

Article

De la performance des estimateurs sur petits domaines autocalés sous le modèle au niveau du domaine de Fay-Herriot

par Yong You, J.N.K. Rao et Mike Hidiroglou

Juin 2013



Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à infostats@statcan.gc.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-877-287-4369 |

Programme des services de dépôt

- | | |
|---------------------------|----------------|
| Service de renseignements | 1-800-635-7943 |
| Télécopieur | 1-800-565-7757 |

Comment accéder à ce produit

Le produit n° 12-001-X au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.gc.ca et de parcourir par « Ressource clé » > « Publications ».

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « À propos de nous » > « Notre organisme » > « Offrir des services aux Canadiens ».

Publication autorisée par le ministre responsable de
Statistique Canada

© Ministre de l'Industrie, 2013

Tous droits réservés. L'utilisation de la présente
publication est assujettie aux modalités de l'entente
de licence ouverte de Statistique Canada
(<http://www.statcan.gc.ca/reference/licence-fra.html>).

This publication is also available in English.

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, ses entreprises, ses administrations et les autres établissements. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Signes conventionnels

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- | | |
|----------------|---|
| . | indisponible pour toute période de référence |
| .. | indisponible pour une période de référence précise |
| ... | n'ayant pas lieu de figurer |
| 0 | zéro absolu ou valeur arrondie à zéro |
| 0 ^s | valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie |
| p | provisoire |
| r | révisé |
| x | confidentiel en vertu des dispositions de la <i>Loi sur la statistique</i> |
| E | à utiliser avec prudence |
| F | trop peu fiable pour être publié |
| * | valeur significativement différente de l'estimation pour la catégorie de référence ($p < 0,05$) |

De la performance des estimateurs sur petits domaines autocalés sous le modèle au niveau du domaine de Fay-Herriot

Yong You, J.N.K. Rao et Mike Hidirolou¹

Résumé

Nous considérons deux méthodes distinctes d'autocalage pour l'estimation des moyennes de petit domaine fondée sur le modèle au niveau du domaine de Fay-Herriot (FH), à savoir la méthode de You et Rao (2002) appliquée au modèle FH et la méthode de Wang, Fuller et Qu (2008) basée sur des modèles augmentés. Nous établissons un estimateur de l'erreur quadratique moyenne de prédiction (EQMP) de l'estimateur de You-Rao (YR) d'une moyenne de petit domaine qui, sous le modèle vrai, est correct jusqu'aux termes de deuxième ordre. Nous présentons les résultats d'une étude en simulation du biais relatif de l'estimateur de l'EQMP de l'estimateur YR et de l'estimateur de l'EQMP de l'estimateur de Wang, Fuller et Qu (WFQ) obtenu sous un modèle augmenté. Nous étudions aussi l'EQMP et les estimateurs de l'EQMP des estimateurs YR et WFQ obtenus sous un modèle mal spécifié.

Mots clés : Modèle augmenté; meilleure prédiction linéaire sans biais empirique; erreur quadratique moyenne de prédiction; erreur de spécification du modèle.

1 Introduction

On parle de petit domaine quand la taille de l'échantillon d'une sous-population ou d'un domaine est trop petite pour produire des estimations directes fiables des paramètres du domaine. Il faut alors utiliser des estimateurs indirects fondés sur un modèle qui s'appuient sur des données d'échantillon empruntées à des domaines apparentés grâce à des modèles de lien afin de réaliser un important gain d'efficacité par rapport aux estimateurs directs. Dans le présent article, nous nous concentrerons sur un modèle de base au niveau du domaine, appelé modèle de Fay-Herriot (FH) (Fay et Herriot 1979), et sur les meilleurs prédicteurs linéaires sans biais empiriques (EBLUP, pour *Empirical Best Linear Unbiased Predictor*) associés des moyennes de petit domaine.

1. Yong You, Division de la recherche et de l'innovation en statistique, Statistique Canada, Ottawa, Canada. Courriel : yong.you@statcan.gc.ca; J.N.K. Rao, École de mathématiques et de statistique, Université Carleton, Ottawa, Canada; Mike Hidirolou, Division de la recherche et de l'innovation en statistique, Statistique Canada, Ottawa, Canada.

