indices échantillonnés, un estimateur de la variance est donné par

$$(1-f)\frac{s^2}{n}, \quad s^2 = \frac{1}{n-1} \sum_{i \in \mathcal{S}} (y_i - \bar{y})^2,$$

où $f = n/N$ est la fraction d'échantillonnage. La méthode de bootstrap le plus simple prend des échantillons de taille n avec remise à partir de $\{y_i : i \in \mathcal{S}\}$, et utilise les propriétés d'échantillonnage de la moyenne bootstrap $\bar{y}^*$ pour estimer ceux de $\bar{y}$ sous répétition du plan d'échantillonnage original. La variance bootstrap de $\bar{y}^*$ est $\mathrm{var}^*(\bar{y}^*) = (n-1)/n \times s^2/n$, ce qui diffère de $\mathrm{var}(\bar{y})$ à moins que n soit grand et que la fraction d'échantillonnage soit petite, ou que $f = 1/n$. Le même problème survient avec l'échantillonnage stratifié, où la variance d'un estimateur global est une somme de contribution des strates individuelles ; voir (15).

Ainsi l'estimateur de variance de bootstrap standard peut s'avérer non convergent dans des plans d'échantillonnage quand $H \to \infty$ et quand les fractions d'échantillonage f_h sont à une distance bornée de zéro. L'échec asymptotique peut paraître sans importance si l'estimateur de variance bootstrap possède une erreur quadratique moyenne faible pour le type d'enquête rencontré en pratique. Cet échec du bootstrap conventionnel a engendré des travaux pour le modifier. Efron (1982) a reconnu le problème d'échellonnage des petites strates, et a suggéré la solution simple convergent à prendre des échantillons bootstrap de taille $n_h - 1$. Mais ceci n'arrange pas l'effet de la fraction d'échantillonnage non négligeable f_h, pour lesquelles des approches plus élaborées sont requises, ce que nous décrivons maintenant.

6.2. Approches bootstrap

Une approche simple dans le cas d'un échantillon non stratifié est de créer une pseudo population utilisant N/n répliques de l'échantillon, et d'ensuite échantillonner à partir de cette population artificielle sans remise (Gross, 1980 ; Chao et Lo, 1985). Pour des échantillons stratifiés (Bickel et Freedman, 1984 ; Booth *et al.*, 1994), la h-ème strate est répliquée N_h/n_h fois, et

l'échantillonnage stratifié est appliqué séparément dans chaque strate. Bickel et Freedman (1984) généralisent cette méthode à la situation où N_h/n_h n'est pas entier, par randomisation entre deux tailles de populations différentes, et dérivent quelques propriétés théoriques. McCarthy et Snowden (1985) décrivent quelques situations où cette approche n'est pas faisable. Sitter (1992b) modifie la méthode bootstrap pour contourner les problèmes de l'estimateur de variance de Bickel et Freedman (1984) en sélectionnant des échantillons sans remise de taille m'_h dans chaque pseudo-population de taille $k_h n_h$, où

$$m'_h = n_h - (1 - f_h) \quad \text{et} \quad k_h = \left(1 - \frac{1 - f_h}{n_h}\right) / f_h, \tag{18}$$

pour obtenir un estimateur de variance non biaisé et convergent, au moins pour une statistique linéaire. Lorsque la fraction d'échantillonnage inverse N_h/n_h n'est pas entière, toutes ces méthodes nécessitent une randomisation entre les pseudo-populations dont les tailles approchent N_h. Les détails de la randomisation sont importants, et en plus d'être quelque peu encombrant à programmer, des erreurs peuvent conduire à l'échec de ces méthodes (Presnell et Booth, 1994). De plus le besoin d'une pseudo-population pour chaque strate requiert de la mémoire additionnelle.

Sans créer une pseudo-population, McCarthy et Snowden (1985) proposent d'imiter le plan d'échantillonnage original en sélectionnant des échantillons de taille m'_h avec remise dans chaque strate avec

$$m'_h = (n_h - 1)/(1 - f_h) \tag{19}$$

de manière à obtenir un estimateur de variance non biaisé et convergent pour un estimateur linéaire. Lorsque m'_h n'est pas entier, une randomisation est à nouveau requise.

Rao et Wu (1988) proposent aussi un bootstrap avec remise, mais ils atteignent la variance standard en rééchellonnant les échantillons bootstrap de taille m'_h selon

$$y^*_h = \bar{y}_h + \sqrt{m'_h(1 - f_h)/(n_h - 1)}(\bar{y}^*_h - \bar{y}_h),$$

où m'_h peut être n'importe quel entier positif. Ceci inclut la méthode de Efron (1982) comme cas particulier. Ils montrent que les estimateurs de distribution sont convergents. Pour estimer les moments de troisième ordre, Rao et Wu (1988) suggèrent de prendre

$$m'_h = \lfloor (1 - f_h)(n_h - 2)^2/(1 - 2f_h)/(n_h - 1) \rfloor,$$

ce qui évite le problème de m'_h non entier comme dans (18) et (19). La méthode a aussi l'avantage de ne pas créer et stocker les pseudo-populations. Cependant pour $m'_h \geqslant n_h$, elle peut conduire à des valeurs impossibles de la statistique bootstrapée.

Le bootstrap mirroir-concordance ('mirror-match bootstrap') (Sitter, 1992a) est un croisement entre le bootstrap avec et sans remise. Dans chaque strate, il produit un échantillon de taille $n^*_h < n_h$ sans remise $k_h = n_h(1 - f^*_h)/\{n^*_h(1 -$

$f_h)\}$ fois. Pour concorder avec les moments de troisième ordre, Sitter (1992a) suggère $n_h^* = f_h n_h$. Ces estimateurs de variance sont convergents pour des statistiques linéaires, et sont équivalents au bootstrap avec remise de McCarthy et Snowden (1985) si $n_h^* = 1$. Une randomisation entre les entiers avoisinants est requise lorsque k_h n'est pas un entier (Sitter, 1992a).

