

## Techniques d'enquête

# Comparaison de certains estimateurs de variance positifs pour le modèle d'estimation sur petits domaines Fay-Herriot

par Susana Rubin-Bleuer et Yong You

Date de diffusion : le 22 juin 2016



Statistique  
Canada

Statistics  
Canada

Canada

---

## 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](http://www.statcan.gc.ca).

Vous pouvez également communiquer avec nous par :

**Courriel** à [STATCAN.infostats-infostats.STATCAN@canada.ca](mailto:STATCAN.infostats-infostats.STATCAN@canada.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

## 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](http://www.statcan.gc.ca) sous « Contactez-nous » > « Normes de service à la clientèle ».

## Note de reconnaissance

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

## Signes conventionnels dans les tableaux

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<sup>s</sup> valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- <sup>p</sup> provisoire
- <sup>r</sup> révisé
- x confidentiel en vertu des dispositions de la *Loi sur la statistique*
- <sup>E</sup> à utiliser avec prudence
- F trop peu fiable pour être publié
- \* valeur significativement différente de l'estimation pour la catégorie de référence ( $p < 0,05$ )

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2016

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

*This publication is also available in English.*

---

# Comparaison de certains estimateurs de variance positifs pour le modèle d'estimation sur petits domaines Fay-Herriot

Susana Rubin-Bleuer et Yong You<sup>1</sup>

## Résumé

La méthode du maximum de vraisemblance restreint (méthode REML pour *restricted maximum likelihood*) est généralement utilisée pour estimer la variance de l'effet aléatoire de domaine sous le modèle de Fay-Herriot (Fay et Herriot 1979) afin d'obtenir le meilleur estimateur linéaire sans biais empirique (estimateur EBLUP pour *empirical best linear unbiased predictor*) d'une moyenne de petit domaine. Lorsque l'estimation REML correspond à zéro, le poids de l'estimateur d'échantillon direct est zéro et l'EBLUP devient un estimateur synthétique, ce qui est rarement souhaitable. Pour résoudre le problème, Li et Lahiri (2011) et Yoshimori et Lahiri (2014) ont élaboré des estimateurs de variance constante par la méthode du maximum de vraisemblance ajusté (méthode ADM pour *adjusted maximum likelihood*), qui produisent toujours des estimations de variance positives. Certains des estimateurs ADM produisent toujours des estimations positives, mais génèrent un biais élevé, ce qui influe sur l'estimation de l'erreur quadratique moyenne (EQM) de l'estimateur EBLUP. Nous proposons d'utiliser un estimateur de variance MIX, défini comme étant une combinaison des méthodes REML et ADM. Nous montrons que cet estimateur est sans biais jusqu'à l'ordre deux et qu'il produit toujours une estimation de variance positive. Nous proposons également un estimateur de l'EQM sous la méthode MIX et montrons au moyen d'une simulation fondée sur un modèle que, dans de nombreuses situations, cet estimateur donne de meilleurs résultats que d'autres estimateurs de l'EQM par « linéarisation de Taylor » récemment proposés.

**Mots-clés :** Estimation de variance; maximum de vraisemblance ajusté; REML; ordre du biais; estimation de l'EQM.

## 1 Introduction

Le modèle de Fay-Herriot (Fay et Herriot 1979) est un modèle de base au niveau du domaine utilisé pour estimer les moyennes de petit domaine lorsque les estimations d'enquête directes disponibles sont imprécises en raison de la petite taille des échantillons. Dans ce modèle, la moyenne de petit domaine est représentée par un terme linéaire non aléatoire dans les covariables, avec un effet aléatoire de domaine. Dans le modèle de Fay-Herriot, on peut obtenir le meilleur estimateur linéaire sans biais (estimateur BLUP pour *best linear unbiased predictor*) d'une moyenne de petit domaine en minimisant l'erreur quadratique moyenne (EQM) dans la classe des estimateurs linéaires sans biais. L'estimateur BLUP correspond à une moyenne pondérée de l'estimateur d'enquête direct et de l'estimateur synthétique de type régression, les poids dépendant de la variance des effets aléatoires de domaine,  $\sigma_v^2$ . En règle générale, cette variance doit être estimée à partir des données sous le modèle de Fay-Herriot. On obtient le meilleur estimateur linéaire sans biais empirique (EBLUP) de la moyenne de petit domaine en remplaçant la variance dans la formule de l'estimateur BLUP par une estimation. De nombreuses méthodes bien connues d'estimation de variance sont utilisées dans ce contexte, mais la plus courante est la méthode du maximum de vraisemblance restreint (méthode REML pour *restricted maximum likelihood*), car elle tient compte de la perte des degrés de liberté attribuable à l'estimation du coefficient de régression. De plus, cette méthode est sans biais jusqu'à l'ordre deux et exige moins d'itérations, car elle converge plus rapidement. Malgré ces caractéristiques

1. Susana Rubin-Bleuer et Yong You, Division de la coopération internationale et des méthodes statistiques institutionnelles, Statistique Canada.  
Courriel : susana.rubin-bleuer@canada.ca; yong.you@canada.ca.

importantes, il arrive parfois, particulièrement lorsque le nombre de domaines,  $m$ , est petit ou modéré, que la méthode REML produise une estimation de variance nulle. Cela donnerait un poids nul à l'estimateur d'enquête direct dans la formule EBLUP, et l'estimateur EBLUP devient donc un estimateur synthétique de type régression. La plupart des praticiens hésitent toutefois à utiliser les estimateurs synthétiques pour les moyennes de petit domaine, car ceux-ci ne tiennent pas compte des données d'enquête et sont souvent très biaisés. Lorsqu'on a affaire à des ensembles de données réels, pour lesquels les modèles ne sont jamais parfaits, une estimation positive pour  $\sigma_v^2$  réduit le biais de l'estimateur EBLUP par rapport au modèle synthétique. Certes, une estimation positive de la variance des effets aléatoires produit un estimateur EBLUP « prudent », en ce sens qu'elle attribue un poids positif à l'estimateur d'enquête direct. Elle peut également être considérée comme la somme de l'estimateur de régression et du terme non nul qui tient compte d'une partie du « biais de modèle ». Cette caractéristique donne lieu à une série de méthodes d'estimation de variance qui donnent des estimations positives.

Dans cet article, nous mettons l'accent sur les estimateurs de variance par maximum de vraisemblable ajusté mis au point par Lahiri et Li (2009), et nous proposons un estimateur de variance MIX. Notre estimateur de variance MIX combine un estimateur REML et une des méthodes d'estimation par maximum de vraisemblance ajusté. Nous proposons également un estimateur de l'EQM de l'EBLUP sous la méthode MIX et examinons les propriétés théoriques et en échantillon fini de l'estimateur de variance MIX et de l'estimateur de l'EQM.

