No 11-522-XIF au catalogue

La série des symposiums internationaux de Statistique Canada - Recueil

Symposium 2006 : Enjeux méthodologiques reliés à la mesure de la santé des populations

2006

Statistics Canada Statistique Canada

Canada

Estimateurs robustes de l'erreur quadratique moyenne de prédiction du MPLSBE de petit domaine dans le modèle de Fay-Herriot

Shijie Chen, P. Lahiri, J.N.K. Rao[1]

Résumé

Dans le présent article, nous calculons un estimateur de deuxième ordre sans biais (ou presque sans biais) de l'erreur quadratique moyenne de prédiction (EQMP) du meilleur prédicteur linéaire sans biais empirique (MPLSBE) d'un total de petit domaine pour une extension, selon l'hypothèse de non-normalité, du modèle bien connu de Fay-Herriot. Plus précisément, nous calculons notre estimateur de l'EQMP en posant essentiellement certaines conditions de moment pour les distributions de l'erreur d'échantillonnage et des effets aléatoires. L'estimateur de l'EQMP de Prasad-Rao fondé sur l'hypothèse de normalité se révèle étonnamment robuste en ce sens qu'il reste un estimateur de deuxième ordre sans biais dans des conditions de non-normalité des effets aléatoires lorsqu'un estimateur simple de la méthode des moments est employé pour la composante de variance et lorsque l'erreur d'échantillonnage suit une distribution normale. Nous montrons que l'estimateur de l'EQMP fondé sur l'hypothèse de normalité n'est plus un estimateur de deuxième ordre sans biais lorsque l'erreur d'échantillonnage suit une distribution non normale ou lorsque la méthode des moments de Fay-Herriot est utilisée pour estimer la composante de variance même si l'erreur d'échantillonnage suit une distribution normale. Il est intéressant de noter que lorsque l'estimateur simple de la méthode des moments est utilisé pour la composante de variance, l'estimateur de l'EQMP que nous proposons n'exige pas une estimation du kurtosis des effets aléatoires. Les résultats d'une étude de simulation sur l'exactitude de l'estimateur de l'EQMP proposé, dans des conditions de non-normalité de la distribution tant de l'erreur d'échantillonnage que des effets aléatoires, sont également présentés.

MOTS CLÉS : erreur quadratique moyenne de prédiction; modèle linéaire mixte; composantes de variance.

1. Introduction

Dans le cadre de l'estimation du revenu par habitant pour de petits domaines (population inférieure à 1 000 habitants) à partir des données du Recensement de la population et du logement de 1970, Fay et Herriot (1979) ont fait appel à un modèle de régression pour petits domaines et ont démontré que le meilleur prédicteur linéaire sans biais empirique (MPLSBE) obtenu donnait lieu à une erreur moyenne inférieure à l'erreur associée aux estimateurs d'enquête traditionnels ou à l'erreur associée à la méthode de remplacement fondée sur les moyennes par comté. Posons Y_i, un estimateur direct du i^{e} total de petit domaine θ_i, et $x_i = (x_{i1}, \cdots, x_{ip})'$, un vecteur $p \times 1$ des variables explicatives, $i = 1, \cdots, m.$ Le modèle Fay-Herriot peut être exprimé sous la forme $Y_i = \theta_i + e_i$ et $\theta_i = x_i'\beta + v_i,$ ou sous la forme du modèle linéaire mixte suivant : $Y_i = x_i'\beta + v_i + e_i,$ où l'erreur d'échantillonnage $\{e_i\}$ et les effets aléatoires $\{v_i\}$ sont indépendants, de sorte que $e_i \overset{ind}{\sim} N(\theta_i, D_i)$ et $v_i \overset{ind}{\sim} N(0, \psi), i = 1, \cdots, m.$ On suppose que les variances d'échantillonnage D_i sont connues, mais que β et ψ doivent être estimés à partir des données $(Y_i, x_i), i = 1, \cdots, m.$ En pratique, les D_i sont estimées de façon externe à l'aide de la méthode de la fonction généralisée de la variance (FGV); voir, entre autres, Wolter (1985), Fay et Herriot (1979), Bell et Otto (1995). Le modèle de Fay-Herriot (FH) a été largement utilisé dans les estimations sur petits domaines et pour résoudre des problèmes connexes, notamment en raison de sa simplicité et de sa capacité à assurer la confidentialité des microdonnées et à produire des estimateurs convergents par rapport au plan de sondage; voir l'ouvrage de Rao (2003), chapitre 7.

S'il est facile d'obtenir le meilleur prédicteur linéaire sans biais empirique (MPLSBE) d'un total ou d'une moyenne de petit domaine dans le modèle de FH, l'estimation précise de son erreur quadratique moyenne de prédiction (EQMP) pose un problème de taille. Un estimateur naïf de l'EQMP est donné par l'EQMP du meilleur prédicteur linéaire sans biais (MPLSB), la variance du modèle ψ étant remplacée par un estimateur approprié. Mais cet estimateur naïf sous-estime généralement l'EQMP réelle du MPLSBE pour deux raisons