6.3. Non-réponse

En cas de non-réponse, le mécanisme d'imputation doit être appliqué à chaque ré-échantillon (Shao et Sitter, 1996). Ainsi ils proposent de ré-imputer de manière répétitive les répondants de l'échantillon bootstrappé pour ajuster le modèle d'imputation et ensuite imputer les non-répondants de l'échantillon bootstrap. C'est pourquoi la méthode est lourde en terme de calcul. Mais elle a l'avantage de fournir des estimateurs de variance convergents pour la médiane et d'autres estimateurs basés sur les quantiles.

6.4. Commentaires

Davison et Hinkley (1997, §3.7) contient des commentaires additionnels et des exemples. Sitter (1992b) utilise la simulation pour comparer diverses méthodes de bootstrap avec et sans remise, et trouve qu'ils sont comparables. La complexité des méthodes mirroir-concordance et de ré-échellonnage nous amène à préférer le bootstrap de rééchantillonnage simple dans les cas où la fraction d'échantillonnage est petite, et d'utiliser l'approche de pseudo-population le cas échéant. Dans le contexte d'enquêtes complexes, Lahiri (2003) et Shao (2003) passent en revue l'application des méthodes bootstrap, et Rust et Rao (1996) des méthodes de réplication. Presnell et Booth (1994) mettent l'accent sur des difficultés théoriques de plusieurs approches.

7. Comparaisons empiriques

7.1. Cas simples

De nombreuse simulations ont été effectuées pour étudier la performance relative des estimateurs de variance dont nous avons discuté.

Kovar *et al.* (1988) ont considéré l'échantillonnage aléatoire simple à un degré stratifié avec remise basé sur des pseudo-données qui ressemblent aux vraies populations avec $H = 32$ strates à partir desquelles des échantillons de tailles $n_h \in \{2, 5\}$ sont prélevés. Les paramètres considérés sont le rapport, le coefficient de régression, le coefficient de corrélation, et la médiane. Sur la base d'une mesure de biais et de stabilité, les estimateurs de variance comparés sont le jackknife, la réplication équilibrée répétée, le bootstrap avec remise par ré-échellonnage, et les méthodes de linéarisation. Les conclusions de leur étude sont les suivantes. La linéarisation et le jackknife se comportent de manière équivalente et ont la meilleure performance en termes de biais relatif et stabilité. Ensuite vient la réplication équilibrée répétée. Le bootstrap par ré-échellonnage de Rao et Wu (1988), la seule méthode bootstrap qu'ils considèrent,

produit la pire performance de toutes. La réplication équilibrée répétée et ce bootstrap tendent à surestimer les variances de statistiques non linéaires. Ils considèrent aussi la couverture d'intervalles de confiance, et concluent que les intervalles jackknife et de linéarisation tendent à sous-couvrir, tandis que ceux basés sur la réplication équilibrée répétée tendent à surcouvrir. Les intervalles de confiance bootstrap plus raffinés (par exemple par studentisation) améliorent la situation, mais tendent également à surcouvrir.

Rao et Shao (1996) ont comparé les méthodes de réplication équilibrée répétée de groupe et de réplication équilibrée répétée de manière répétitive de groupe lorsque $n_h = 48$ dans chacunes des $H = 5$ strates ; la fraction d'échantillonnage est d'environ 1/10. L'échantillonnage aléatoire simple au premier degré avec remise est utilisé à l'intérieur de chaque strate. Les conclusions sont que le jackknife produit de bons résultats pour un estimateur ponctuel lisse, et que, en accord avec la théorie, il n'est pas convergent pour la médiane. La méthode de réplication de groupe équilibrée répétée a un faible biais relatif, mais est très instable. Cette instabilité est corrigée par la réplication équilibrée répétée groupée de manière répétitive, mais au prix de plus de calculs.

Sitter (1992b) a comparé diverses méthodes de bootstrap, à savoir le bootstrap avec remise, le bootstrap de ré-échellonnage, une version du bootstrap mirroir-concordance et une version du bootstrap sans remise, sous échantillonnage aléatoire simple stratifié à un degré sans remise. Il a conclu que les méthodes bootstrap sont comparables entre elles pour l'estimation de la variance de la médiane.

7.2. Calage

7.2.1. Etude néérlandaise

Boonstra et Nieuwenbroek (2003) ont effectué une comparaison des variances de la linéarisation de Taylor et de la réplication équilibrée répétée pour des estimateurs de régression généralisée de totaux et de rapport de totaux à partir de deux enquêtes différentes : un plan de sondage deux-par-strate d'une population d'entreprises, et l'autre un plan à deux degrés d'une population de personnes.

La première simulation englobe 2000 échantillons pris dans 84 strates de taille N_h entre 10–20 : deux observations ont été prises dans chaque strate pour un total de 168 observations. Le calage est basé sur les valeurs de deux variables catégoriques en utilisant un estimateur de régression du type décrit dans §2.2, et les estimateurs de variance sont calculés en utilisant la linéarisation de Taylor, la linéarisation de Taylor avec des poids de régression, et la réplication équilibrée répétée avec divers facteurs de Fay. Les estimateurs de Taylor ne tiennent pas compte de l'aléatoire des poids de régression. Les conclusions générales sont que toutes les méthodes sous-estiment légèrement les variances, et que la méthode de Taylor standard tend à sous-estimer le plus. La seule exception à ceci est la réplication équilibrée répétée standard, qui surestime quelque peu les vraies variances. Toutes les méthodes fournissent

des intervalles de confiance à 95 % avec une couverture de 90–93 %; ceci est une conséquence de la sous-estimation de la variance.