Morris (2006) et Lahiri et Li (2009) ont proposé des estimateurs de variance par maximum de vraisemblable ajusté découlant de l'optimisation de la vraisemblance profilée et de la vraisemblance résiduelle ajustée par un facteur  $h(\sigma_v^2)$ ,  $\sigma_v^2 > 0$ . Li et Lahiri (2011) ont proposé deux méthodes d'estimation de variance (méthodes AM.LL et AR.LL, associées respectivement à la vraisemblance profilée et à la vraisemblance résiduelle) qui garantissent des estimations positives avec un facteur d'ajustement  $h_{LL}(\sigma_v^2) = \sigma_v^2$ . Yoshimori et Lahiri (2014) ont proposé deux autres méthodes d'estimation de variance (méthodes AM.YL et AR.YL) par ajustement de la vraisemblance profilée et de la vraisemblance résiduelle avec le facteur

$$h_{YL}(\sigma_v^2) = \left\{ \arctan \left[ \frac{\sum_{i=1}^m \sigma_v^2}{(\sigma_v^2 + \psi_i)} \right] \right\}^{1/m}$$

où  $\psi_i$  est la variance d'échantillonnage pour le  $i^{\circ}$  domaine. Il est bien connu que les estimateurs LL sont biaisés, particulièrement lorsque le nombre de domaines est faible ou modéré (Lahiri et Pramanik 2011). La méthode YL, qui ajuste la vraisemblance profilée, produit aussi un estimateur biaisé de  $\sigma_v^2$ . Le biais de l'estimateur de variance n'affecte toutefois pas l'EQM de l'EBLUP. En effet, l'approximation asymptotique d'ordre deux de l'EQM montre que l'EQM dépend de la variance asymptotique et non du biais de l'estimateur de variance. Cependant, le biais des estimateurs de variance affecte les estimateurs de l'EQM par linéarisation de Taylor et peut produire des estimateurs de l'EQM présentant un biais négatif. Il est alors souhaitable d'examiner d'autres estimateurs de variance positifs.

Yuan (2009) a été le premier à mentionner la méthode consistant à combiner les estimateurs de variance AM.LL et REML pour le modèle Fay-Herriot, mais il n'a pas étudié ses propriétés, empiriques ou autres.

Rubin-Bleuer, Yung et Landry (2010, 2011 et 2012) ont effectué des comparaisons empiriques d'un estimateur de variance MIX dans un modèle chronologique et transversal au niveau du domaine, tandis que Rubin-Bleuer et You (2012) ont étudié les propriétés asymptotiques et en échantillon fini de l'estimateur de variance MIX pour le modèle de Fay-Herriot.

Dans cet article, nous formalisons la méthode MIX pour le modèle de Fay-Herriot et prouvons que l'estimateur de variance MIX est sans biais jusqu'à l'ordre deux. En outre, nous proposons un estimateur de l'EQM par linéarisation de Taylor. Nous examinons également les résultats empiriques de l'estimateur MIX pour un nombre faible ou modéré de domaines. En ce qui concerne l'estimation de l'EQM, Rubin-Bleuer et You (2012) et Molina, Rao et Datta (2015) ont tous proposé des estimateurs d'EQM « fractionnés » différents sous l'estimation de variance MIX. Nous montrons que les estimateurs de l'EQM de Rubin-Bleuer et You (2012) et de Molina et coll. (2015) sont sans biais jusqu'à l'ordre deux. Ces estimateurs d'EQM « fractionnés » ont été assujettis à une règle pour les populations qui a donné des estimations nulles sous l'estimation de variance REML et à une autre règle pour les populations qui a donné des estimations positives sous l'estimation de variance REML. Les deux articles susmentionnés ont montré que, pour un petit nombre de domaines, ces estimateurs « fractionnés » ont donné de bons résultats empiriques en termes de biais relatif moyen. Or, ce résultat pourrait être trompeur, car les estimateurs de l'EQM présentent généralement un biais négatif pour les populations où l'estimation de variance REML est nulle, et un biais positif pour les populations où l'estimation REML est positive, le biais s'annulant en moyenne. Étant donné ce qui précède, nous proposons un nouvel estimateur de l'EQM, et nous le comparons à d'autres pour les populations où l'estimation REML est nulle.

Dans la section 2, nous présentons le modèle de Fay-Herriot, l'estimateur EBLUP de la moyenne de petit domaine et une approximation d'ordre deux de l'EQM de l'EBLUP sous le modèle. Dans la section 3, nous décrivons l'estimateur REML et les estimateurs de variance \*.LL et \*.YL. Dans la section 4, nous présentons un estimateur de variance MIX général et prouvons que son biais est du même ordre que celui de l'estimateur REML. Nous proposons un estimateur sans biais (jusqu'à l'ordre deux) de l'EQM sous la méthode MIX. Dans la section 5, nous menons une étude empirique afin de comparer les différents estimateurs de variance. Il est à noter que nous avons défini l'estimateur de variance MIX comme étant une combinaison de l'estimateur REML et d'un des estimateurs de variance par maximum de vraisemblable ajusté, mais l'estimateur de variance MIX que nous avons choisi pour cette étude combine l'estimateur REML et l'estimateur de variance AM.LL. Nous avons sélectionné cette combinaison, parce que Li et Lahiri (2011) ont déclaré que la méthode de la vraisemblance profilée ajustée avait donné de meilleurs résultats que celle de la vraisemblance résiduelle ajustée (AR.LL) et que le facteur d'ajustement des estimateurs de variance de Yoshimori et Lahiri (2014) était trop proche de zéro (en termes logarithmiques) pour améliorer de façon significative la méthode REML. Enfin, dans la section 6, nous présentons et analysons les résultats de la simulation et tirons des conclusions.

## 2 EBLUP et EQM de l'EBLUP sous le modèle de Fay-Herriot

Soit  $y_i, i = 1, \dots, m$ , les estimateurs d'enquête directs des moyennes de petit domaine  $\theta_i, i = 1, \dots, m$ . Le modèle de Fay-Herriot se compose des modèles d'échantillonnage et de lien suivants :

$$\text{Modèle d'échantillonnage : } y_i = \theta_i + e_i, e_i | \theta_i \stackrel{\text{i.d.}}{\sim} (0, \psi_i), \quad i = 1, \dots, m, \quad (2.1)$$

$$\text{Modèle de lien : } \theta_i = \mathbf{z}'_i \boldsymbol{\beta} + v_i, \quad v_i \stackrel{\text{i.i.d.}}{\sim} (0, \sigma_v^2), \quad \sigma_v^2 > 0, \quad i = 1, \dots, m, \quad (2.2)$$

où les  $e_i$  sont les erreurs d'échantillonnage, indépendamment distribuées de moyenne de zéro et de variances d'échantillonnage « connues »  $\psi_i$ ,  $\mathbf{z}_i$  ( $p \times 1$ ) sont des vecteurs connus de valeurs de covariables;  $\boldsymbol{\beta}$  est un vecteur  $p \times 1$  de coefficients de régression fixes inconnus; et  $v_i$  sont des effets aléatoires indépendants et identiquement distribués avec une moyenne de zéro et une variance de modèle  $\sigma_v^2$ . La combinaison de (2.1) et (2.2) donne :

$$y_i = \mathbf{z}'_i \boldsymbol{\beta} + v_i + e_i, \quad i = 1, \dots, m, \quad (2.3)$$

avec des erreurs de modèle et d'échantillonnage. Les  $y_i, i = 1, \dots, m$ , peuvent être considérés comme des résultats dans l'espace conjoint plan de sondage-modèle (Rubin-Bleuer et Schioppa-Kratina 2005).