[1] Shijie Chen, RTI International; P. Lahiri, Université du Maryland, College Park; J.N.K. Rao, Université Carleton

principales. Premièrement, il ne tient pas compte de la variabilité accrue attribuable à l'estimation de ψ, et cette sous-estimation est d'ordre $O(m^{-1})$, pour une grande valeur de m. Deuxièmement, l'estimateur naïf de l'EQMP sous-estime même l'EQMP réelle du MPLSB, cette sous-estimation étant d'ordre $O(m^{-1})$. Prasad et Rao (1990) soulignent l'importance de tenir compte de ces deux sources de sous-estimation. L'utilisation de la méthode de linéarisation de Taylor produit un estimateur de deuxième ordre sans biais (ou presque sans biais) de l'EQMP du MPLSBE lorsque la composante de variance est estimée par une méthode simple des moments. Le biais de cet estimateur de l'EQMP est d'ordre $o(m^{-1})$. Le calcul de l'estimateur de l'EQMP de Prasad et Rao comporte essentiellement deux étapes. Premièrement, on obtient un développement de deuxième ordre de l'EQMP sous une forme correcte en omettant tous les termes d'ordre $o(m^{-1})$. La seconde étape consiste à estimer cette approximation de deuxième ordre de l'EQMP sous une forme correcte de manière à ce que le biais soit d'ordre inférieur, c.-à-d. $o(m^{-1})$. Datta et Lahiri (2000) ont développé l'estimateur de l'EQMP de Prasad et Rao afin de couvrir différentes méthodes d'estimation de ψ. Datta, Rao et Smith (2005) ont obtenu un estimateur de l'EQMP presque sans biais lorsque ψ était estimé par la méthode des moments de Fay-Herriot (1979). Das, Jiang et Rao (2004) ont généralisé la méthode de Taylor en l'appliquant à des modèles linéaires mixtes généraux et ont obtenu des estimateurs de l'EQMP presque sans biais.

Il convient de noter que le calcul du MPLSBE, à l'aide d'un estimateur de moment de ψ, n'exige pas une hypothèse de normalité. Toutefois, pour l'estimation de l'EQMP, Prasad et Rao (1990), Datta et Lahiri (2000), Das, Jiang et Rao (2004), Datta, Rao et Smith (2005) et d'autres se sont appuyés sur l'hypothèse de normalité. Lahiri et Rao (1995) font exception à cet égard et posent l'hypothèse de normalité pour l'erreur d'échantillonnage $\{e_i\}$, mais remplacent l'hypothèse de normalité des effets aléatoires $\{v_i\}$ par certaines conditions de moment. Ces auteurs montrent que l'estimateur de l'EQMP du MPLSBE de Prasad-Rao fondé sur l'hypothèse de normalité reste un estimateur de deuxième ordre sans biais dans ces conditions de non-normalité lorsque l'estimateur simple de la méthode des moments de ψ est employé. Il s'agit là d'un résultat étonnant, qui démontre la robustesse de l'estimateur de l'EQMP de Prasad-Rao sous des conditions non spécifiées de non-normalité des effets aléatoires $\{v_i\}$. Ce résultat tient-il toujours lorsque l'erreur d'échantillonnage $\{e_i\}$ suit une distribution non normale ou lorsque ψ est estimé par l'estimateur de moment de ψ proposé par Fay et Herriot (1979)?

À la section 2, nous examinons brièvement le meilleur prédicteur linéaire sans biais empirique. À la section 3, nous obtenons une approximation de deuxième ordre de l'EQMP du MPLSBE sans hypothèse de normalité. À la section 4, nous proposons un estimateur de deuxième ordre sans biais (ou presque sans biais) de l'EQMP du MPLSBE, à l'aide de l'approximation de l'EQMP obtenue à la section précédente. Enfin, nous présentons à la section 5 certains résultats empiriques d'une brève étude de simulation. Par souci de simplicité, nous ne présentons pas ici la démonstration de nos résultats.

2. MPLSBE

Le modèle de Fay-Herriot, sans l'hypothèse de normalité, peut être exprimé sous la forme $Y_i = \theta_i + e_i$ et $\theta_i = x_i'\beta + v_i, i = 1, \cdots, m,$ ou encore sous la forme d'un modèle linéaire mixte,

$$Y_i = x_i'\beta + v_i + e_i\,, \quad i = 1, \cdots, m\,, \tag{1}$$

où l'erreur d'échantillonnage $\{e_i\}$ et les effets aléatoires $\{v_i\}$ ne sont pas corrélés, de sorte que $e_i \sim [0, D_i, \kappa_{ei}]$ et $v_i \sim [0, \psi, \kappa_v]$, $[\mu, \sigma^2, \kappa]$ représentant une loi de probabilité de moyenne μ, de variance σ^2 et de kurtosis κ. Nous définissons le kurtosis de la distribution par $\kappa = \mu_4/\sigma^4 - 3$, où σ^2 et μ_4 représentent la variance et le quatrième moment central de la distribution, respectivement. Nous supposons que $[\beta, \psi, \kappa_v]$ n'est pas connu, mais que $[D_i, \kappa_{ei}]$ est connu.

Posons $X' = (x_1, \cdots, x_m)$, et $\Sigma(\psi) = diag\{\psi + D_j; j = 1, ..., m\}$, et β peut être estimé par $\hat{\beta}(\psi) = [X'\Sigma^{-1}(\psi)X]^{-1} X'\Sigma^{-1}(\psi)Y$, un estimateur par les moindres carrés pondérés de β pour un ψ

donné. Le MPLSB de θ_i sous le modèle de FH (1) est donné par :

$$\hat{\theta}_i(Y_i;\psi) = B_i Y_i + (1-B_i)x_i'\hat{\beta}(\psi),$$

où $B_i = \frac{\psi}{D_i+\psi}, i = 1,\ldots,m.$ Le MPLSBE de θ_i est ensuite obtenu par

$$\hat{\theta}_i(Y_i;\hat{\psi}) = \hat{B}_i Y_i + (1-\hat{B}_i)x_i'\hat{\beta}(\hat{\psi}) =: \hat{\theta}_i,$$

où $\hat{B}_i = \frac{\hat{\psi}}{D_i+\hat{\psi}}, i = 1,\ldots,m,$ et $\hat{\psi}$ est un estimateur de moment de $\psi.$ Notons que le MPLSB et le MPLSBE, fondés sur $\hat{\psi},$ n'exigent pas une hypothèse de normalité de $\{e_i\}$ et de $\{v_i\}$.