Dans une deuxième étude, Boonstra et Nieuwembroek (2003) ont créé une population artificielle en utilisant des données à partir de l'enquête néérlandaise à deux degrés sur la population active et l'ont utilisée pour comparer diverses formes de réplication équilibrée répétée. La population artificielle comprend 188,216 personnes dans 70 strates contenant chacune de 10 à plusieurs centaines d'unités d'échantillonnage primaires. Chacune de ces unités contient de 15 à 40 ménages. Cinq cents échantillons sont tirés selon un plan de sondage à deux degrés avec échantillonnage aléatoire simple avec remise à chaque degré, avec des fractions d'échantillonnage de 1/5 à chaque degré produisant des échantillons généraux de 4 % de la population. Les estimateurs de variance utilisés sont les estimateurs de Taylor correspondant à ceux mentionnés dans le paragraphe précédant, et la réplication équilibrée répétée groupée avec un facteur de Fay de $\varepsilon = 0.57$ et des strates artificielles pour un total de 120 strates et de 120 rééchantillons.

Le Tableau 1 montre que la réplication équilibrée répétée est de manière appréciable moins efficace que la linéarisation de Taylor, qui est peu affectée par l'utilisation des poids calés. L'utilisation de strates artificielles supplémentaires améliore légèrement la réplication équilibrée répétée.

Tableau 1. — Résultat représentatif de l'étude néérlandaise de Boonstra et Nieuwenbroek (2003). Performances des estimateurs de variance par linéarisation de Taylor standard et par réplication équilibrée répétée (RER) groupée. BR désigne le biais relatif et REQM la racine de l'erreur quadratique moyenne des 500 estimations.

Statistique	Méthodes d'estimation de variance	BR (%)	REQM relatif (%)
Total	Taylor standard	2.2	15.4
	RER groupée	–0.1	29.2
	RER groupée, avec strates artificielles	2.1	22.5
Rapport	Taylor standard	0.1	4.8
	Taylor standard, avec poids de régression	–0.4	4.8
	RER groupée	–0.9	25.0
	RER groupée, avec strates artificielles	0.6	13.7

7.2.2. Études britanniques et suisses

Deux exercices de simulation majeurs entrepris pour comparer les performances du rééchantillonnage et des méthodes plus standard ont utilisé des données simulées basées sur des enquêtes suisses et britanniques sur la population active (Canty et Davison, 1999). Celles-ci comportent des vagues de participants se chevauchant, et les statistiques considérées sont les totaux, les rapports, et différences de totaux et rapports entre deux vagues successives. Le calage par ajustement itératif proportionnel ajuste les poids aux

marges des trois variables catégoriques. L'effet de données manquantes n'est pas considéré.

La population de l'étude britannique consiste en environ 60, 000 adresses dans la première vague, avec autour de 12, 000 nouvelles adresses dans la deuxième vague utilisées pour remplacer 12, 000 adresses de la première vague ; soit environ 20% d'unités changées entre deux vagues. Trente strates sont utilisées, chacune contenant quelques individus présents seulement dans la première vague, quelques uns présents seulement dans la deuxième vague, et certains présents dans les deux vagues, dans les proportions 1 : 1 : 3. La fraction d'échantillonnage est d'environ $f = 1/48$, donnant une taille d'échantillon de 1250 dans chaque vague échantillonnée. Un total de 5000 échantillons de la population artificielle suivent ce schéma d'échantillonnage, et différentes estimations de variance sont calculées pour une variété de statistiques lisses, basées sur chacun de ces échantillons. Pour les totaux et pour les différences de totaux entre deux vagues, l'estimateur utilisé est $\hat{\tau} = \sum_{h,j,k} w_{hjk} y_{hjk}$, où la somme est sur les strates h, les adresses j, et les personnes k vivant à cette adresse.

Les méthodes d'estimation de variance utilisées sont :

- une méthode standard basée sur une formule obtenue en traitant $\hat{\tau}$ comme une variable binomiale avec dénominateur $n = \sum_h n_h$ le nombre total d'adresses échantillonnées, c'est-à-dire (Canty et Davison, 1999, équation (1))
$$d_p^2 \frac{\hat{\tau}(1-\hat{\tau})}{n} \left(\sum_{h,j,k} w_{hjk} \right),$$
où d_p est un effet de plan ;
- la linéarisation, où (10) est utilisée avec
$$l_{hj} = n_h \sum_k w_{hjk} e_{hjk} - \sum_i \sum_k w_{hik} e_{hik},$$
avec le résidu $e_{hjk} = y_{hjk} - \hat{y}_{hjk}$ d'une régression de la variable y sur X dont les lignes contiennent les variables de calage x_{hjk} utilisant la matrice diagonale Ω des poids originaux ω_{hjk} pour l'inclusion de l'individu k de l'adresse j de la strate h, et les poids recalés sont
$$w_{hjk} = \omega_{hjk} \left\{ 1 + (c^T - 1^T \Omega X)(X^T \Omega X)^{-1} x_{hjk}^T \right\};$$
- le jackknife groupé, avec 10 groupes dans chacune des 30 strates, de sorte que chaque recalcul de la variances nécessite 300 calculs de la statistique ;
- la réplication groupée équilibrée répétée, avec les 30 strates subdivisées en groupe suivant la présence de l'unité dans la première vague, la deuxième vague, ou les deux vagues, donnant un total de 90 répliques ; et
- le bootstrap standard, approprié étant donnée la faible fraction d'échantillonnage, avec 100 répliques.