Ces EBLUP ne concordent pas forcément avec les estimateurs directs que l'on privilégie en pratique pour les agrégats parce qu'ils sont fondés sur des échantillons de suffisamment grande taille pour satisfaire les exigences de fiabilité et qu'ils ne dépendent pas de modèles. Par conséquent, on procède souvent à un calage pour contraindre les estimations à concorder. Le présent article porte sur deux méthodes d'autocalage fondées sur des modifications des prédicteurs EBLUP. La première méthode (WFQ), que l'on doit à Wang, Fuller et Qu (Wang et coll. 2008), s'appuie sur un modèle FH augmenté et sur les EBLUP associés, appelés estimateurs WFQ. La deuxième méthode (YR), également proposée par Wang, Fuller et Qu, est fondée sur une approche utilisée par You et Rao (2002) dans le contexte de modèles au niveau de l'unité. Les estimateurs YR s'obtiennent en modifiant les estimateurs optimaux des paramètres de régression utilisés dans les EBLUP afin d'imposer la concordance. Étant donné les modifications des EBLUP, les erreurs quadratiques moyennes de prédiction (EQMP) des estimateurs WFQ et YR autocalés seront plus importantes.

Le présent article a deux grands objectifs. Le premier consiste à comparer les EQMP et leurs estimateurs pour les estimateurs YR et WFQ des moyennes de petit domaine. À la section 2, nous examinons les expressions des EQMP et des estimateurs associés pour l'EBLUP et pour l'estimateur WFQ. À la section 3, nous développons les expressions de l'EQMP et de son estimateur pour l'estimateur YR. À la section 4, nous comparons les EQMP et leurs estimateurs au moyen d'une étude en simulation.

Le deuxième objectif est d'examiner les propriétés des estimateurs WFQ et YR et des estimateurs de leur EQMP lorsque le modèle de lien est mal spécifié dans le modèle FH en raison de l'omission d'une variable. Wang, Fuller et Qu ont aussi étudié l'effet de l'erreur de spécification du modèle de lien dans un cas particulier, pour lequel ils ont montré que l'estimateur YR produit un grand biais, mais non l'estimateur WFQ. Cependant, ce résultat tenait au fait que, dans leur étude en simulation, la variable d'augmentation était fortement corrélée avec une covariable omise (coefficient de corrélation $\rho = 0,983$). Par conséquent, leur modèle

augmenté était, en réalité, proche du modèle vrai, d'où la performance supérieure de l'estimateur WFQ. À la section 4, nous considérons une variable omise différente, qui est faiblement corrélée à la variable d'augmentation ($\rho = -0,116$), de sorte que le modèle augmenté est également mal spécifié. Nous comparons les biais, ainsi que les EQMP et leurs estimateurs pour les EBLUP et les estimateurs YR et WFQ obtenus sous le modèle mal spécifié.

2 Estimateurs EBLUP et WFQ

Supposons que nous ayons m petits domaines et des estimateurs directs sans biais sous le plan, y_i , des moyennes de domaine $\theta_i, i = 1, \dots, m$. Le modèle FH se rapporte à (y_i, θ_i) et aux variables auxiliaires au niveau du domaine associées $x_i = (x_{i1}, \dots, x_{ip})'$ avec $x_{i1} = 1$. Sous l'hypothèse d'échantillonnage indépendant dans les divers domaines, le modèle FH peut s'écrire sous forme d'un modèle linéaire mixte donné par

$$y_i = \theta_i + e_i = x_i' \beta + v_i + e_i, i = 1, \dots, m, \quad (2.1)$$

où $\theta_i = x_i' \beta + v_i$ est le modèle de lien, et e_i est l'erreur d'échantillonnage de moyenne 0 et de variance connue σ_e^2 , qui est indépendante de l'effet aléatoire propre au domaine v_i . L'échantillonnage est indépendant dans les divers domaines et les v_i sont supposés indépendants et identiquement distribués de moyenne 0 et de variance σ_v^2 .

Le meilleur prédicteur linéaire sans biais (BLUP, pour *Best Linear Unbiased Predictor*) de θ_i sous le modèle « vrai » (2.1) est donné par

$$\tilde{\theta}_i = \gamma_i y_i + (1 - \gamma_i) x_i' \tilde{\beta}, \quad (2.2)$$

où $\gamma_i = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2)$ et $\tilde{\beta}$ est l'estimateur par les moindres carrés pondérés (MCP) optimal de β donné par

$$\tilde{\beta} = \left(\sum_{i=1}^m \gamma_i x_i x_i' \right)^{-1} \left(\sum_{i=1}^m \gamma_i x_i y_i \right), \quad (2.3)$$