Dans le modèle (2.3), l'estimateur EBLUP de la moyenne de petit domaine  $\theta_i$  est donné par :

$$\hat{\theta}_i(\hat{\sigma}_v^2) = \mathbf{z}'_i \hat{\boldsymbol{\beta}}(\hat{\sigma}_v^2) + \hat{\gamma}_i [y_i - \mathbf{z}'_i \hat{\boldsymbol{\beta}}(\hat{\sigma}_v^2)] = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i) \mathbf{z}'_i \hat{\boldsymbol{\beta}}(\hat{\sigma}_v^2), \quad i = 1, \dots, m, \quad (2.4)$$

où  $\hat{\sigma}_v^2$  est un estimateur convergent de  $\sigma_v^2$ ,

$$\hat{\gamma}_i = \hat{\sigma}_v^2 / (\hat{\sigma}_v^2 + \psi_i), \quad \text{et} \quad \hat{\boldsymbol{\beta}}(\hat{\sigma}_v^2) = \left[ \sum_{i=1}^m \mathbf{z}_i \mathbf{z}'_i / (\hat{\sigma}_v^2 + \psi_i) \right]^{-1} \left[ \sum_{i=1}^m \mathbf{z}_i y_i / (\hat{\sigma}_v^2 + \psi_i) \right]. \quad (2.5)$$

Pour calculer l'erreur quadratique moyenne (EQM) de l'EBLUP, nous posons les conditions de régularité suivantes :

- 1) Les  $\psi_i$  ont une borne supérieure et sont loin de zéro;
- 2) Les  $\mathbf{z}_i, 1 \leq i \leq m$  sont bornés;
- 3)  $\liminf \lambda_{\min} (1/m \sum_i \mathbf{z}_i \cdot \mathbf{z}'_i) > 0$  où  $\lambda_{\min}(A)$  = valeur propre minimum de la matrice  $A$ .

Sous la normalité des erreurs d'échantillonnage  $e_i$  associées au modèle (2.3) et les conditions de régularité ci-dessus, une approximation d'ordre deux de l'EQM est donnée par :

$$\text{EQM} \{ \hat{\theta}_i(\hat{\sigma}_v^2) \} = g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + g_{3i}(\sigma_v^2) + o\left(\frac{1}{m}\right), \quad (2.6)$$

avec  $g_{1i}(\sigma_v^2) = \gamma_i \psi_i$ ,  $g_{2i}(\sigma_v^2) = (1 - \gamma_i)^2 \mathbf{z}'_i \left[ \sum_{i=1}^m \mathbf{z}_i \mathbf{z}'_i / (\sigma_v^2 + \psi_i) \right]^{-1} \mathbf{z}_i$  et

$$g_{3i}(\sigma_v^2) = (\psi_i)^2 \bar{V}(\hat{\sigma}_v^2) / (\sigma_v^2 + \psi_i)^3, \quad (2.7)$$

où  $\bar{V}(\hat{\sigma}_v^2)$  est la variance asymptotique de  $\hat{\sigma}_v^2$  (Das, Jiang et Rao 2004).

### 3 Examen des méthodes REML et du maximum de vraisemblance ajusté

#### 3.1 Méthode REML

Nous examinons le modèle combiné de Fay-Herriot (2.3) où  $\sigma_v^2 > 0$ . On obtient l'estimateur de variance REML de  $\sigma_v^2$  en maximisant la fonction de vraisemblance résiduelle pour  $\sigma_v^2$  :

$$L_{\text{REML}}(\sigma_v^2) \propto \left[ \sum_{i=1}^m \mathbf{z}_i \mathbf{z}_i' / (\sigma_v^2 + \psi_i) \right]^{-1/2} \prod_{i=1}^m (\sigma_v^2 + \psi_i)^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y} \right\}$$

où  $\mathbf{y} = (y_1, \dots, y_m)'$ ,  $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{Z} (\mathbf{Z}' \mathbf{V}^{-1} \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{V}^{-1}$ ,  $\mathbf{V} = \text{Var}(\mathbf{y})$ , et  $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_m)'$ . (Cressie 1992; Datta et Lahiri 2000; Rao 2003, chapitre 6). L'estimateur de variance REML est donné par :

$$\hat{\sigma}_{\text{vREML}}^2 = \max(\tilde{\sigma}_{\text{vREML}}^2, 0), \quad (3.1)$$

où  $\tilde{\sigma}_{\text{vREML}}^2$  est la valeur convergente de l'algorithme REML. Le biais asymptotique et la variance de l'estimateur REML jusqu'à l'ordre deux sont donnés respectivement par :

$$\text{Biais}(\hat{\sigma}_{\text{vREML}}^2) = o\left(\frac{1}{m}\right) \text{ et } V(\hat{\sigma}_{\text{vREML}}^2) = \frac{2}{\text{tr}(\mathbf{V}^{-2})} + o\left(\frac{1}{m}\right). \quad (3.2)$$

Un estimateur sans biais d'ordre deux de l'EQM de l'EBLUP sous l'estimation de variance REML est donné par (Datta et Lahiri 2000; Chen et Lahiri 2008, 2011) :

$$\text{eqm}\{\hat{\theta}_i(\hat{\sigma}_{\text{vREML}}^2)\} = \begin{cases} g_{1i}(\hat{\sigma}_{\text{vREML}}^2) + g_{2i}(\hat{\sigma}_{\text{vREML}}^2) + 2g_{3i}(\hat{\sigma}_{\text{vREML}}^2) & \text{si } \hat{\sigma}_{\text{vREML}}^2 > 0 \\ g_{2i}(0) & \text{si } \hat{\sigma}_{\text{vREML}}^2 = 0. \end{cases} \quad (3.3)$$

**Remarque 3.1.** Quand  $\hat{\sigma}_v^2 = 0$ , l'EBLUP est réduit à l'estimateur synthétique. Cependant, lorsque

$$\hat{\sigma}_v^2 = 0, g_{1i}(\hat{\sigma}_v^2) = 0, g_{2i}(\hat{\sigma}_v^2) = \mathbf{z}_i' \left[ \sum_{i=1}^m \mathbf{z}_i \mathbf{z}_i' / \psi_i \right]^{-1} \mathbf{z}_i,$$

et  $g_{3i}(\hat{\sigma}_v^2) = \bar{V}(\hat{\sigma}_v^2) / \psi_i > 0$ , c'est-à-dire que  $\text{eqm}\{\hat{\theta}_i(\hat{\sigma}_v^2)\}$  n'est pas une fonction continue de  $\hat{\sigma}_v^2$ . Nous verrons dans l'étude empirique que, lorsque nous procédons au conditionnement sur  $\{\hat{\sigma}_v^2 = 0\}$ , l'estimateur de l'EQM en (3.3) présente un biais négatif significatif, à moins que le rapport signal/bruit sous-jacent  $\sigma_v^2 / \psi_i$  ne soit négligeable.