L'effet de repondération est étudié pour le jackknife, pour la réplication équilibrée répétée, et pour le bootstrap. Trois ensembles de poids sont utilisés : ceux pour l'échantillon original, qui sont 'incorrects' parce qu'ils ne tiennent pas compte des changements dus au rééchantillonnage, et ceux obtenus après une et cinq itérations de l'algorithme d'ajustement proportionnel itératif utilisé pour le calage.

Le tableau 2 présente des résultats caractéristiques de cette étude. Toutes ces méthodes de rééchantillonnage tiennent compte du calage, mais la méthode standard ne le fait pas, ce qui explique qu'elle sous-estime la vraie variance. Le bootstrap et la linéarisation ont un erreur quadratique moyenne plus petit que le jackknife et que la réplication équilibrée répétée. La Figure 1 montre un résumé graphique des résultats correspondant à l'estimation du taux de chômage, un rapport pour lequel à la fois le numérateur et le dénominateur sont aléatoires, expliquant la différence dans l'échelle entre cette figure et Tableau 2. La performance relativement faible du jackknife et de la réplication équilibrée répétée est due à leur instabilité : la sous-estimation est parfois grande, bien que le biais systématique est faible. La bonne performance de la méthode standard résulte de sa faible variabilité ; elle sous-estime la vraie variance uniformément.

Tableau 2. — Résumé de statistiques pour les écarts-types du total de chômage et du changement dans le total de chômage (Canty et Davison, 1999). BR désigne le biais relatif, ET l'écart type standard et EQM l'erreur quadratique moyenne des 500 estimations. RER désigne la réplication équilibrée répétée.

Statistique	Méthode	Biais	BR (%)	ET	EQM
Total	standard	–35.2	–6.1	25.4	1880
	bootstrap	–14.6	–2.5	55.0	3230
	jackknife	45.3	7.8	96.8	11400
	linéarisation	–13.6	–2.4	38.4	1650
	RER	28.0	4.8	79.6	7100
Changement dans le total	standard	–42.7	–7.2	21.6	2290
	bootstrap	–8.2	–1.4	55.5	3140
	jackknife	64.3	10.8	107.0	15600
	linéarisation	–4.7	–0.8	41.8	1770
	RER	55.5	9.3	85.5	10400

L'avantage principal des méthodes de rééchantillonnage est de tenir compte de la variabilité supplémentaire introduite par calage. L'inconvénient est que le calage doit être effectuée pour chaque estimation de rééchantillonnage, ce qui peut représenter un fardeau de calcul supplémentaire. Lorsque le calage nécessite des itérations, il est naturel de se demander combien d'étapes itératives sont requises. Cette étude montre que le bootstrap et la réplication équilibrée répétée sont stables au nombre d'itérations de l'alogrithme d'ajustement

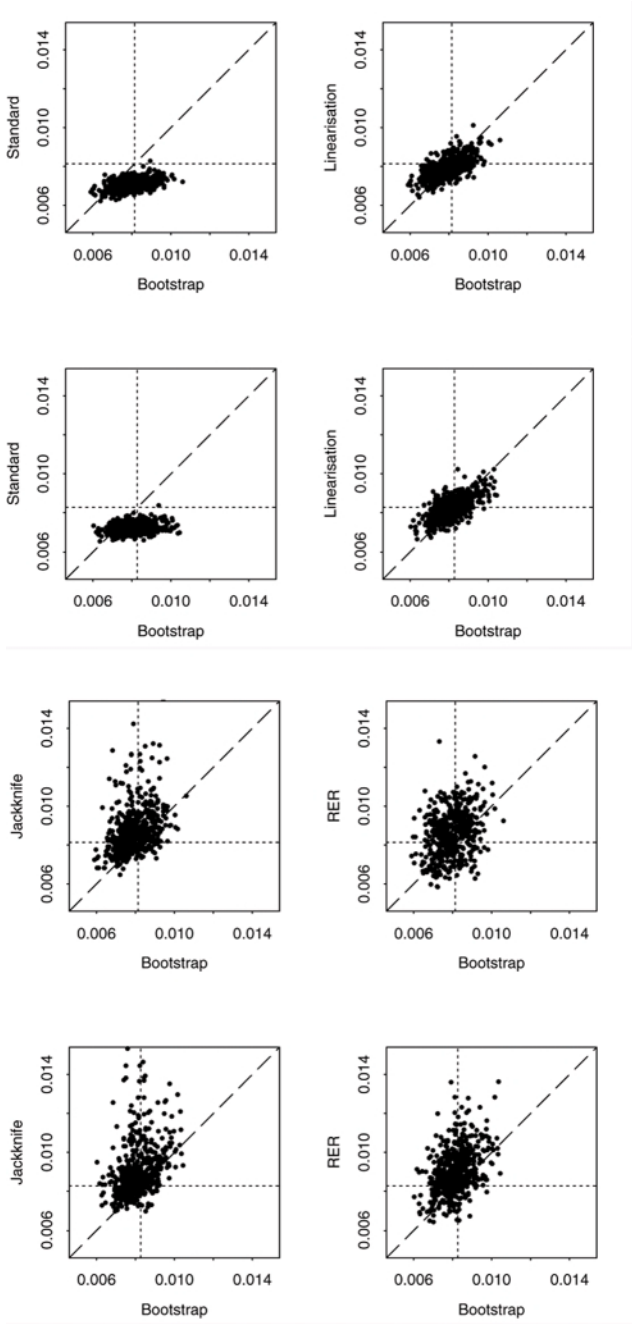

FIG 1. — Écarts-types pour le taux de chômage (ligne du haut) et changement du taux de chômage (ligne du bas) ; la droite trait-tillée $y = x$, les lignes pointillés sont les 'vrais' écarts-types d'échantillonnage. De Canty et Davison (1999).

proportionel itératif. Par contre le jackknife présente des propriétés plutôt étranges, à savoir que cinq pas itératifs produisent des estimateurs fortement variables.