voir Rao (2003, page 116). L'estimateur $\tilde{\theta}_i$ dépend de la variance sous le modèle σ_v^2 inconnue, et en remplaçant σ_v^2 dans (2.2) par un estimateur approprié $\hat{\sigma}_v^2$, nous obtenons l'EBLUP :

$$\hat{\theta}_i^{\text{EBLUP}} = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i) x_i' \hat{\beta}, \quad (2.4)$$

où $\hat{\gamma}_i = \hat{\sigma}_v^2 / (\hat{\sigma}_v^2 + \sigma_i^2)$ et $\hat{\beta}$ s'obtient à partir de (2.3) en remplaçant γ_i par $\hat{\gamma}_i$. Dans le présent article, nous utilisons l'estimateur du maximum de vraisemblance restreint (REML pour *Restricted Maximum Likelihood*) de σ_v^2 , en supposant que v_i et e_i suivent une loi normale. La somme pondérée $\sum_{i=1}^m w_i \hat{\theta}_i^{\text{EBLUP}}$ des EBLUP (2.4) ne concorde pas nécessairement avec l'estimateur direct pondéré correspondant de l'agrégat $\sum_{i=1}^m w_i y_i$, où les w_i sont des poids spécifiés au préalable de façon que $\sum_{i=1}^m w_i y_i$ soit un estimateur de l'agrégat (total ou moyenne) convergent sous le plan. Un écart important entre $\sum_{i=1}^m w_i \hat{\theta}_i^{\text{EBLUP}}$ et $\sum_{i=1}^m w_i y_i$ peut être le signe d'un échec du modèle dont il faudra s'occuper avant de procéder au calage, comme l'a souligné le rédacteur associé.

Un estimateur de l'erreur quadratique moyenne de prédiction $\text{EQMP}(\hat{\theta}_i^{\text{EBLUP}}) = E(\hat{\theta}_i^{\text{EBLUP}} - \theta_i)^2$ correct jusqu'aux termes de deuxième ordre, sous l'estimation par la méthode du REML, est donné par

$$\text{eqmp}(\hat{\theta}_i^{\text{EBLUP}}) = g_{1i} + g_{2i} + 2g_{3i}, \quad (2.5)$$

où $g_{1i} = \hat{\gamma}_i \sigma_i^2$ est le terme principal d'ordre $O(1)$, et g_{2i} et g_{3i} sont des termes d'ordre plus faible $O(m^{-1})$ qui tiennent compte de la variabilité de $\hat{\beta}$ et $\hat{\sigma}_v^2$ respectivement (Rao 2003, page 128). Nous avons

$$g_{2i} = (1 - \hat{\gamma}_i)^2 x_i' \hat{V}(\tilde{\beta}) x_i = \hat{\sigma}_v^2 (1 - \hat{\gamma}_i)^2 x_i' \left(\sum_{i=1}^m \hat{\gamma}_i x_i x_i' \right)^{-1} x_i$$

et

$$g_{3i} = 2(\sigma_i^2)^2 (\hat{\sigma}_v^2 + \sigma_i^2)^{-3} \left\{ \sum_{i=1}^m (\hat{\sigma}_v^2 + \sigma_i^2)^{-2} \right\}^{-1},$$

où $\hat{V}(\tilde{\beta})$ est l'estimateur de $V(\tilde{\beta}) = \sigma_v^2 \left(\sum_{i=1}^m \gamma_i x_i x_i' \right)^{-1}$. L'estimateur de l'EQMP (2.5) est presque sans biais en ce sens que

$$E\{\text{eqmp}(\hat{\theta}_i^{\text{EBLUP}})\} = \text{EQMP}(\hat{\theta}_i^{\text{EBLUP}}) + O(m^{-2}).$$

Wang, Fuller et Qu ont obtenu l'estimateur EBLUP, $\hat{\eta}_i^{\text{EBLUP}}$, sous le modèle FH augmenté qui suit :

$$y_i = x_i' \delta + w_{ei} \lambda + u_i + e_i \equiv \eta_i + e_i, \quad (2.6)$$

où les effets aléatoires u_i sont indépendants avec $E(u_i) = 0$ et $\text{Var}(u_i) = \sigma_u^2$, et sont indépendants de e_i . La variable auxiliaire d'augmentation utilisée est $w_{ei} = w_i \sigma_i^2$. Ils ont montré que l'estimateur EBLUP de η_i sous le modèle augmenté (2.6), $\hat{\eta}_i^{\text{EBLUP}} \equiv \hat{\theta}_i^{\text{WFQ}}$, possède la propriété d'autocalage au sens où il satisfait