#### 3.2 Méthodes du maximum de vraisemblance ajusté

On obtient les estimateurs de variance par maximum de vraisemblance ajusté en optimisant soit la vraisemblance profilée ajustée (AM pour *adjusted maximum*), soit la vraisemblance résiduelle ajustée (AR pour *adjusted residual*) avec le facteur  $h(\sigma_v^2)$ . Comme il est mentionné dans l'introduction, les estimateurs

AM.LL et AR.LL utilisent le facteur d'ajustement  $h_{LL}(\sigma_v^2) = \sigma_v^2$ , tandis que les estimateurs AM.YL et AR.YL utilisent le facteur d'ajustement

$$h_{YL}(\sigma_v^2) = \left\{ \arctan \left[ \sum_{i=1}^m \sigma_v^2 / (\sigma_v^2 + \psi_i) \right] \right\}^{1/m}.$$

Nous désignons par  $\hat{\sigma}_{vAM.LL}^2$  et  $\hat{\sigma}_{vAM.YL}^2$  les estimateurs de variance obtenus par maximisation des fonctions de vraisemblance profilée ajustée, pour  $\sigma_v^2$  :

$$L_{AM.*}(\sigma_v^2) \propto h(\sigma_v^2) \cdot \prod_{i=1}^m (\sigma_v^2 + \psi_i)^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y} \right\}, \quad (3.4)$$

où  $h(\sigma_v^2) = h_{LL}(\sigma_v^2)$  et  $h(\sigma_v^2) = h_{YL}(\sigma_v^2)$  pour AM.LL et AM.YL respectivement. La matrice  $\mathbf{P}$  est comme en (3.1). Le biais des estimateurs AM jusqu'à l'ordre deux (désigné par  $\approx$ ) correspond à :

$$B(\hat{\sigma}_{vAM.LL}^2) \approx \frac{\text{tr}\{\mathbf{P} - \mathbf{V}^{-1}\} + 2/\sigma_v^2}{\text{tr}(\mathbf{V}^{-2})} = O\left(\frac{1}{m}\right) \quad \text{et} \quad B(\hat{\sigma}_{vAM.YL}^2) \approx \frac{\text{tr}\{\mathbf{P} - \mathbf{V}^{-1}\}}{\text{tr}(\mathbf{V}^{-2})} = O\left(\frac{1}{m}\right), \quad (3.5)$$

(Li et Lahiri 2011; Yoshimori et Lahiri 2014). On obtient les estimateurs de variance AR.LL et AR.YL, désignés par  $\hat{\sigma}_{vAR.LL}^2$  et  $\hat{\sigma}_{vAR.YL}^2$ , en maximisant les fonctions de vraisemblance résiduelle ajustée (AR) pour  $\sigma_v^2$  :

$$L_{AR.*}(\sigma_v^2) \propto h(\sigma_v^2) \cdot \left| \sum_{i=1}^m \mathbf{z}_i \mathbf{z}_i' / (\sigma_v^2 + \psi_i) \right|^{-1/2} \prod_{i=1}^m (\sigma_v^2 + \psi_i)^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{y}' \mathbf{P} \mathbf{y} \right\} \quad (3.6)$$

où  $h(\sigma_v^2) = h_{LL}(\sigma_v^2)$  et  $h(\sigma_v^2) = h_{YL}(\sigma_v^2)$  pour AR.LL et AR.YL respectivement, et  $\mathbf{P}$  est comme en (3.1). Les biais asymptotiques des estimateurs AR sont donnés respectivement par :

$$B(\hat{\sigma}_{vAR.LL}^2) \approx \frac{2/\sigma_v^2}{\text{tr}(\mathbf{V}^{-2})} = O\left(\frac{1}{m}\right) \quad \text{et} \quad B(\hat{\sigma}_{vAR.YL}^2) = o\left(\frac{1}{m}\right). \quad (3.7)$$

Sous les conditions de régularité données dans la section 2 et sous  $\sigma_v^2 > 0$ , les deux LL et les deux estimateurs de variance YL existent et sont  $\sqrt{m}$ -convergents (Li et Lahiri 2011; Yoshimori et Lahiri 2014). Lahiri et ses coauteurs ont proposé les estimateurs de l'EQM suivants :

$$\text{eqm}\{\hat{\theta}_i(\cdot)\} = g_{1i}(\cdot) + g_{2i}(\cdot) + 2g_{3i}(\cdot) - \psi_i^2 \cdot B(\cdot) / (\cdot + \psi_i)^2 \quad (3.8)$$

où l'argument en  $(\cdot)$  ci-dessus est soit  $\hat{\sigma}_{vAM.LL}^2$ ,  $\hat{\sigma}_{vAR.LL}^2$  ou  $\hat{\sigma}_{vAM.YL}^2$  sous les estimateurs de variance AM.LL, AR.LL et AM.YL respectivement et sous  $\hat{\sigma}_{vAR.YL}^2$  :

$$\text{eqm}\{\hat{\theta}_i(\hat{\sigma}_{vAR.YL}^2)\} = g_{1i}(\hat{\sigma}_{vAR.YL}^2) + g_{2i}(\hat{\sigma}_{vAR.YL}^2) + 2g_{3i}(\hat{\sigma}_{vAR.YL}^2). \quad (3.9)$$

Les estimateurs (3.8) et (3.9) sont sans biais jusqu'à l'ordre deux.



**Remarque 3.2.** Il n'est pas nécessaire que les erreurs d'échantillonnage aient une distribution normale pour assurer la convergence et la normalité asymptotique des estimateurs LL et YL (voir, par exemple, Rubin-Bleuer et coll. 2011).

### 3.3 Algorithmes d'optimisation

Vu les données, la fonction de vraisemblance REML peut atteindre sa valeur maximale à  $\sigma_v^2 = 0$ , même lorsque la valeur sous-jacente réelle de  $\sigma_v^2$  est positive. Par ailleurs, les vraisemblances LL et YL atteignent toujours leur valeur maximale à  $\sigma_v^2 > 0$ . Pourtant, la vraisemblance résiduelle YL est très proche de la vraisemblance REML. Des études empiriques montrent que l'algorithme de score sous AR.YL donne  $\hat{\sigma}_{\text{VAR.YL}}^2 = 0$  dans une proportion presque aussi importante que sous REML pour les ensembles de données suivant un modèle de Fay-Herriot avec une variance sous-jacente réelle faible mais non nulle. Cela se produit lorsque l'algorithme de score passe à côté de la valeur maximale positive de la vraisemblance AR.YL et produit une valeur nulle (pour plus de détails, voir l'annexe B). Pour éviter ce problème, nous utilisons une méthode de grille pour l'optimisation (Estevao 2014). Dans notre étude, nous établissons la limite supérieure de l'intervalle de recherche à  $1000 \times \sigma_v^2$ , car nous connaissons  $\sigma_v^2$  a priori. Pour les applications avec des données réelles, nous suggérons d'obtenir une estimation initiale  $\hat{\sigma}_{\text{vAM.LL}}^2$  en utilisant la méthode de score et de fixer la limite supérieure à  $1000 \times \hat{\sigma}_{\text{vAM.LL}}^2$ , puis d'augmenter graduellement la limite jusqu'à ce que l'estimation de variance se situe dans l'intervalle de recherche.