7.3. Imputation

Kovar *et al.* (1988) ont effectué une étude de simulation dans le but de tester la méthode de réplication équilibrée répétée ajustée au mécanisme d'imputation. L'échantillonnage est stratifié par grappes avec $n_h = 2$ pour $H = 32$ strates, et des taux de données manquantes de 10%, 20%, 30%, 40%, et 50%. Pour la moyenne et la médiane, la méthode ajustée de réplication équilibrée répétée produit de bons biais relatifs et coefficients de variation, contrairement à celui non ajusté qui sous-estime la variance.

7.4. Imputation et calage

Les simulations ci-dessus ne traitent pas du cas de données calées et imputées. Sur la base d'une simulation réaliste fondée sur l'enquête du budget des ménages suisses de l'année 1998 (Renfer, 2001), nous considérons l'estimateur calé et imputé de Horvitz–Thompson du total des dépenses pour le pain et pour les produits céréaliers, sur la base de données complètes de $N = 9275$ ménages dans $H = 7$ strates de tailles variées. Nous disposons également d'un ensemble de 14 variables auxiliaires sur chaque ménage, dont 10 marges de population sont connues. Pour la simulation, nous considérons les $N = 9275$ ménages comme la population totale, de dépense totale connue.

Nous effectuons alors un échantillonnage stratifié sans remise et avec probabilités d'inclusion égales 1/8 à l'intérieur de 6 strates, et 3/8 dans l'autre strate, donnant une taille d'échantillon de 1332. Nous créons la non-réponse partielle de manière aléatoire équiprobable à travers l'échantillon entier. Nous calculons les estimations sur chacun des 500 échantillons ainsi simulés.

Le bootstrap utilise 100 répliques obtenu par la procédure de Shao et Sitter (1996), c'est-à-dire, avec des réponses manquantes imputées de manière déterministe en utilisant un modèle linéaire ajusté aux répondants pleinement bootstrappés, et avec le jeu de données imputé calé aux poids par régression linéaire. Une simulation préalable a montré que 100 répliques étaient raisonnables.

Pour que le jackknife ait la même complexité de calcul que le bootstrap, nous utilisons à peu près le même nombre de blocs de suppression que lorsqu'on applique le jackknife de bloc avec remise, soit 13 blocs choisis aléatoirement dans chaque strate, donnant environ 91 calculs en tout pour chaque estimation de variance jackknife.

Nous avons appliqué deux formes de réplication équilibrée répétée, la première utilisant une séparation aléatoire de chaque strate en deux moitiés pour chaque réplication ; aucun facteur de Fay n'était utilisé mais les poids pour les observations incluses dans la réplique étaient multipliées par un facteur de deux avant le calage. La deuxième forme, la réplication équilibrée répétée groupée de manière répétitive, prend la moyenne des estimations de variance avec 13 répétitions.

L'estimateur de linéarisation est celui donné par (10) et (14).

Nous avons appliqué les formules standard pour l'imputation multiple, en utilisant 30 imputations aléatoires à partir d'un modèle linéaire ajusté aux données complètes ; pour l'imputation paramétrique nous avons utilisé un modèle d'erreur normal homoscedastique, où les valeurs des paramètres de régression et de variance changent aléatoirement et indépendemment suivant les distributions ajustées normales et chi-carrées entre simulations ; tandis que les erreurs d'imputations non paramétriques ont été simulées suivant un résidu bootstrap basé sur un modèle (Davison et Hinkley, 1997, p. 262).

Les Figures 2 et 3 comparent les performances de ces techniques d'estimation de variance pour des taux de données manquantes de 0%, 20%, 40%, and 60%. Le jackknife de bloc sous-estime les vraies variances, qui sont systématiquement surestimées par la réplication équilibrée répétée groupée de manière répétitive.

La réplication équilibrée répétée standard est clairement fortement variable, en accord avec les résultats de Rao et Shao (1996), mais la réplication équilibrée répétée groupée de §5.4 est plus stable, bien qu'elle surestime la variance lorsque beaucoup de données manquent.

La linéarisation fonctionne bien pour de faible taux de données manquantes, et de manière générale elle produit des variances qui sont plutôt trop basses mais assez stables. Le bootstrap se comporte bien en terme à la fois de biais et de variance, et semble être la meilleure de ces méthodes. Les méthodes d'imputation multiple produisent des erreurs dans des directions opposées, lorsqu'il manque une fraction substantielle de données.

De manière générale l'approche bootstrap de Shao et Sitter (1996) et la méthode de linéarisation de §3.1 semblent les meilleures en termes de biais et de stabilité. Si on se préoccupe du temps calcul, l'avantage va à la linéarisation, qui est jusqu'à cinquante fois plus rapide que les autres méthodes considérées dans cette étude.

Une simulation séparée a été effectuée pour vérifier l'effet de l'utilisation de strates artificielles pour la réplication équilibrée répétée. Chacune des $H = 7$ strates a été divisée en quatre parties, et les estimateurs de variance ont été calculés pour 1000 échantillons avec des réponses complètes. Les résultats montrent que l'introduction des strates artificielles améliore beaucoup les estimateurs de variances de réplication équilibrée répétée.