Prasad et Rao (1990) ont proposé l'estimateur simple de la méthode des moments de ψ qui suit :

$$\hat{\psi}_{PR} = max\left\{0, (m-p)^{-1}\sum_{j=1}^{m}\{(Y_u - x_u'\hat{\beta}_{\text{OLS}})^2 - (1-h_{jj})D_j\}\right\},$$

où $\hat{\beta}_{\text{OLS}} = (X'X)^{-1}X'Y,$ l'estimateur par les moindres carrés ordinaires de $\beta,$ et $h_{jj} = x_j'(X'X)^{-1}x_j,$ le levier du j^{e} petit domaine $(j = 1,\cdots,m).$ Fay et Herriot (1979) ont obtenu un estimateur de moment différent, $\hat{\psi}_{FH},$ en résolvant par itération l'équation suivante pour ψ :

$$\frac{1}{m-p}Y'Q(\psi)Y - 1 = 0, \tag{2}$$

où $Y'Q(\psi)Y = \sum_{j=1}^{m}(D_j+\psi)^{-1}\{Y_j - x_j'\hat{\beta}(\psi)\}^2$ est la somme résiduelle pondérée des carrés. Pfeffermann et Nathan (1981) ont proposé une méthode des moments semblable dans le contexte de l'analyse de régression de données d'enquête. Pour le cas particulier $D_i = D\ (i = 1,\cdots,m)$, $\hat{\psi}_{PR} = \hat{\psi}_{FH}.$ Dans le présent article, nous nous attachons au MPLSBE fondé sur $\hat{\psi}_{PR}$ ou $\hat{\psi}_{FH}.$ Les estimateurs $\hat{\psi}_{PR}$ et $\hat{\psi}_{FH}$ sont généralement des estimateurs convergents pour des valeurs élevées de m sous les conditions de régularité suivantes :

- $(r.1)\ 0 < D_L \le D_j \le D_U < \infty,\ j = 1,\cdots,m,$
- $(r.2)\ \sup_{j\ge 1} h_{jj} = O(\frac{1}{m}).$

Sous des conditions de non-normalité et de régularité, le biais de $\hat{\psi}_{PR}$ est d'ordre $o(m^{-1}).$ Cependant, à moins que $D_i = D\ (i = 1,\cdots,m)$, le biais de $\hat{\psi}_{FH}$ est d'ordre $O(m^{-1}),$ même sous des conditions de normalité, et est donné par

$$E[\hat{\psi}_{FH} - \psi] = b(\hat{\psi}_{FH};\psi,\kappa_v) + o(m^{-1}), \tag{3}$$

où

$$\begin{aligned}
b(\hat{\psi}_{FH};\psi,\kappa_v) &= b_N(\hat{\psi}_{FH};\psi) + \alpha(\hat{\psi}_{FH};\psi,\kappa_v)\\
b_N(\hat{\psi}_{FH};\psi) &= \frac{2[m\text{tr}(\Sigma^{-2}) - \{\text{tr}(\Sigma^{-1})\}^2]}{[\text{tr}(\Sigma^{-1})]^3}\\
\alpha(\hat{\psi}_{FH};\psi,\kappa_v) &= \frac{[\text{tr}(\Sigma^{-2})]^2 - \text{tr}(\Sigma^{-3})\text{tr}(\Sigma^{-1})}{[\text{tr}(\Sigma^{-1})]^3}\psi^2\kappa_v + \frac{\text{tr}(D^2\Phi\Sigma^{-2})\text{tr}(\Sigma^{-2}) - \text{tr}(\Sigma^{-1})\text{tr}(D^2\Phi\Sigma^{-3})}{[\text{tr}(\Sigma^{-1})]^3},\\
D &= Diag\{D_j;\, j = 1,\cdots,m\},\\
\Phi &= Diag\{\kappa_{ej};\, j = 1,\cdots,m\}.
\end{aligned}$$

Ci-dessus, $b_N(\hat{\psi}_{FH};\psi)$ représente le biais de $\hat{\psi}_{FH}$ jusqu'à l'ordre $O(m^{-1})$ sous des conditions de normalité (voir Datta, Rao et Smith, 2005), et $\alpha(\hat{\psi}_{FH};\psi,\kappa_v)$ représente l'effet additionnel de la non-normalité sur le biais.

La variance de $\hat{\psi}$ peut être exprimée ainsi

$$var(\hat{\psi}) = var_N(\hat{\psi}) + \eta(\hat{\psi};\psi,\kappa_v), \tag{4}$$

où $var_N(\hat{\psi})$ représente la variance de $\hat{\psi}$ jusqu'à l'ordre $O(m^{-1})$ sous des conditions de normalité, et $\eta(\hat{\psi},\psi,\kappa_v)$ représente l'effet additionnel de la non-normalité sur la variance. À partir des travaux de Datta, Rao et Smith (2005), nous obtenons

$$var_N(\hat{\psi}_{PR}) = 2m^{-2}\sum_{j=1}^{m}(\psi + D_j)^2 = 2m^{-2}\text{tr}(\Sigma^{-2}),$$

$$var_N(\hat{\psi}_{FH}) = 2m\left\{\sum_{j=1}^{m}(\psi + D_j)^{-1}\right\}^{-2} = 2m\left\{\text{tr}(\Sigma^{-1})\right\}^{-2}.$$