$$\sum_{i=1}^m w_i \hat{\theta}_i^{\text{WFQ}} = \sum_{i=1}^m w_i y_i.$$

L'EBLUP $\hat{\theta}_i^{\text{WFQ}}$ sous le modèle augmenté (2.6) s'obtient à partir de (2.4) en remplaçant x_i' par (x_i', w_{ei}) , $\hat{\sigma}_v^2$ par $\hat{\sigma}_u^2$ et $\hat{\beta}'$ par $(\hat{\delta}', \hat{\lambda})$. L'estimateur de $\text{EQMP}(\hat{\theta}_i^{\text{WFQ}})$ s'obtient de manière similaire à partir de (2.5). Sous le modèle vrai (2.1), l'EQMP de $\hat{\theta}_i^{\text{WFQ}}$ sera plus grande que celle de $\hat{\theta}_i^{\text{EBLUP}}$, mais son estimateur, $\text{eqmp}(\hat{\theta}_i^{\text{WFQ}})$, restera presque sans biais sous le modèle vrai, comme l'ont souligné le rédacteur associé et un examinateur, parce que le modèle vrai est un cas particulier du modèle augmenté avec $\lambda = 0$.

3 Estimateur YR

Wang, Fuller et Qu ont appliqué la méthode de You et Rao (2002) au modèle (2.1) et obtenu un estimateur de θ_i donné par

$$\hat{\theta}_i^{\text{YR}} = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i) x_i' \hat{\beta}^{\text{YR}}, \quad (3.1)$$

où $\hat{\beta}^{\text{YR}}$ s'obtient à partir de

$$\tilde{\beta}^{\text{YR}} = \left\{ \sum_{i=1}^m w_i (1 - \gamma_i) x_i x_i' \right\}^{-1} \left\{ \sum_{i=1}^m w_i (1 - \gamma_i) x_i y_i \right\} \quad (3.2)$$

en remplaçant γ_i par $\hat{\gamma}_i$. Notons que l'estimateur YR (3.1) a la même forme que l'EBLUP $\hat{\theta}_i^{\text{EBLUP}}$, mais qu'il utilise un estimateur non optimal pour β . Les estimateurs YR $\hat{\theta}_i^{\text{YR}}$ ont la propriété d'autocalage, c'est-à-dire $\sum_{i=1}^m w_i \hat{\theta}_i^{\text{YR}} = \sum_{i=1}^m w_i y_i$, puisque, en vertu de (3.2),

$$\sum_{i=1}^m w_i (1 - \gamma_i) (y_i - x_i' \tilde{\beta}^{\text{YR}}) = 0.$$

Cependant, l'EQMP de $\hat{\theta}_i^{\text{YR}}$ sera légèrement plus grande que celle de $\hat{\theta}_i^{\text{EBLUP}}$ fondé sur $\hat{\beta}$, parce que $\hat{\beta}$ est asymptotiquement plus efficace que $\tilde{\beta}^{\text{YR}}$.

Dans le cas de $\hat{\theta}_i^{\text{EBLUP}}$, l'estimateur de EQMP($\hat{\theta}_i^{\text{YR}}$) contient les termes g_{1i} , g_{2i} et g_{3i} . Nous devons estimer la variance de $\tilde{\beta}^{\text{YR}}$ afin d'obtenir le terme g_{2i} dans l'estimateur de EQMP($\hat{\theta}_i^{\text{YR}}$); les autres termes g_{1i} et g_{3i} ne sont pas affectés. Il découle de (3.2) que

$$V(\tilde{\beta}^{\text{YR}}) = \sigma_v^2 \left\{ \sum_{i=1}^m w_i (1 - \gamma_i) x_i x_i' \right\}^{-1} \left\{ \sum_{i=1}^m w_i^2 (1 - \gamma_i)^2 \gamma_i^{-1} x_i x_i' \right\} \left\{ \sum_{i=1}^m w_i (1 - \gamma_i) x_i x_i' \right\}^{-1}. \quad (3.3)$$

L'estimateur $\hat{V}(\tilde{\beta}^{\text{YR}})$ s'obtient en substituant $\hat{\sigma}_v^2$ et $\hat{\gamma}_i$ à σ_v^2 et γ_i dans (3.3).