8. Discussion

Il semble de manière générale que les estimateurs de variance bootstrap et de linéarisation sont des plus prometteurs pour l'utilisation dans les enquêtes complexes. Le bootstrap a l'avantage d'être un outil à utilités multiples, d'être applicable dans beaucoup de situations, et d'être utilisable à la fois pour des statistiques lisses et non lisses. De plus, et contrairement au jackknife ou à la réplication équilibrée répétée, le nombre de recalculs requis est plutôt contrôlé par l'utilisateur que déterminé par la méthode. Ses inconvénients principaux

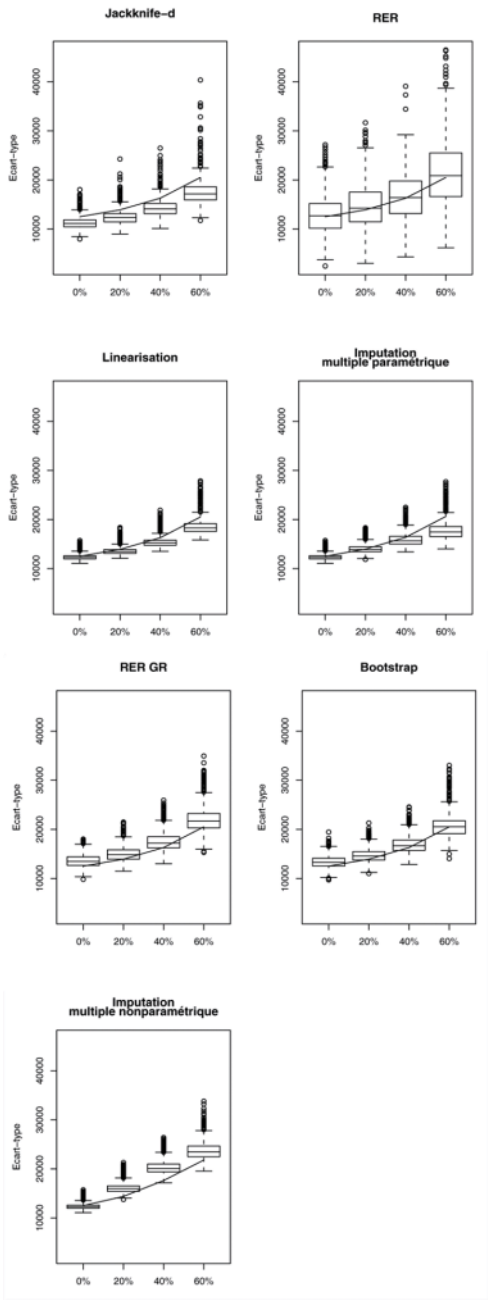

FIG 2. — Comparaison entre estimateurs de variance, en fonction de la proportion de données manquantes, pour estimateurs calés et imputés. Simulation basée sur l'enquêtes du budget des ménages suisses en 1998. La courbe montre les vraies variances, estimée à partir de 10,000 simulations, et les boxplots montrent les écart-types calculés pour 500 échantillons.

sont le fardeau de calcul qu'il nécessite, particulièrement lorsqu'il est utilisé avec l'imputation, et le fait qu'il nécessite une programmation spéciale lorsque les fractions d'échantillonnages f_h sont grandes.

La linéarisation demande des dérivations analytiques, mais elle n'entraîne pas de rééchantillonnage et est ainsi beaucoup plus rapide que le bootstrap. Son application peut être compliquée pour des statistiques non lisses, mais la règle d'enchaînement peut être utilisée pour obtenir des valeurs d'influence pour des estimateurs complexes.

La réplication équilibrée répétée et le jackknife sont comparables, mais se comportent en général moins bien que les autres méthodes. Il semble nécessaire de les ajuster pour obtenir de meilleures performances.

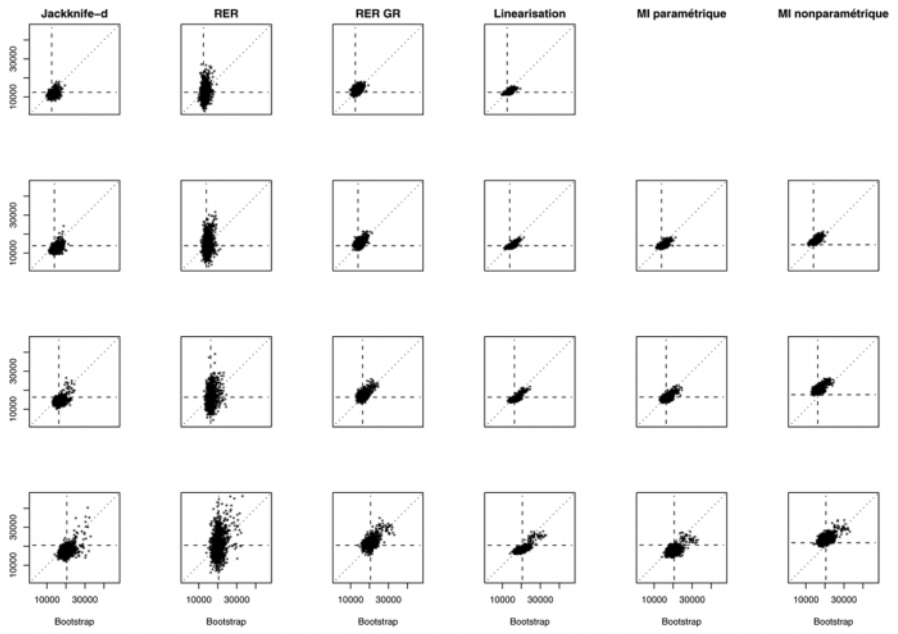

FIG 3. — Comparaison des erreurs standard de rééchantillonnage en présence de calage et d'imputation ; de haut en bas 0%, 20%, 40%, 60% non-réponses unité. Les traits-tillés sont les 'vrais' écarts-types d'échantillonnage, et la ligne pointillée montre $x = y$. Simulation basée sur l'enquête de budget des ménages suisses de 1998.