On peut montrer que

$$\eta(\hat{\psi}_{PR};\psi,\kappa_v) = m^{-1}\left\{\kappa_v\psi^2 + \frac{1}{m}\sum_{j=1}^{m}\kappa_{ej}D_j^2\right\} = m^{-1}\left\{\kappa_v\psi^2 + m^{-1}\text{tr}(D^2\Phi)\right\}$$

$$\eta(\hat{\psi}_{FH};\psi,\kappa_v) = \left\{\sum_{j}(\psi + D_j)^{-1}\right\}^{-2}\sum_{j=1}^{m}\left\{(\psi + D_j)^{-2}[\kappa_v\psi^2 + \kappa_{ej}D_j^2]\right\}$$
$$= [\text{tr}(\Sigma^{-1})]^{-2}\left\{\text{tr}(\Sigma^{-2})\kappa_v\psi^2 + \text{tr}(D^2\Phi\Sigma^{-2})\right\}.$$

Lorsque $\{e_i\}$ et $\{v_i\}$ sont normaux, nous avons $var(\hat{\psi}_{PR}) = var_N(\hat{\psi}_{PR}), var(\hat{\psi}_{FH}) = var_N(\hat{\psi}_{FH})$ et $var(\hat{\psi}_{FH}) \le var(\hat{\psi}_{PR})$ dans des conditions d'égalité pour le cas équilibré $D_i = D, (i = 1,\cdots,m),$ voir Datta, Rao et Smith (2005). Il est intéressant de noter que le dernier résultat ne s'applique pas aux cas de non-normalité. Pour le cas équilibré $D_i = D\ (i = 1,\cdots,m),$ nous avons $b(\hat{\psi}_{PR};\psi,\kappa_v) = b(\hat{\psi}_{FH};\psi,\kappa_v) = 0$ et $var(\hat{\psi}_{PR}) = var(\hat{\psi}_{FH}) = 2m^{-1}(\psi + D)^2 + m^{-1}\{\kappa_v\psi^2 + m^{-1}D^2\text{tr}(\Phi)\}$ simplement parce que dans cette situation $\hat{\psi}_{PR} = \hat{\psi}_{FH}$.

3. Approximation de l'EQMP

L'EQMP *(MSPE)* du MPLSBE $\hat{\theta}_i$ est donnée par $MSPE(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2,$ où l'espérance est établie en fonction de la distribution marginale de Y sous le modèle de non-normalité de Fay-Herriot (1). La non-normalité n'a pas d'effet sur l'EQMP du MPLSB de $\hat{\theta}_i(Y_i,\psi)$ et cette dernière est donnée par

$$MSPE[\hat{\theta}_i(Y_i,\psi)] = g_{1i}(\psi) + g_{2i}(\psi),$$

où

$$g_{1i}(\psi) = \frac{\psi D_i}{\psi + D_i},$$

$$g_{2i}(\psi) = \frac{D_i^2}{(\psi + D_i)^2}var[\hat{\beta}(\psi)] = \frac{D_i^2}{(\psi + D_i)^2}x_i'\left[X'\Sigma^{-1}(\psi)X]^{-1}\right]^{-1}x_i.$$

Nous cherchons à obtenir une approximation de l'EQMP du MPLSBE dans des conditions de non-normalité qui tienne compte de l'estimation de ψ et qui soit exacte au deuxième ordre, c.-à-d. exacte jusqu'à l'ordre $O(m^{-1})$.

Nous décomposons l'EQMP (MSPE) du MPLSBE $\hat{\theta}_i$ ainsi

$$MSPE[\hat{\theta}_i(Y_i,\hat{\psi})] = MSPE[\hat{\theta}_i(Y_i,\psi)] + E[\hat{\theta}_i(Y_i,\hat{\psi}) - \hat{\theta}_i(Y_i,\psi)]^2$$
$$+2E[\hat{\theta}_i(Y_i,\hat{\psi}) - \hat{\theta}_i(Y_i,\psi)][\hat{\theta}_i(Y_i,\psi) - \theta_i]. \quad (5)$$

où $\hat{\theta}_i(Y_i,\hat{\psi}) = \hat{\theta}_i$ et $\hat{\theta}_i(Y_i,\psi)$ est le MPLSB. Le terme des produits croisés en (5) est nul sous des conditions de normalité de $\{v_i\}$ et $\{e_i\}$; (voir Kacker et Harville, 1984), mais d'ordre $O(m^{-1})$ sous des conditions de non-normalité et donc non négligeable dans le modèle de non-normalité de FH (1). Nous obtenons les approximations suivantes des deux derniers termes de (5).

Résultat 1 : Dans le modèle de non-normalité de FH (1) et sous les conditions de régularité (r.1), (r.2) et (r.3) : $\sup_{j\geq 1} E|v_i|^{8+\delta}$, $0<\delta<1$, nous avons

$$(i) E[\hat{\theta}_i(Y_i,\hat{\psi})-\hat{\theta}_i(Y_i,\psi)]^2 = g_{3i}(\psi,\kappa_v)+o(m^{-1}),$$

$$(ii) E[\hat{\theta}_i(\hat{\theta}_i(Y_i,\hat{\psi})-\hat{\theta}_i(Y_i,\psi)][\hat{\theta}_i(Y_i,\psi)-\theta_i] = g_{4i}(\psi,\kappa_v)+o(m^{-1}),$$

où

$$g_{3i}(\psi,\kappa_v)=\frac{D_i^2}{(\psi+D_i)^3}var(\hat{\psi}),$$

$$g_{4i}(\psi,\kappa_v)=\frac{\psi D_i^2}{m(\psi+D_i)^3}\left[D_i\kappa_{ei}-\psi\kappa_v\right]c(\hat{\psi};\psi),$$

$c(\hat{\psi}_{PR};\psi)=1$ et $c(\hat{\psi}_{FH};\psi)=m(\psi+D_i)^{-1}\left\{\sum_j(\psi+D_j)^{-1}\right\}^{-1}$.