L'estimateur de EQMP($\hat{\theta}_i^{\text{YR}}$) est donné par

$$\text{eqmp}(\hat{\theta}_i^{\text{YR}}) = g_{1i} + g_{2i}^{\text{YR}} + 2g_{3i}, \quad (3.4)$$

où

$$g_{2i}^{\text{YR}} = (1 - \hat{\gamma}_i)^2 x_i' \hat{V}(\tilde{\beta}^{\text{YR}}) x_i.$$

L'estimateur de l'EQMP (3.4) est presque sans biais sous le modèle vrai (2.1), comme l'estimateur de l'EQMP (2.5) de $\hat{\theta}_i^{\text{EBLUP}}$.

Remarque : Tout estimateur $\hat{y}_i$ of θ_i peut être ajusté comme il suit :

$$\hat{y}_i^a = \hat{y}_i + a_i \left(\sum_{i=1}^m w_i y_i - \sum_{i=1}^m w_i \hat{y}_i \right)$$

pour a_i spécifié de manière à satisfaire la contrainte de calage $\sum_{i=1}^m w_i \hat{y}_i^a = \sum_{i=1}^m w_i y_i$, où $\sum_{i=1}^m w_i a_i = 1$. En particulier, nous pouvons utiliser $\hat{y}_i = \hat{\theta}_i^{\text{EBLUP}}$ pour obtenir l'estimateur EBLUP ajusté. Comme l'ont mentionné WFQ, $\hat{\theta}_i^{\text{YR}}$ ainsi que $\hat{\theta}_i^{\text{WFQ}}$ sont des estimateurs de la forme $\hat{y}_i^a$ parce que $\sum_{i=1}^m w_i y_i - \sum_{i=1}^m w_i \hat{y}_i$ est égal à zéro si l'on fixe $\hat{y}_i$ égal à $\hat{\theta}_i^{\text{YR}}$ ou à $\hat{\theta}_i^{\text{WFQ}}$. Tout ensemble d'estimateurs $\{\hat{y}_i\}$ qui satisfait $\sum_{i=1}^m w_i y_i = \sum_{i=1}^m w_i \hat{y}_i$ possède la propriété d'autocalage.

4 Étude en simulation

Nous avons réalisé une petite étude en simulation du biais et de l'EQMP des estimateurs YR et WFQ, ainsi que du biais relatif et du coefficient de variation (CV) des estimateurs YR et WFQ de l'EQMP, en utilisant les conditions de simulation de WFQ. Comme ces auteurs, nous avons considéré $m = 50$ petits domaines (États), avec les tailles d'échantillon n_i variant de 7 à 58 et les variances d'échantillonnage connues $\sigma_i^2 = 16 / n_i$. Les poids w_i utilisés dans la contrainte de calage sont donnés par $w_i = \left(\sum_{i=1}^{50} P_i \right)^{-1} P_i$, où P_i est la population de l'État i indiquée dans le tableau 4.1 de WFQ.

Dans WFQ, le modèle vrai est donné par

$$y_i = \beta_0 + \beta_1 z_i + v_i + e_i, \quad (4.1)$$

avec $\beta_0 = 6$, $\beta_1 = 3$, $v_i \sim N(0, \sigma_v^2 = 1)$ et $e_i \sim N(0, 16 / n_i)$. WFQ ont utilisé $z_i = P_i^{0,2} - \left(\sum_{i=1}^{50} P_i^{0,2} \right) / 50$ comme valeurs des covariables dans le modèle vrai, et les z_i choisies sont fortement corrélées avec la variable augmentée pondérée associée $w_{ei} = w_i \sigma_i^2$; coefficient de corrélation $\rho = 0,983$.

Dans WFQ, le modèle mal spécifié choisi est le modèle de la moyenne

$$y_i = \mu + \tilde{v}_i + e_i \quad (4.2)$$

avec $\tilde{v}_i \sim N(0, \tilde{\sigma}_v^2)$ et $e_i \sim N(0, 16 / n_i)$, et en raison de ce choix, le modèle augmenté associé

$$y_i = \delta_0 + \lambda w_i \sigma_i^2 + u_i + e_i \quad (4.3)$$

est presque identique au modèle vrai (4.1).