La linéarisation et le bootstrap s'appliquent aux estimateurs de changement basés sur des enquêtes de panel en séparant les unités en trois parties : celles présentes seulement à la première occasion d'échantillonnage ($k = 1$), celles présentes lors des deux occasions ($k = 2$), et celles présentes seulement à la deuxième occasion ($k = 3$) (Canty et Davison, 1999). Chacune des H strates originales est séparée en trois parties et l'estimateur est calculé en fonction. Si nous ignorons les corrections pour les fractions d'échantillonnage, l'estimateur

de variance de linéarisation est donné par

$$\sum_{h=1}^{H}\left\{\frac{1}{n_{h1}(n_{h1}-1)}\sum_{j=1}^{n_{h1}}(l_{hj}^{1})^2+\frac{1}{n_{h2}(n_{h2}-1)}\sum_{j=1}^{n_{h2}}(l_{hj}^{2})^2+\frac{1}{n_{h3}(n_{h3}-1)}\sum_{j=1}^{n_{h3}}(l_{hj}^{3})^2\right\},$$

où la h-ème strate contient n_{h1} unités présentes à la première occasion, n_{h2} présentes aux deux occasions, et n_{h3} présentes à la deuxième occasion seulement, et où l_{hj}^{k} sont les influence empiriques correspondantes. Un argument similaire s'applique au bootstrap, pour lequel on applique le rééchantillonnage avec ces mêmes sous-strates, et aux autres méthodes de rééchantillonnage.

Remerciements

Cette recherche a été soutenue par l'Union Européenne, dans le contexte du projet DACSEIS (www.dacseis.de), et a été présentée lors du quatrième Colloque francophone sur les sondages en mai 2005 à Québec. Nous remercions vivement les organisateurs du Colloque ainsi que les autres membres de l'équipe DACSEIS. Nous remercions également Andreas Kaufmann pour son aide précieuse. Les commentaires pertinents des réviseurs et de l'éditeur ont beaucoup amélioré l'article.

Références

Berger Y. G. et Rao J. N. K. (2006) Adjusted jackknife for imputation under unequal probability sampling without replacement. *Journal of the Royal Statistical Society series B* **68**, 531–547.

Berger Y. G. et Skinner C. J. (2003) Variance estimation for a low income proportion. *Applied Statistics* **52**, 457–468.

Berger Y. G. et Skinner C. J. (2005) A jackknife variance estimator for unequal probability sampling. *Journal of the Royal Statistical Society, Series B* **67**, 79–89.

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

Binder D. A. (1996) Linearization methods for single phase and two-phase samples : A cookbook approach. *Survey Methodology* **22**, 17–22.

Boonstra H. J. H. et Nieuwenbroek N. (2003) An empirical comparison of BRR and linearization variance estimators. Centraal Bureau voor de Statistiek, The Netherlands.

Booth J. G., Butler R. W. et Hall P. (1994) Bootstrap methods for finite populations. *Journal of the American Statistical Association* **89**, 1282–1289.

Campbell C. (1980) A different view of finite population estimation. In *Proceedings of the Section on Survey Research Methods*, pp. 319–324. Alexandria, Virginia : American Statistical Association.

Canty A. J. et Davison A. C. (1999) Resampling-based variance estimation for labour force surveys. *The Statistician* **48**, 379–391.

CHAO M. T. et LO S. H. (1985) A bootstrap method for finite populations. *Sankyā A* **47**, 399–405.

CHEN J. et SHAO J. (2001) Jackknife variance estimation for nearest-neighbor imputation. *Journal of the American Statistical Association* **96**, 260–269.

COCHRAN W. G. (1977) *Sampling Techniques*. Third edition. New York : Wiley.

DAVISON A. C. et HINKLEY D. V. (1997) *Bootstrap Methods and Their Application*. Cambridge : Cambridge University Press.

DEVILLE J. C. (1999) Variance estimation for complex statistics and estimators : Linearization and residual techniques. *Survey Methodology* **25**, 193–203.

DEVILLE J. C. et SÄRNDAL C. E. (1992) Calibration estimators in survey sampling. *Journal of the American Statistical Association* **87**, 376–382.

DEVILLE J. C., SÄRNDAL C. E. et SAUTORY O. (1993) Generalized raking procedures in survey sampling. *Journal of the American Statistical Association* **88**, 1013–1020.

DIPPO C. S., FAY R. E. et MORGANSTEIN D. H. (1984) Computing variances from complex samples with replicate weights. In *Proceedings of the Section on Survey Research Methods*, pp. 489–494. Washington DC : American Statistical Association.

DROESBEKE J.-J., FICHET B. et TASSI P. (eds) (1987) *Les sondages*. First edition. Economica.

EFRON B. (1979) Bootstrap methods : Another look at the jackknife. *Annals of Statistics* **7**, 1–26.

EFRON B. (1982) *The Jackknife, the Bootstrap, and other Resampling Plans*. Philadephia : SIAM.

FAY R. E. (1989) Theory and application of replicate weighting for variance calculations. In *Proceedings of the Social Statistics Section*, pp. 212–217. American Statistical Association.

FAY R. E. (1996) Alternative paradigms for the analysis of imputed survey data. *Journal of the American Statistical Association* **91**, 490–498.