Par conséquent, le développement de deuxième ordre de l'EQMP du MPLSBE $\hat{\theta}_i$ est donné par

$$\begin{aligned} & AEQMP_i \\ &= g_{1i}(\psi)+g_{2i}(\psi)+g_{3i}(\psi,\kappa_v)+2g_{4i}(\psi,\kappa_v) \qquad (6)\\ &=\frac{\psi D_i}{\psi+D_i}+\frac{D_i^2}{(\psi+D_i)^2}var[\hat{\beta}(\psi)]+\frac{D_i^2}{(\psi+D_i)^3}var(\hat{\psi})+\frac{2\psi D_i^2}{m(\psi+D_i)^3}\left[D_i\kappa_{ei}-\psi\kappa_v\right]c(\hat{\psi};\psi)\\ &= AMSPE_{i,N}+\frac{D_i^2}{(\psi+D_i)^3}\eta(\hat{\psi};\psi,\kappa_v)+2g_{4i}(\psi,\kappa_v),. \end{aligned}$$

où $AEQMP_{i,N}$ est l'approximation de l'EQMP dans des conditions de normalité telle qu'elle est donnée par Prasad et Rao (1990) ainsi que par Datta, Rao et Smith (2005). Le terme $g_{3i}(\psi,\kappa_v)$ correspond à l'incertitude accrue attribuable à l'estimation de la composante de variance ψ et le terme $2g_{4i}(\psi,\kappa_v)$ est requis pour tenir compte de la non-normalité. Sous les conditions de régularité, $g_{1i}(\psi)$ est le terme principal [d'ordre $O(1)$] et les autres termes sont tous d'ordre $O(m^{-1})$. Notons que la non-normalité a une incidence tant sur $var(\hat{\psi})$ que sur le terme des produits croisés $2E[\hat{\theta}_i(\hat{\psi},Y)-\hat{\theta}_i(\psi,Y)][\hat{\theta}_i(\psi,Y)-\theta_i]$. Lorsque $\{e_i\}$ aussi bien que $\{v_i\}$ sont normaux, l'approximation ci-dessus se réduit à l'approximation de Prasad-Rao (1990) quand $\hat{\psi}=\hat{\psi}_{PR}$, et à l'approximation de Datta-Rao-Smith (2005) quand $\hat{\psi}=\hat{\psi}_{FH}$. Lorsque $\{e_i\}$ suit une distribution normale et que $\hat{\psi}=\hat{\psi}_{PR}$, l'approximation de l'EQMP (6) se réduit à l'approximation de Lahiri-Rao (1995).

4. Estimateur presque sans biais de l'EQMP

L'approximation de deuxième ordre de l'EQMP, *AEQMP$_i$*, donnée en (6), fait intervenir les paramètres inconnus ψ et κ_v. Supposons que $\hat{\kappa}_v$ est un estimateur convergent de κ_v. Alors $g_{2i}(\hat{\psi})$, $g_{3i}(\hat{\psi},\hat{\kappa}_v)$, et $g_{4i}(\hat{\psi},\hat{\kappa}_v)$ sont des estimateurs de deuxième ordre sans biais (ou presque sans biais) de $g_{2i}(\psi)$, $g_{3i}(\psi,\kappa_v)$, et de $g_{4i}(\psi,\kappa_v)$ respectivement, puisque les dernières fonctions de ψ et κ_v sont déjà d'ordre $O(m^{-1})$. Cependant, l'estimation du terme principal $g_{1i}(\psi)$ en (6) requiert une attention particulière, celle-ci étant d'ordre $O(1)$.

Sous les conditions de régularité (r.1) et (r.2), nous pouvons montrer que

$$E[g_{1i}(\hat{\psi})]=g_{1i}(\psi)-g_{3i}(\psi,\kappa_v)+g_{5i}(\psi,\kappa_v)+o(m^{-1}), \qquad (7)$$

où $g_{5i}(\psi,\kappa_v)=\frac{D_i^2}{(\psi+D_i)^2}b(\hat{\psi};\psi,\kappa_v)$. À partir de (6) et (7), un estimateur de deuxième ordre sans biais (ou presque sans biais) de l'EQMP est donné par

$$mspe_i = g_{1i}(\hat{\psi})+g_{2i}(\hat{\psi})+2g_{3i}(\hat{\psi},\hat{\kappa}_v)+2g_{4i}(\hat{\psi},\hat{\kappa}_v)-g_{5i}(\hat{\psi},\hat{\kappa}_v). \qquad (8)$$