## 4 Estimateur de variance MIX

### 4.1 Estimation de variance

L'estimateur de variance MIX est une procédure qui commence par calculer l'estimation de variance REML et qui la remplace par l'estimation de variance par le maximum de vraisemblance ajusté seulement lorsque l'estimation REML est négative. L'estimateur de variance MIX est toujours positif et sans biais jusqu'à un terme d'ordre  $o(1/m)$ . L'estimateur de variance MIX de  $\sigma_v^2$  est défini par :

$$\hat{\sigma}_{\text{vMIX}}^2 = \begin{cases} \hat{\sigma}_{\text{vREML}}^2 & \text{si } \hat{\sigma}_{\text{vREML}}^2 > 0 \\ \hat{\sigma}_{\text{vadj}}^2 & \text{si } \hat{\sigma}_{\text{vREML}}^2 = 0, \end{cases} \quad (4.1)$$

où  $\hat{\sigma}_{\text{vadj}}^2$  est un des estimateurs de la vraisemblance ajustée définis dans la section 3.

**Remarque 4.1.** L'estimateur de variance MIX présente automatiquement certaines des propriétés communes partagées par l'estimateur REML et l'estimateur de variance par le maximum de vraisemblance ajusté. Par exemple, il est pair et invariant de translation. Ainsi, sous l'hypothèse de normalité des erreurs d'échantillonnage, l'approximation d'ordre deux (2.6) de l'EQM de l'EBLUP est également valide : le théorème 4.1 ci-après montre que l'EQM de l'EBLUP sous l'estimateur de variance MIX hérite des mêmes propriétés asymptotiques que l'EQM sous l'estimateur de variance REML.

**Théorème 4.1.** Sous les conditions de régularité 1 à 3 données dans la section 2 et l'hypothèse que  $\sigma_v^2 > 0$ , l'EQM de l'EBLUP sous l'estimateur de variance MIX est égale à l'EQM sous l'estimateur de variance REML jusqu'à l'ordre deux. Le théorème découle du fait que la variance asymptotique de  $\hat{\sigma}_{vMIX}^2$  coïncide avec la variance asymptotique de  $\hat{\sigma}_{vREML}^2$  (pour obtenir plus de détails, voir l'annexe A).

**Théorème 4.2.** Sous les conditions du théorème 4.1, le  $\text{Bias}(\hat{\sigma}_{vMIX}^2) = o(1/m)$ . La preuve est présentée à l'annexe A.

## 4.2 Estimation de l'EQM

Le fait que l'estimateur MIX  $\hat{\sigma}_{vMIX}^2$ , est sans biais jusqu'à l'ordre deux est essentiel pour démontrer que l'estimateur d'EQM que nous proposons est lui aussi sans biais jusqu'à l'ordre deux.

**Corollaire 4.2.** L'estimateur de l'EQM de l'EBLUP sous  $\hat{\sigma}_{vMIX}^2$  donné par :

$$\text{eqm}[\hat{\theta}_i(\hat{\sigma}_{vMIX}^2)] = g_{1i}(\hat{\sigma}_{vMIX}^2) + g_{2i}(\hat{\sigma}_{vMIX}^2) + 2g_{3i}(\hat{\sigma}_{vMIX}^2) \quad (4.2)$$

est sans biais jusqu'à l'ordre deux. Étant donné que  $\hat{\sigma}_{vMIX}^2$  est sans biais jusqu'à l'ordre deux, le résultat suit le modèle de Datta et Lahiri (2000).

## 4.3 Autres estimateurs de l'EQM

Dans l'équation qui suit, l'estimateur de variance MIX est la combinaison des estimateurs REML et AM.LL.

Rubin-Bleuer et You (2012) ont suggéré un autre estimateur de l'EQM, qui est lui aussi sans biais jusqu'à l'ordre deux. Il s'agit d'un estimateur de l'EQM « fractionné » qui prend la forme :

$$\text{eqm}^*[\hat{\theta}_i(\hat{\sigma}_{vMIX}^2)] = \begin{cases} g_{1i}(\hat{\sigma}_{vMIX}^2) + g_{2i}(\hat{\sigma}_{vMIX}^2) + 2g_{3i}(\hat{\sigma}_{vMIX}^2) & \text{si } \hat{\sigma}_{vMIX}^2 = \hat{\sigma}_{vREML}^2, \\ g_{1i}(\hat{\sigma}_{vMIX}^2) + g_{2i}(\hat{\sigma}_{vMIX}^2) \\ + 2g_{3i}(\hat{\sigma}_{vMIX}^2) - (1 - \hat{\gamma}_{iMIX})^2 \cdot \text{Biais}(\hat{\sigma}_{vMIX}^2) & \text{si } \hat{\sigma}_{vMIX}^2 = \hat{\sigma}_{vAM.LL}^2. \end{cases} \quad (4.3)$$

L'estimateur  $\text{eqm}^*$  présente un biais relatif moyen (BRM) plus faible que l'estimateur de l'EQM donné en (4.2). Le BRM est plus faible parce que l'EQM est surestimée lorsque le REML est positif et sous-estimée lorsque le REML est nul. L'estimateur  $\text{eqm}^*$  est généralement satisfaisant, mais il peut prendre des valeurs négatives pour un ensemble particulier de données.

Molina et coll. (2015) ont proposé deux estimateurs de l'EQM pour l'EBLUP sous la méthode MIX. Dans l'équation qui suit, le test préliminaire proposé de l'hypothèse de variance nulle est représenté par le sigle TP, et les estimateurs sont :

$$\text{eqm}_0\{\hat{\theta}_i(\hat{\sigma}_{vMIX}^2)\} = \begin{cases} \text{eqm}\{\hat{\theta}_i(\hat{\sigma}_{vREML}^2)\} & \text{si } \hat{\sigma}_{vREML}^2 > 0 \\ g_{2i}(0) & \text{si } \hat{\sigma}_{vREML}^2 = 0 \end{cases} \quad (4.4)$$

et

$$\text{eqm}_{\text{TP}} \left\{ \hat{\theta}_i \left( \hat{\sigma}_{v\text{MIX}}^2 \right) \right\} = \begin{cases} \text{eqm} \left\{ \hat{\theta}_i \left( \hat{\sigma}_{v\text{REML}}^2 \right) \right\} & \text{si } \hat{\sigma}_{v\text{REML}}^2 > 0 \text{ et le TP est rejeté} \\ g_{2i} (0) & \text{si } \hat{\sigma}_{v\text{REML}}^2 = 0 \text{ ou le TP n'est pas rejeté.} \end{cases} \quad (4.5)$$

La justification de  $\text{eqm}_0$  et  $\text{eqm}_{\text{TP}}$  se fonde sur l'EQM du BLUP où  $\sigma_v^2 = 0$ . Molina et coll. (2015) ont montré dans une étude empirique que les estimateurs de l'EQM proposés donnaient de bons résultats en moyenne lorsque  $\sigma_v^2$  et le nombre de domaines  $m$  étaient faibles.