GROSS S. (1980) Median estimation in sample surveys. In *Proceedings of the Section on Survey Research Methods*, pp. 181–184. Alexandria, VA : American Statistical Association.

GUPTA V. K. et NIGAM A. K. (1987) Mixed orthogonal arrays for variance estimation with unequal numbers of primary selections per stratum. *Biometrika* **74**, 735–742.

GURNEY M. et JEWETT R. S. (1975) Constructing orthogonal replications for standard errors. *Journal of the American Statistical Association* **70**, 819–821.

HALL P. G. (2003) A short prehistory of the bootstrap. *Statistical Science* **18**, 158–167.

HAMPEL F. R., RONCHETTI E. M., ROUSSEEUW P. J. et STAHEL W. A. (1986) *Robust Statistics : The Approach Based on Influence Functions*. New York : Wiley.

HARTIGAN J. A. (1969) Using subsample values as typical values. *Journal of the American Statistical Association* **64**, 1303–1317.

HUBER P. J. (1981) *Robust Statistics*. New York : Wiley.

JUDKINS D. R. (1990) Fay's method of variance estimation. *Journal of Official Statistics* **6**, 223–239.

KIM J. K. et SITTER R. R. (2003) Efficient replication variance estimation for two-phase sampling. *Statistica Sinica* **13**, 641–653.

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.

KREWSKI D. et RAO J. N. K. (1981) Inference from stratified samples : Properties of the linearization, jackknife and balanced repeated replication methods. *Annals of Statistics* **9**, 1010–1019.

LAHIRI P. (2003) On the impact of bootstrap in survey sampling and small-area estimation. *Statistical Science* **18**, 199–210.

MCCARTHY P. J. (1969) Pseudo-replication : Half samples. *Review of the Interational Statistics Institute* **37**, 239–264.

MCCARTHY P. J. et SNOWDEN C. B. (1985) The bootstrap and finite population sampling. *Vital and Health Statistics* **2**, 2–95.

MILLER R. G. (1974) The jackknife - A review. *Biometrika* **61**, 1–15.

NIELSEN S. F. (2003) Proper and improper multiple imputation (with discussion). *International Statistical Review* **71**, 593–627.

POLITIS D. N., ROMANO J. P. et WOLF M. (1999) *Subsampling*. New York : Springer-Verlag.

PRESNELL B. et BOOTH J. G. (1994) Resampling methods for sample surveys. Technical Report 470, Department of Statistics, University of Florida, Gainesville.

QUENOUILLE M. H. (1949a) Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society, Series B* **11**, 68–84.

QUENOUILLE M. H. (1949b) Notes on bias in estimation. *Biometrika* **43**, 353–360.

RAO J. N. K. (1988) Variance estimation in sample surveys. In *Handbook of Statistics, Volume 6*, ed. P. R. K. amd C. R. Rao, pp. 427–447. Amsterdam : Elsevier Science.

RAO J. N. K. et SHAO J. (1992) Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* **79**, 811–822.

RAO J. N. K. et SHAO J. (1996) On balanced half-sample variance estimation in stratified random sampling. *Journal of the American Statistical Association* **91**, 343–348.

RAO J. N. K. et SHAO J. (1999) Modified balanced repeated replication for complex survey data. *Biometrika* **86**, 403–415.

RAO J. N. K. et WU C. F. J. (1985) Inference from stratified samples : Second-order analysis of three methods for nonlinear statistics. *Journal of the American Statistical Association* **80**, 620–630.

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) Some recent work on resampling methods for complex surveys. *Survey Methodology* **18**, 209–217.

RENFER J.-P. (2001) *Description and process of the Household and Budget Survey of 1998 (HBS 1998)*. Swiss Federal Statistical Office. 1-19.

RUBIN D. B. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York : Wiley.

RUBIN D. B. (1996) Multiple imputation after 18+ years. *Journal of the American Statistical Association* **91**, 473–489. with discussion, 507-515, and rejoinder, 515-517.

RUST K. F. et RAO J. N. K. (1996) Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* **5**, 283–310.

SHAO J. (2003) Impact of the bootstrap on sample surveys. *Statistical Science* **18**, 191–198.

SHAO J., CHEN Y. et CHEN Y. (1998) Balanced repeated replication for stratified multistage survey data under imputation. *Journal of the American Statistical Association* **93**, 819–831.

SHAO J. et RAO J. N. K. (1994) Standard errors for low income proportions estimated from stratified multistage samples. *Sankyā B* **55**, 393–414.

SHAO J. et SITTER R. R. (1996) Bootstrap for imputed survey data. *Journal of the American Statistical Association* **91**, 1278–1288.

SHAO J. et TU D. (1995) *The Jackknife and Bootstrap*. New York : Springer-Verlag.

SHAO J. et WU C. F. J. (1989) A general theory for jackknife variance estimation. *Annals of Statistics* **17**, 1176–1197.

SHAO J. et WU C. F. J. (1992) Asymptotic properties of the balanced repeated replication method for sample quantiles. *Annals of Statistics* **20**, 1571–1593.

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.

SITTER R. R. (1993) Balanced repeated replications based on orthogonal multi-arrays. *Biometrika* **80**, 211–221.

TUKEY J. W. (1958) Bias and confidence in not quite large samples (abstract). *Annals of Mathematical Statistics* **29**, 614.

WOLTER K. M. (1985) *Introduction to Variance Estimation*. New York : Springer-Verlag.

WU C. F. J. (1991) Balanced repeated replications based on mixed orthogonal arrays. *Biometrika* **78**, 181–188.

ZHANG P. (2003) Multiple imputation : Theory and method. *International Statistical Review* **71**, 581–592.