Lorsque $\hat{\psi}=\hat{\psi}_{PR}$, , nous avons

$$\begin{aligned} mspe_i^{PR} &= mspe_{i,N}^{PR}+\frac{2D_i^2}{m(\hat{\psi}+D_i)^3}\left[\hat{\psi}D_i\kappa_{ei}+\frac{1}{m}\sum_{j=1}^{m}\kappa_{ej}D_j^2\right] \\ &= mspe_{i,N}^{PR}+\frac{2D_i^2}{m(\hat{\psi}+D_i)^3}\left[\hat{\psi}D_i\kappa_{ei}+m^{-1}\mathrm{tr}(D^2\Phi)\right], \end{aligned} \qquad (9)$$

où $mspe_{i,N}^{PR}=g_{1i}(\hat{\psi})+g_{2i}(\hat{\psi})+\frac{2D_i^2}{m^2(\hat{\psi}+D_i)^3}\sum_{j=1}^{m}(\hat{\psi}+D_j)^2$ est l'estimateur de l'EQMP fondé sur l'hypothèse de normalité initialement proposé par Prasad et Rao (1990). Il est intéressant de noter que l'estimateur de l'EQMP (9) n'exige pas une estimation de κ_v bien que κ_v intervienne dans l'approximation de l'EQM (6). Pour une distribution normale de $\{e_i\}$ mais une distribution non normale non spécifiée de $\{v_i\}$, $mspe_i^{PR}=mspe_{i,N}^{PR}$ (Lahiri et Rao, 1995). Notons que la robustesse de l'estimateur de l'EQMP de Prasad-Rao ne tient pas dans des conditions de non-normalité de $\{e_i\}$. Lorsque $\kappa_{ej}>0,\ j=1,\cdots,m$, $mspe_i^{PR}$ donne une surestimation quand $mspe_{i,N}^{PR}$ donne une surestimation.

Lorsque $\hat{\psi}=\hat{\psi}_{FH}$, nous obtenons

$$mspe_i^{FH}=mspe_i^{DRS}+\frac{2D_i^2}{(\hat{\psi}_{FH}+D_i)^3}\eta(\hat{\psi}_{FH};\hat{\psi}_{FH},\hat{\kappa}_v)+2g_{4i}(\hat{\psi}_{FH},\hat{\kappa}_v)-\frac{D_i^2}{(\psi+D_i)^2}\alpha(\hat{\psi}_{FH};\hat{\psi}_{FH},\hat{\kappa}_v) \qquad (10)$$

où

$$mspe_i^{DRS}=g_{1i}(\hat{\psi}_{FH})+g_{2i}(\hat{\psi}_{FH})+\frac{2D_i^2}{(\hat{\psi}_{FH}+D_i)^3}evar_N(\hat{\psi}_{FH})-\frac{D_i^2}{(\hat{\psi}_{FH}+D_i)^2}b_N(\hat{\psi}_{FH};\hat{\psi}_{FH})$$

est l'estimateur fondé sur l'hypothèse de normalité proposé par Datta, Rao et Smith (2005). Dans l'équation ci-dessus, $evar_N(\hat{\psi}_{FH})$ est obtenu à partir de $var_N(\hat{\psi}_{FH})$, $\hat{\psi}_{FH}$ étant utilisé au lieu de ψ. Lorsque $\{e_i\}$ et $\{v_i\}$ suivent tous deux une distribution normale, $mspe_i^{FH}$, donné par (10), se réduit à $mspe_i^{DRS}$. Toutefois, lorsque $\{e_i\}$ suit une distribution normale mais non $\{v_i\}$, $mspe_i^{FH}$ ne correspond pas à $mspe_i^{DRS}$ à moins que $D_i=D\ (i=1,\cdots,m)$. Par conséquent, $mspe_i^{DRS}$, contrairement à $mspe_i^{PR}$, n'est pas robuste dans des conditions de non-normalité de $\{v_i\}$, même lorsque $\{e_i\}$ suit une distribution normale. Il est facile de vérifier que pour le cas équilibré $D_i=D\ (i=1,\cdots,m)$, $mspe_i^{FH}=mspe_i^{PR}$.

Nous proposons maintenant deux estimateurs différents de κ_v. Pour justifier le premier estimateur, notons tout d'abord que

$$E[\hat{\psi}_{FH}]^2=Var[\hat{\psi}_{FH}]+[E(\hat{\psi}_{FH})]^2 \qquad (11)$$

Ensuite, à partir des approximations de deuxième ordre de $Var[\hat{\psi}_{FH}]$ et de $E(\hat{\psi}_{FH})$, nous obtenons

$$E[\hat{\psi}_{FH}]^2=\psi^2+k(\psi)+l(\psi)\kappa_v+o(m^{-1}), \qquad (12)$$

où

$$k(\psi) = 2m[\mathrm{tr}(\Sigma^{-1})]^{-2} + 2\psi b_N(\hat{\psi}_{FH};\psi) + \mathrm{tr}(D^2\Phi\Sigma^{-2})\left\{[\mathrm{tr}(\Sigma^{-1})]^{-2} + 2\psi\,\mathrm{tr}(\Sigma^{-2})[\mathrm{tr}(\Sigma^{-1})]^{-3}\right\}$$
$$-2\psi\,\mathrm{tr}(D^2\Phi\Sigma^{-3})[\mathrm{tr}(\Sigma^{-1})]^{-2}$$
$$l(\psi) = \psi^2[\mathrm{tr}(\Sigma^{-1})]^{-2}\left[\mathrm{tr}(\Sigma^{-2}) + 2\psi[\mathrm{tr}(\Sigma^{-1})]^{-1}\{[\mathrm{tr}(\Sigma^{-2})]^2 - \mathrm{tr}(\Sigma^{-1})\mathrm{tr}(\Sigma^{-3})\}\right]$$

Pour obtenir un estimateur de moment de κ_v nous résolvons

$$\hat{\psi}^2_{FH} = \hat{\psi}^2_{FH} + k(\hat{\psi}_{FH}) + l(\hat{\psi}_{FH})\kappa_v \tag{13}$$

pour κ_v. La solution analytique de κ_v est alors donnée par $\hat{\kappa}_v = -k(\hat{\psi}_{FH})/l(\hat{\psi}_{FH})$ si $\hat{\psi}_{FH} > 0$ et par 0 dans les autres cas. Puisqu'il s'agit d'une fonction lisse et que $\hat{\psi}_{FH}$ est convergent par rapport à ψ, nous observons une convergence de l'estimateur $\hat{\kappa}_v$.