**Remarque 4.2.** Les estimateurs  $\text{eqm}_0$  et  $\text{eqm}_{\text{TP}}$  sont eux aussi sans biais jusqu'à l'ordre deux (l'annexe contient une preuve abrégée de cette propriété). Notre argument contre  $\text{eqm} \left\{ \hat{\theta}_i \left( \hat{\sigma}_v^2 \right) \right\}$  (en 3.3) s'applique aussi à  $\text{eqm}_0$  et à  $\text{eqm}_{\text{TP}}$  : pour un nombre modéré de domaines, le pourcentage de populations où  $\hat{\sigma}_{v\text{REML}}^2 = 0$  peut être significatif, même si  $\sigma_v^2 / \psi_i$  n'est pas négligeable. Dans ce cas-ci, l'estimateur de l'EQM de l'EBLUP doit aussi tenir compte de la variation due à l'estimation de variance ou à la sous-estimation du risque.

## 5 Conditions de simulation et mesures de performance

### 5.1 Conditions de simulation

Nous avons réalisé une simulation Monte Carlo fondée sur un modèle, en suivant l'exemple de Rubin-Bleuer et You (2012), afin d'examiner la performance en échantillon fini des différentes méthodes. Les estimations « directes »  $(y_1, \dots, y_m)$  où  $m = 15, m = 45$  et  $m = 100$ , sont générées à partir du modèle de Fay-Herriot en (2.3) où  $\boldsymbol{\beta}' = (5, 4, 3, 2, 1)$  et les covariables  $\mathbf{z}'_i = (1, z_{i2}, \dots, z_{ip})$ , générées une fois à partir de distributions normales  $z_{ik} \sim k + N(1, 1)$ ,  $k = 2, \dots, 5$ ,  $i = 1, \dots, m$ , et maintenues fixes sur les populations répétées. Les effets aléatoires normaux indépendants de domaine  $v_i$  sont générés avec la variance  $\sigma_v^2 = 1$ . Des erreurs d'échantillonnage indépendantes  $e_i$ , sont générées avec des variances d'échantillonnage  $\psi_i \triangleq 50/n_i$ , où  $n_i$  est la taille de l'échantillon pour le domaine  $i, i = 1, \dots, m$ . Il y a cinq groupes de variances d'échantillonnage déterminés par  $n_i = 3, 5, 7, 10$  ou  $15$ , où les rapports signal/bruit  $\sigma_v^2 / \psi_i = 0,06; 0,1; 0,14; 0,2$  et  $0,3$ , respectivement. Ainsi, lorsque  $m = 100$ , il y a 20 domaines par rapport signal/bruit. Nous avons d'abord généré 50 000 ensembles d'estimateurs directs pour chaque cas, puis calculé l'EBLUP et l'EQM Monte Carlo réelle de l'EBLUP à l'aide des estimateurs de variance REML, AM.LL, MIX, AM.YL et AR.YL. Nous n'avons pas étudié l'estimateur AR.LL en raison de sa performance médiocre mentionnée par Li et Lahiri (2011). Nous avons ensuite généré 10 000 ensembles d'estimateurs directs indépendamment des 50 000 premiers. Pour chaque ensemble généré, nous avons calculé les cinq estimateurs de variance. Pour l'estimateur de variance MIX, nous avons examiné trois des quatre estimateurs de l'EQM par linéarisation qui font l'objet d'une discussion dans la section 4. Comme il arrive souvent que les estimateurs de l'EQM par linéarisation n'estiment pas le biais de façon exacte, nous avons également jeté un coup d'œil à l'estimateur bootstrap paramétrique de l'EQM

(BP EQM) corrigé pour le biais en utilisant la méthode de Pfeffermann et Glickman (2004) ainsi que l'estimateur BP naïf de l'EQM avec 500 répétitions chacune (voir l'annexe B pour la construction des poids bootstrap). Les mesures de performance Monte Carlo sont définies ci-après.

1. L'EQM de l'EBLUP,  $\overline{\text{EQM}}_\ell(\hat{\theta}_i)$ , par groupe de variances d'échantillonnage :

$$\text{EQM}(\hat{\theta}_i) = \frac{1}{50\,000} \sum_{r=1}^{50\,000} (\hat{\theta}_i^{(r)} - \theta_i^{(r)})^2, \quad \overline{\text{EQM}}_\ell(\hat{\theta}_i) = \frac{5}{m} \sum_{i \in \{j: \psi_j = 50/n_\ell\}} \text{EQM}(\hat{\theta}_i), \quad \ell = 1, \dots, 5.$$

2.  $E(\hat{\sigma}_v^2) = \sum_{r=1}^{10\,000} \hat{\sigma}_v^{2(r)} / 10\,000$ ,  $V(\hat{\sigma}_v^2) = \sum_{r=1}^{10\,000} (\hat{\sigma}_v^{2(r)} - E(\hat{\sigma}_v^2))^2 / 10\,000$ , où  $\hat{\sigma}_v^{2(r)}$  est la valeur de  $\hat{\sigma}_v^2$  pour la  $r^{\text{e}}$  simulation ( $r = 1, \dots, 10\,000$ ).

3. Le biais relatif moyen (BRM) de l'EQM par groupe de variances d'échantillonnage :

$$\text{BRM}_\ell(\text{eqm}) = \frac{5}{m} \sum_{i \in \{j: \psi_j = 50/n_\ell\}} \text{BR}(\text{eqm}(\hat{\theta}_i)), \quad \ell = 1, \dots, 5,$$

$$\text{où } \text{BR}(\text{eqm}(\hat{\theta}_i)) = \left[ \sum_{r=1}^{10\,000} \text{eqm}(\hat{\theta}_i^{(r)}) / 10\,000 - \text{EQM}(\hat{\theta}_i) \right] / \text{EQM}(\hat{\theta}_i).$$

4. La racine de l'EQM relative des estimateurs de l'EQM par groupe de variances d'échantillonnage :

$$\text{REQMR}_\ell(\text{eqm}) = \left( \frac{5}{m} \sum_{i \in \{j: \psi_j = 50/n_\ell\}} \frac{\sum_{r=1}^{10\,000} (\text{eqm}(\hat{\theta}_i^{(r)}) - \text{EQM}(\hat{\theta}_i))^2 / 10\,000}{\text{EQM}(\hat{\theta}_i)} \right)^{1/2}.$$

Nous examinons également le biais des estimateurs conditionnels de l'EQM étant donné que  $\{\hat{\sigma}_{\text{vREML}}^2 = 0\}$ , car ce sont les populations pour lesquelles les estimateurs positifs ont été élaborés.