Afin de refléter l'erreur de spécification au moyen du modèle augmenté, nous avons produit 50 valeurs $(z_1, \dots, z_{50})$ à partir de $N(1, 1)$ et les avons maintenues fixes dans les diverses simulations. Le coefficient de corrélation résultant entre les valeurs produites z_i et $w_i \sigma_i^2$ est égal à -0,116. Nous avons ensuite pris le modèle correspondant

$$y_i = \beta_0 + \beta_1 z_i + v_i + e_i \quad (4.4)$$

avec $\beta_0 = 6$, $\beta_1 = 3$, $v_i \sim N(0, 1)$ et $e_i \sim N(0, 16 / n_i)$ comme étant le modèle vrai et le modèle de la moyenne (4.2) comme étant le modèle mal spécifié. Nous avons considéré deux scénarios : i) le modèle vrai est donné par (4.4) et il est modélisé et spécifié correctement; ii) le modèle vrai est donné par (4.4) et il est modélisé et spécifié incorrectement comme étant le modèle de la moyenne (4.2).

Nous avons créé $R = 10\,000$ jeux de données $\{(y_i^{(r)}, z_i), i = 1, \dots, 50; r = 1, \dots, R\}$ sous le modèle vrai (4.4), où $y_i^{(r)} = \beta_0 + \beta_1 z_i + v_i^{(r)} + e_i^{(r)}$ avec $v_i^{(r)}$ tiré de la loi $N(0, 1)$ et $e_i^{(r)}$ de la loi $N(0, 16 / n_i)$, respectivement. Soit $\theta_i^{(r)} = \beta_0 + \beta_1 z_i + v_i^{(r)}$ la valeur de la vraie moyenne θ_i et $\hat{\theta}_i^{(r)}$ la valeur associée d'un estimateur $\hat{\theta}_i$ (EBLUP, YR ou WFQ) pour la r^e exécution de la simulation sous le scénario (i) ou (ii). Alors, le biais absolu et l'EQMP simulés de $\hat{\theta}_i$ sont donnés par

$$|B(\hat{\theta}_i)| = \left| R^{-1} \sum_{r=1}^R (\hat{\theta}_i^{(r)} - \theta_i^{(r)}) \right|$$

et

$$EQMP(\hat{\theta}_i) = R^{-1} \sum_{r=1}^R (\hat{\theta}_i^{(r)} - \theta_i^{(r)})^2.$$

Le biais relatif absolu (BRA) d'un estimateur de l'EQMP, $eqmp(\hat{\theta}_i)$, est calculé comme étant $|BR\{eqmp(\hat{\theta}_i)\}|$, où

$$BR\{eqmp(\hat{\theta}_i)\} = \frac{1}{R} \sum_{r=1}^R \{eqmp(\hat{\theta}_i^{(r)}) - EQMP(\hat{\theta}_i)\} / EQMP(\hat{\theta}_i),$$

où $\text{eqmp}(\hat{\theta}_i^{(r)})$ est la valeur de $\text{eqmp}(\hat{\theta}_i)$ pour la r^{e} exécution de la simulation. De même, le coefficient de variation (CV) de $\text{eqmp}(\hat{\theta}_i)$ est calculé comme

$$\text{CV}\{\text{eqmp}(\hat{\theta}_i)\} = \left[\frac{1}{R} \sum_{r=1}^R \{\text{eqmp}(\hat{\theta}_i^{(r)}) - \text{EQMP}(\hat{\theta}_i)\}^2 \right]^{\frac{1}{2}} / \text{EQMP}(\hat{\theta}_i).$$

Nous résumons les résultats de la simulation en présentant le premier quartile, le troisième quartile, la médiane et la moyenne des valeurs de $|B(\hat{\theta}_i)|$, $\text{EQMP}(\hat{\theta}_i)$, $|BR\{\text{eqmp}(\hat{\theta}_i)\}|$ %, et $\text{CV}\{\text{eqmp}(\hat{\theta}_i)\}$ %. Nous avons également regroupé les domaines en fonction des variances d'échantillonnage, $\sigma_i^2 = 16 / n_i$, en trois groupes, puis nous avons calculé les mesures susmentionnées séparément pour chaque groupe. Les résultats étaient très semblables à ceux obtenus sans groupement. Par conséquent, pour simplifier, nous ne présentons les résultats que pour le cas non groupé.