Pour obtenir un deuxième estimateur de κ_v, nous remplaçons ψ et $var(\hat{\psi}_{FH})$ dans (4) par $\hat{\psi}_{FH}$ et $v_{WJ} = \sum_{u=1} w_u(\hat{\psi}_{FH,(-u)} - \hat{\psi}_{FH})^2$, un estimateur jackknife pondéré de $var(\hat{\psi}_{FH})$ envisagé par Chen et Lahiri (2006), où $\hat{\psi}_{FH,(-u)}$ est l'estimateur de ψ de Fay-Herriot, calculé à partir de toutes les données sauf celles du u e petit domaine. L'équation obtenue est donnée par :

$$k^{\text{å}}(\hat{\psi}_{FH}) + l^{\text{å}}(\hat{\psi}_{FH})\kappa_v = 0, \tag{14}$$

où

$$k^{\text{å}}(\psi) = 2m + \mathrm{tr}(D^2\Phi\Sigma^{-2}) - \{\mathrm{tr}(\Sigma^{-1}\}^2 v_{WJ}$$
$$l^{\text{å}}(\psi) = \mathrm{tr}(\Sigma^{-2})\psi^2.$$

Après avoir résolu (14) pour κ_v, nous obtenons un autre estimateur analytique de κ_v : $\hat{\kappa}_v^{\text{å}} = -k^{\text{å}}(\hat{\psi}_{FH})/l^{\text{å}}(\hat{\psi}_{FH})$ si $\hat{\psi}_{FH} > 0$ et 0 dans les autres cas. Puisqu'il s'agit d'une fonction lisse et que $\hat{\psi}_{FH}$ et v_{WJ} sont convergents par rapport à ψ et $var(\hat{\psi}_{FH})$ respectivement, nous observons une convergence de l'estimateur $\hat{\kappa}_v^{\text{å}}$.

5. Étude de simulation

Nous examinons dans cette section l'exactitude d'échantillon fini de l'estimateur proposé de l'EQMP, $eqmp_i$, du MPLSBE, dans le cadre d'une simulation de Monte Carlo, pour le cas particulier $x_i'\beta = \mu$ et $D_j = D, \kappa_{ej} = \kappa_e\ (j = 1, \cdots, m).$ L'EQMP étant invariante à la translation (c.-à-d. qu'elle ne change pas lorsque Y_i est remplacé par $Y_i - \mu$), nous posons $\mu = 0$ sans perte de généralité. Nous choisissons aussi les valeurs de paramètre suivantes : $m = 30, 60$ et neuf combinaisons de $\kappa_v = 0, 3, 6$ et $\kappa_e = 0, 3, 6.$ Notons que $\kappa = 0, 3,$ et 6 correspondent respectivement à une distribution normale, une double exponentielle et une exponentielle *décalée* ayant une moyenne de zéro. Nous posons $D = \psi = 1,$ de sorte que $B = 0{,}5$.

Nous générons 10 000 ensembles indépendants d'observations $\{v_i, e_i, i = 1, \cdots, m\}$ pour chacun des cas avec des paramètres spécifiés. Nous calculons ensuite les valeurs simulées de *EQMP*, *E*[estimateur de l'*EQMP*], et *EQM*[estimateur de l'*EQMP*] = *E*[estimateur de l'*EQMP* – *EQMP*]^2 à partir des 10 000 ensembles de données $\{Y_i = v_i + e_i, i = 1, \cdots, m\}$ générés ainsi, puis nous établissons la moyenne pour les petits domaines. Nous comparons trois estimateurs différents de l'EQMP : l'estimateur naïf, l'estimateur de Prasad-Rao et l'estimateur proposé fondé sur $\hat{\psi}_{PR}$. Notons que dans le cas équilibré $\hat{\psi}_{PR} = \hat{\psi}_{FH}$ et il n'est pas nécessaire d'estimer κ_v.

Le tableau 1 présente le biais relatif (BR) en pourcentage de chacun des estimateurs de l'EQMP. Le BR de l'estimateur de l'EQMP est calculé comme suit

BR = [*E*(estimateur de l'*EQMP*) moyen – *EQMP* moyenne] / (*EQMP* moyenne),

où la moyenne est établie sur l'ensemble des petits domaines. Notons que dans le cas équilibré $D_i = D$, théoriquement E[$eqmp_1$] = …E[$eqmp_m$] pour tout estimateur de l'EQMP $eqmp_i$, et l'EQMP de l'estimateur du MPLSBE est la même pour l'ensemble des petits domaines. Par conséquent, l'utilisation du BR présenté ci-dessus est appropriée dans le cas équilibré. Pour chaque estimateur de l'EQMP, le tableau 1 montre que le BR absolu diminue à mesure que le nombre de petits domaines, m , augmente. Dans tous les cas, l'estimateur naïf mène à une sous-estimation. Les résultats de l'estimateur de l'EQMP de Prasad-Rao et de l'estimateur de l'EQMP proposé sont presque identiques, et le BR est négligeable lorsque l'erreur d'échantillonnage $\{e_i\}$ suit une distribution normale, ce qui concorde avec la théorie antérieure (Lahiri et Rao, 1995). Lorsque $\{e_i\}$ suit une distribution non normale, l'estimateur de l'EQMP de Prasad-Rao donne lieu à une sous-estimation, pouvant atteindre dans certains cas 10 % pour $m = 30$. Fait intéressant, la sous-estimation diminue graduellement à mesure que la valeur de m s'approche de 60. Par ailleurs, l'estimateur de l'EQMP proposé corrige la sous-estimation dans tous les cas, mais mène à une surestimation qui diminue considérablement à mesure que m s'approche de 60.
Le tableau 2 présente l'erreur quadratique moyenne relative en pourcentage (EQMR) des estimateurs de l'EQMP. L'EQMR est calculée ainsi