5. Le biais relatif moyen des estimateurs conditionnels de l'EQM :

$$\text{BRM}_C = \frac{5}{m} \sum_{i \in \ell} E[\text{eqm}(\hat{\theta}_i) | \hat{\sigma}_{\text{vREML}}^2 = 0] / E[(\hat{\theta}_i - \theta_i)^2 | \hat{\sigma}_{\text{vREML}}^2 = 0] - 1.$$

## 6 Résultats de la simulation et analyse

### 6.1 Distribution Monte Carlo des estimateurs de variance

Le tableau 6.1 montre que l'estimateur de variance REML présente le biais le plus faible ( $\sigma_v^2 = 1$ ) et la variance la plus élevée. L'efficacité plus faible du REML pourrait être attribuable au fait qu'il ne s'agit pas d'une fonction lisse des données causée par sa définition fractionnée (3.1). L'estimateur MIX hérite d'une partie de cette efficacité faible. Les autres estimateurs de variance ont une variabilité plus faible et un biais

positif plus élevé, mais l'espérance conditionnelle des estimateurs AM.YL et AR.YL étant donné que  $\hat{\sigma}_{\text{vREML}}^2 = 0$  est proche de zéro. Le biais inconditionnel de l'estimateur AM.LL est plus élevé que celui du MIX. Selon la définition de l'estimateur MIX, les biais conditionnels des estimateurs MIX et AM.LL coïncident. En outre, le MIX converge plus rapidement que les autres estimateurs. Par exemple, étant donné la distribution des probabilités sur les 10 000 estimations de la variance où  $m = 45$ , nous avons calculé la probabilité que les estimations se situent dans un intervalle contenant  $\sigma_v^2 = 1$ . La probabilité que les estimations se situent entre 0,6 et 1,4 est de 0,47 pour le MIX et de 0,16 pour l'AM.YL. Par ailleurs, la probabilité que les estimations soient inférieures à 0,2 est de 0,05 pour le MIX et de 0,53 pour l'AM.YL.

**Tableau 6.1**  
Espérance, variance et espérance et variance conditionnelle de  $\hat{\sigma}_v^2$

Méthode	$m$	$E(\hat{\sigma}_v^2)$	$V(\hat{\sigma}_v^2)$	%REML = 0	$E(\hat{\sigma}_v^2/\text{REML} = 0)$	$V(\hat{\sigma}_v^2/\text{REML} = 0)$
REML	15	1,48	3,38	43 %	N/A	N/A
	45	1,21	1,67	29 %	N/A	N/A
	100	1,07	0,81	16 %	N/A	N/A
AM.LL	15	2,80	1,37	43 %	1,80	0,11
	45	1,88	1,01	29 %	0,94	0,03
	100	1,49	0,51	16 %	0,63	0,01
MIX	15	2,28	1,87	43 %	1,80	0,11
	45	1,48	1,31	29 %	0,94	0,03
	100	1,17	0,66	16 %	0,63	0,01
AR.YL	15	1,66	2,99	43 %	0,27	0,01
	45	1,24	1,72	29 %	0,06	0,00
	100	1,08	0,80	16 %	0,02	0,00
AM.YL	15	0,52	0,84	43 %	0,10	0,00
	45	0,65	0,85	29 %	0,03	0,00
	100	0,76	0,59	16 %	0,01	0,00

## 6.2 EQM réelle de l'EBLUP, biais relatif moyen et racine de l'EQM relative moyenne des estimateurs de l'EQM

Tous les estimateurs de variance sont convergents et asymptotiquement normaux, la variance convergeant au même rythme. Leurs biais sont différents : ceux des estimateurs REML, AR.YL et MIX sont de l'ordre  $o(1/m)$ , tandis que ceux des estimateurs AM.LL et AM.YL sont de l'ordre  $O(1/m)$ . Le biais inhérent aux trois dernières méthodes ont un impact sur l'estimation de l'EQM de l'EBLUP, même pour un nombre modéré de domaines.

Pour  $m = 100$ , les tableaux 6.2a et 6.2b montrent que l'EQM de l'EBLUP diminue à mesure que  $\sigma_v^2/\psi_i$  augmente, et que cette relation se maintient quel que soit le nombre de domaines. Nous observons que l'EQM de  $\hat{\theta}_i$  sous les estimateurs de variance REML et MIX est légèrement plus élevée que le reste des EQM en raison de la plus grande variabilité inhérente à ces estimateurs de variance. Le tableau 6.2a présente les résultats pour l'estimateur de l'EQM par linéarisation de Taylor et les deux estimateurs paramétriques de l'EQM sous les estimateurs de variance REML, AM.LL, AR.YL et AM.YL. Le tableau 6.2b présente les résultats pour les estimateurs suivants de l'EQM sous l'estimation de variance MIX : RB\_Y1 défini en (4.3), RB\_Y2 défini en (4.2), M\_et\_coll défini en (4.5), BP EQM et BP EQM naïf. Parmi les estimateurs de

l'EQM de Taylor, RB\_Y1 et M\_et\_coll sous la méthode MIX présentent le biais le plus faible. Parmi les estimateurs bootstrap de l'EQM, BP sous MIX et BP naïf sous AR.YL présentent le biais le plus faible. Quant à la racine de l'erreur quadratique moyenne relative (REQMR) des estimateurs de l'EQM, elle diminue à mesure que  $\sigma_v^2/\psi_i$  augmente. Les différences entre l'estimateur RB\_Y2 de l'EQM sous le MIX et l'estimateur de l'EQM de Taylor sous l'AM.YL semblent faibles mais consistantes. Alors que le RB\_Y1, le M\_et\_coll et les estimateurs naïfs de l'EQM sous la méthode MIX présentent un BRM plus faible que le RB\_Y2 sous la même méthode, et que le BRM de l'estimateur de Taylor et de l'estimateur BP naïf sous la méthode AR.YL est plus faible que celui du RB\_Y2 sous la méthode MIX, c'est le contraire pour la REQMR. Cela s'explique en partie par le biais conditionnel négatif extrême de ces estimateurs de l'EQM (c'est-à-dire les estimateurs RB\_Y1 et M\_et\_coll sous la méthode MIX et les estimateurs de Taylor et BP naïf sous la méthode AR.YL), comme le montre le tableau 6.3. Même pour  $m = 100$ , une proportion relativement élevée (16 %) des populations donnent  $\hat{\sigma}_{vREML}^2 = 0$  et, dans ces populations, les estimations obtenues au moyen de la plupart des méthodes d'estimation de variance et la plupart des estimateurs de l'EQM sont les plus inférieures à la valeur réelle. C'est-à-dire que, pour ces estimateurs de l'EQM, les estimateurs conditionnels ne donnent pas de bons résultats. L'estimateur BP EQM semble corriger pour le biais de façon satisfaisante, mais il est plus variable que le BP EQM naïf. Lorsque nous incluons le BRM, la REQMR et le  $BRM_C$  dans l'évaluation, c'est le RB\_Y2 sous la méthode MIX, suivi de près par l'estimateur BP naïf, qui l'emporte sous la méthode MIX qui semble donner les meilleurs résultats. Nous pourrions donc conclure à la supériorité des estimateurs RB\_Y2 et naïf sous la méthode MIX pour  $m = 100$ , ce qui est un nombre modéré de domaines pour les données de ce genre.