Le tableau 4.1 donne les mesures sommaires pour le biais absolu des estimateurs EBLUP, YR et WFQ, sous les scénarios (i) et (ii). Comme prévu, les valeurs des mesures du biais absolu sont toutes négligeables pour les trois estimateurs sous le scénario (i), confirmant l'absence de biais des estimateurs sous ce scénario. Par ailleurs, le biais absolu est grand relativement à l'erreur-type sous le scénario (ii) pour les trois estimateurs, ce qui donne une beaucoup plus grande EQMP que sous le scénario (i). Il est intéressant de noter, lorsque l'on examine les valeurs pour le scénario (ii) au tableau 4.1, que le biais absolu est presque le même pour les divers estimateurs, ce qui donne à penser que l'autocalage ne réduit peut-être pas le biais absolu en cas d'erreur de spécification du modèle. Une explication plausible est que l'estimateur YR donne le même poids, $\hat{\gamma}_i$, à l'estimateur direct; de même, l'estimateur WFQ attribue un poids proche de $\hat{\gamma}_i$. Nos résultats ne contredisent pas la conclusion de WFQ selon laquelle leur estimateur réduit le biais absolu, car leur modèle augmenté a quasiment éliminé l'erreur de spécification du modèle.

Tableau 4.1
Mesures sommaires du biais absolu des estimateurs EBLUP, YR et WFQ

Mesure	EBLUP	YR	WFQ
Scénario (i)			
1 ^{er} quartile	0,002	0,002	0,002
Médiane	0,005	0,005	0,005
Moyenne	0,007	0,007	0,007
3 ^e quartile	0,009	0,010	0,009
Scénario (ii)			
1 ^{er} quartile	0,092	0,085	0,083
Médiane	0,144	0,153	0,152
Moyenne	0,254	0,255	0,251
3 ^e quartile	0,316	0,311	0,314

Il convient aussi de mentionner que l'estimateur EBLUP ajusté $\hat{y}_i^a$ défini dans la remarque en utilisant $\hat{y}_i = \hat{\theta}_i^{\text{EBLUP}}$ pourrait donner de meilleurs résultats que les estimateurs YR et WFQ en ce qui concerne le biais absolu sous le scénario (ii). Par exemple, si l'estimateur EBLUP donne une sous-estimation dans la plupart des domaines, le terme de correction $a_i \left(\sum_{i=1}^m w_i y_i - \sum_{i=1}^m w_i \theta_i^{\text{EBLUP}} \right)$ sera positif et compensera donc la sous-estimation. Par ailleurs, le terme de correction est nul pour les estimateurs avec autocalage YR et WFQ.

Le tableau 4.2 donne les mesures sommaires de l'EQMP des estimateurs EBLUP, YR et WFQ, sous les scénarios (i) et (ii). Il montre que, sous le scénario (i), les EQMP des trois estimateurs EBLUP, YR et WFQ sont pratiquement les mêmes, excepté que celle de l'EBLUP est légèrement plus faible que prévu. Sous le scénario (ii), nous observons de nouveau très peu de différence entre les EQMP des trois estimateurs. Toutefois, l'erreur de spécification du modèle les accroît chacune considérablement. Par exemple, pour le troisième quartile (correspondant aux faibles valeurs de n_i), la valeur passe de 0,621 à 1,100 pour l'estimateur EBLUP, de 0,626 à 1,094 pour l'estimateur YR et de 0,634 à 1,117 pour l'estimateur WFQ. Cette augmentation de l'EQMP est due au biais induit par l'erreur de spécification du modèle.

Tableau 4.2
Mesures sommaires pour l'EQMP des estimateurs EBLUP, YR et WFQ

Mesure	EBLUP	YR	WFQ
Scénario (i)			
1 ^{er} quartile	0,403	0,406	0,403
Médiane	0,471	0,476	0,472
Moyenne	0,499	0,505	0,507
3 ^e quartile	0,621	0,626	0,634
Scénario (ii)			
1 ^{er} quartile	0,562	0,567	0,567
Médiane	0,751	0,748	0,752
Moyenne	0,885	0,886	0,883
3 ^e quartile	1,100	1,094	1,117

Le tableau 4.3 donne les mesures sommaires du biais relatif absolu en pourcentage et du CV en pourcentage des estimateurs de l'EQMP associés aux estimateurs EBLUP, YR et WFQ, sous le scénario (i) correspondant à l'absence d'erreur de spécification du modèle vrai (4.4). Le tableau 4.4 donne les résultats pour les estimateurs de l'EQMP sous le scénario (ii), qui correspond à l'erreur de spécification du modèle vrai (4.4) sous forme du modèle de la moyenne (4.2).