EQMR = [racine carrée moyenne de l'EQM (estimateur de l'*EQMP*)] / (*EQMP* moyenne),

où la moyenne est établie sur l'ensemble des petits domaines. Le tableau montre que l'estimateur de l'EQMP proposé donne les meilleurs résultats, pour ce qui est de l'EQMR, lorsque l'erreur d'échantillonnage suit une distribution non normale, tandis que l'estimateur proposé et celui de Prasad-Rao produisent des résultats presque identiques dans le cas d'une distribution normale. L'estimateur naïf de l'EQMP donne lieu à une EQMR relativement importante en raison de la valeur élevée du carré du biais.

6. Remerciements

Les travaux de S. Chen ont été partiellement financés par les fonds à l'intention des professionnels offerts par RTI International, RTP NC. J.N.K. Rao a bénéficié d'une subvention versée par le Conseil de recherches en sciences naturelles et en génie du Canada. Enfin, les auteurs tiennent à remercier Huilin Li pour l'assistance informatique qu'elle leur a fournie.

Références

Bell, W.R. et Otto, M.C. (1995), "Sampling Error Modelling of Poverty and Income Statistics for States," In *Proceedings of the Section on Survey Research Methods,* Washington, D.C. American Statistical Association.

Chen, S. et Lahiri, P. (2006). On mean squared prediction error estimation in small area estimation problems, unpublished manuscript.

Das, K., Jiang, J. et Rao, J. N. K. (2004). Mean squared error of empirical predictor. *Ann. Statist.* 32, 818-840.

Datta, G.S. et Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10, 613-627.

Datta, G.S., Rao, J.N.K. et Smith, D.D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, 92, 183-196.

Fay, R. E., et Herriot, R. A. (1979). Estimates of Income for Small Places: an Application of James-Stein Procedure to Census Data. *Journal of American Statistical Association* 74, 269-277.

Kackar, R.N., et Harville, D.A. (1984). Approximations for the Standard Errors of Estimation of Fixed and Random Effects in Mixed Linear Models. *Journal of the American Statistical Association* 79, 853-862.

Lahiri, P., et Rao, J.N.K. (1995). Robust estimation of mean squared error of small area estimators. *Journal of the American Statistical Association* 90, 758-766.

Pfeffermann, D. et Nathan, G. (1981). Regression analysis of data from a cluster sample. *Journal of American Statistical Association* 76, 681-689.

Prasad, N.G.N., et Rao, J.N.K. (1990). The Estimation of Mean Squared Error of Small Area Estimators. *Journal of American Statistical Association* 85, 163-171.

Rao, J.N.K. (2003). *Small Area Estimation*. New York: Wiley.

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

Tableau 1. *Valeurs simulées du biais relatif (BR) en pourcentage des estimateurs de l'erreur quadratique moyenne pour* $\psi = D = 1$. *La composante de variance est estimée par la méthode de l'estimation des moments de Prasad-Rao.*

		Distribution de v					
		Normale		Double exponentielle		Exponentielle	
Distribution de e		m=30	m=60	m=30	m=60	m=30	m=60
	Naïf	-12,1	-6,64	-6,66	-2,92	-12,29	-6,34
Normale	Prasad-Rao	0,86	-0,11	0,53	0,54	1,61	0,54
	Proposé	0,86	-0,11	0,53	0,54	1,61	0,54
	Naïf	-17,9	-10,67	-9,6	-4,96	-17,59	-10,82
Double exponentielle	Prasad-Rao	-5,35	-4,22	-2,59	-1,52	-4,21	-4,16
	Proposé	5,76	1,1	1,82	0,42	8,27	1,53
	Naïf	-22,13	-14,17	-11,91	-6,67	-22,65	-14,18
Exponentielle	Prasad-Rao	-9,86	-7,9	-4,85	-3,25	-9,69	-7,6
	Proposé	12,4	2,61	4,33	0,66	15,05	3,86

Tableau 2. *Valeurs simulées de l'erreur quadratique moyenne relative (EQMR) en pourcentage des estimateurs de l'erreur quadratique moyenne pour* $\psi = D = 1$. *La composante de variance est estimée par la méthode de l'estimation des moments de Prasad-Rao.*

		Distribution de v					
		Normale		Double exponentielle		Exponentielle	
Distribution de e		m=30	m=60	m=30	m=60	m=30	m=60
	Naïf	4,29	2,03	2,48	1,05	5,99	3,2
Normale	Prasad-Rao	2,6	1,57	1,62	0,87	3,87	2,61
	Proposé	2,6	1,57	1,62	0,87	3,87	2,61
	Naïf	5,95	2,95	3,04	1,34	7,59	4,2
Double exponentielle	Prasad-Rao	3,22	2,13	1,83	1,04	4,47	3,21
	Proposé	1,77	1,57	1,14	0,86	2,61	2,39
	Naïf	7,58	3,88	3,72	1,61	9,3	5,02
Exponentielle	Prasad-Rao	4,04	2,75	2,18	1,21	5,25	3,73
	Proposé	1,61	1,41	0,8	0,8	2,24	1,99