Tableau 4.3
Mesures sommaires pour le biais relatif absolu (%) et le CV (%) des estimateurs de l'EQMP associés aux estimateurs EBLUP, YR et WFQ sous le scénario (i) : Modèle (4.4) spécifié correctement

Mesure	EBLUP	YR	WFQ
Biais relatif absolu (BRA) %			
1 ^{er} quartile	0,678	0,650	0,627
Médiane	1,186	1,216	1,173
Moyenne	1,354	1,349	1,372
3 ^e quartile	1,674	1,716	1,863
Coefficient de variation (CV) %			
1 ^{er} quartile	12,93	12,44	13,19
Médiane	15,64	15,40	15,78
Moyenne	16,02	15,69	15,78
3 ^e quartile	19,39	19,07	19,05

Tableau 4.4

Mesures sommaires pour le biais relatif absolu (%) et le CV (%) des estimateurs de l'EQMP associés aux estimateurs EBLUP, YR et WFQ sous le scénario (ii) : Modèle (4.4) mal spécifié

Mesure	EBLUP	YR	WFQ
Biais relatif absolu (BRA) %			
1 ^{er} quartile	4,848	4,850	4,857
Médiane	7,653	7,776	7,466
Moyenne	10,412	10,469	10,221
3 ^e quartile	13,423	13,611	13,776
Coefficient de variation (CV) %			
1 ^{er} quartile	5,510	5,579	4,700
Médiane	7,894	7,317	7,610
Moyenne	10,52	10,55	10,17
3 ^e quartile	13,42	13,03	13,11

Pour ce qui est des mesures du biais relatif absolu (BRA) et du CV des estimateurs de l'EQMP, le tableau 4.3 montre que, sous le scénario (i), le BRA en pourcentage est très faible ($< 2\%$) pour les trois estimateurs de l'EQMP, ce qui confirme que ces estimateurs sont presque sans biais. En outre, le CV en pourcentage est pratiquement le même pour les trois estimateurs de l'EQMP. Sous le scénario (ii), le tableau 4.4 montre que le BRA en pourcentage augmente considérablement, mais est pratiquement le même pour les trois estimateurs de l'EQMP. Par exemple, la valeur pour le troisième quartile passe de 1,674 à 13,423 dans le cas de l'EBLUP, de 1,716 à 13,611 dans le cas de l'estimateur YR et de 1,863 à 13,776 dans le cas de l'estimateur WFQ. De plus, le CV en pourcentage est presque le même pour les trois estimateurs de l'EQMP. La comparaison des valeurs du CV du tableau 4.4 (scénario (ii)) aux valeurs correspondantes du tableau 4.3 (scénario (i)) permet de constater que les CV sont effectivement beaucoup plus faibles sous le scénario (ii) que sous le scénario (i). Cela tient au fait que l'EQMP (utilisée dans le dénominateur du calcul du CV) est nettement plus grande sous le scénario (ii) que sous le scénario (i).

5 Conclusion

Nous avons étudié les propriétés de l'EBLUP et de deux estimateurs avec autocalage, YR et WFQ, sous le modèle au niveau du domaine de Fay-Herriot. Nous avons présenté un estimateur presque sans biais de l'EQMP de l'estimateur YR à la section 3. Notre étude en simulation indique que les trois méthodes donnent des résultats très semblables en ce qui concerne l'EQMP des estimateurs, ainsi que le biais relatif et le CV des estimateurs de l'EQMP. Toutefois, sous une erreur grossière de spécification du modèle due à l'omission d'une variable, l'EQMP augmente considérablement pour les trois estimateurs, sauf si le modèle augmenté de WFQ élimine presque complètement l'erreur de spécification. En pratique, il est difficile de tenir compte de l'erreur de spécification du modèle due à des variables omises inconnues. Il est également intéressant de souligner que l'autocalage pourrait ne pas réduire le biais sous une erreur de spécification du modèle que le modèle augmenté de WFQ ne prend pas en compte.

Remerciements

Nous remercions un examinateur et le rédacteur associé de leurs nombreux commentaires et suggestions constructifs.

Bibliographie

- Fay, R.E., et Herriot, R.A. (1979). Estimation of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 268-277.
- Rao, J.N.K. (2003). *Small Area Estimation*, New York : John Wiley & Sons, Inc.
- Wang, J., Fuller, W.A. et Qu, Y. (2008). Estimation pour petits domaines sous une contrainte. *Techniques d'enquête*, 34, 1, 33-40.
- You, Y., et Rao, J.N.K. (2002). A pseudo empirical best linear unbiased prediction approach to small area estimation using survey weights. *The Canadian Journal of Statistics*, 30, 431-439.