Tableau 6.2a

EQM, BRM et REQMR (pourcentage) des estimateurs de l'EQM,  $m = 100$ 

Méthode	$\sigma_v^2/\psi_i$	EQM	Estimateur de l'EQM de Taylor		Estimateur BP		Estimateur BP naïf	
			BRM	REQMR	BRM	REQMR	BRM	REQMR
REML	0,06	135,4	5,1	71,1	-4,4	80,7	1,6	69,9
	0,1	132,1	5,3	64,7	-4,7	74,0	-0,2	63,0
	0,14	119,5	6,0	61,9	-5,5	71,3	-1,8	59,9
	0,2	119,2	6,5	53,6	-5,8	62,4	-3,4	51,7
	0,3	106,6	8,2	46,7	-6,8	55,0	-5,6	44,8
AM.LL	0,06	134,9	6,1	75,4	8,2	66,9	31,3	63,8
	0,1	131,2	6,8	68,1	7,8	59,5	27,5	55,7
	0,14	118,3	8,1	64,6	7,8	55,6	26,5	51,2
	0,2	117,6	8,4	55,4	6,5	46,7	21,6	42,1
	0,3	104,5	10,2	46,7	5,5	38,8	18,2	34,0
AR.YL	0,06	135,4	6,6	69,3	-4,3	80,2	2,1	69,4
	0,1	132,0	7,4	61,9	-4,5	73,4	0,3	62,5
	0,14	119,4	9,0	58,0	-5,3	70,6	-1,2	59,3
	0,2	119,0	10,6	48,2	-5,6	61,8	-2,9	51,1
	0,3	106,4	14,7	38,5	-6,6	54,3	-5,1	44,1
AM.YL	0,06	134,7	10,0	63,2	-12,3	81,0	-19,6	65,9
	0,1	131,3	12,0	56,6	-12,5	75,2	-19,7	61,2
	0,14	118,8	15,0	53,1	-13,7	73,3	-21,4	59,8
	0,2	118,6	18,1	44,8	-13,4	65,2	-20,7	53,5
	0,3	106,4	25,2	38,4	-14,4	58,8	-21,7	48,6

**Tableau 6.2b**  
EQM, BRM et REQMR (pourcentage) des estimateurs de l'EQM,  $m = 100$

	$\sigma_v^2/\psi_i$	EQM	RB_Y1		RB_Y2		M_et_coll		Estimateur BP		Estimateur BP naïf	
			BRM	REQMR	BRM	REQMR	BRM	REQMR	BRM	REQMR	BRM	REQMR
MIX	0,06	135,4	2,7	75,7	13,6	63,0	5,2	71,1	-3,0	75,3	8,8	62,4
	0,1	132,1	3,6	68,3	14,9	56,1	5,3	64,7	-3,2	68,3	6,6	55,4
	0,14	119,5	4,9	64,7	16,0	52,4	6,0	61,9	-3,9	65,1	5,3	51,8
	0,2	119,1	6,3	55,2	16,7	43,8	6,5	53,6	-4,4	56,3	2,9	43,7
	0,3	106,5	9,4	46,2	19,9	36,0	8,3	46,7	-5,4	48,6	0,6	36,7

**Tableau 6.3**  
EQM<sub>C</sub> ( $E[(\hat{\theta}_i - \theta_i)^2 | \hat{\sigma}_{v,REML}^2 = 0]$ ) et BRM<sub>C</sub> (pourcentage),  $m = 100$

Méthode	$\sigma_v^2/\psi_i$	EQM <sub>C</sub>	Estimateur de l'EQM de Taylor			Estimateur BP	Estimateur BP naïf
REML	0,06	135,6					
	0,1	133,0					
	0,14	121,5					
	0,2	120,4					
	0,3	108,0					
AM.LL	0,06	135,0					
	0,1	132,2					
	0,14	120,2					
	0,2	118,8					
	0,3	105,9					
AR.YL	0,06	135,5					
	0,1	132,9					
	0,14	121,4					
	0,2	120,2					
	0,3	107,8					
AM.YL	0,06	134,9					
	0,1	132,1					
	0,14	120,4					
	0,2	119,6					
	0,3	107,6					
			RB_Y1	RB_Y2	M_et_coll	Estimateur BP	Estimateur BP naïf
MIX	0,06	135,0	-92,0	-22,0	-76,4	-46,0	-27,0
	0,1	132,2	-85,2	-17,7	-74,3	-42,7	-25,9
	0,14	120,2	-85,4	-15,0	-78,3	-43,3	-27,0
	0,2	118,8	-74,4	-7,6	-72,8	-37,6	-23,9
	0,3	105,9	-65,9	1,5	-73,1	-34,6	-22,6

Les tableaux 6.4a et 6.4b ci-dessous présentent les résultats pour  $m = 45$  avec 9 domaines par  $\sigma_v^2/\psi_i$ . L'estimateur AM.YL donne des EQM plus petites que le MIX, les différences ne dépassant pas 2 %. Le biais des estimateurs de variance augmente à mesure que le nombre de domaines diminue, ce qui a un impact sur les estimateurs de l'EQM. En effet, les BRM de tous les estimateurs de l'EQM ont augmenté. En particulier, le BRM des estimateurs de l'EQM de Taylor sous l'estimation de variance YL et LL et le BRM de l'estimateur RB\_Y2 ont augmenté de 100 % par rapport au BRM avec 100 domaines. En ce qui concerne la REQMR, l'estimateur de l'EQM de Taylor sous AM.YL a une REQMR légèrement inférieure à celle du RB\_Y2 sous la méthode MIX pour une  $\sigma_v^2/\psi_i$  très faible. En général, la variabilité (en termes de REQMR) du RB\_Y2 est plus faible que celle de l'estimateur de Taylor sous LL et YL et des estimateurs RB\_Y1 et

M\_et\_coll. Cela pourrait être attribuable en partie à la sous-estimation des EQM pour les populations avec des estimations REML nulles, dont le pourcentage tourne autour de 30 % lorsque  $m = 45$ . Le tableau 6.5 est plus instructif à cet égard : étant donné  $\hat{\sigma}_{vREML}^2 = 0$ , RB\_Y1 et M\_et\_coll conduisent à une sous-estimation importante.

Tableau 6.4a

EQM, BRM et REQMR (pourcentage) des estimateurs de l'EQM,  $m = 45$  domaines

Méthode	$\sigma_v^2/\psi_i$	$\overline{\text{EQM}}$	Estimateur de l'EQM de Taylor		Estimateur BP		Estimateur BP naïf	
			BRM	REQMR	BRM	REQMR	BRM	REQMR
REML	0,06	171,4	11,8	94,7	-4,7	107,0	6,2	89,2
	0,1	174,1	11,9	83,9	-5,3	93,8	3,0	76,2
	0,14	171,3	12,6	74,5	-5,4	81,9	1,1	65,3
	0,2	166,6	13,9	63,4	-5,8	66,7	-1,2	52,0
	0,3	128,9	20,1	63,0	-7,0	61,4	-3,1	46,7
AM.LL	0,06	171,1	15,5	100,0	16,0	84,9	43,5	83,3
	0,1	173,4	16,8	87,0	14,4	71,1	36,7	68,5
	0,14	170,4	17,7	75,7	12,6	59,7	30,7	56,7
	0,2	165,3	18,2	61,7	9,9	46,2	23,5	43,2
	0,3	127,5	25,6	55,0	10,0	39,7	22,6	36,6
AR.YL	0,06	171,1	17,2	89,9	-3,7	105,0	8,0	87,6
	0,1	173,6	19,6	76,9	-4,3	91,8	4,8	74,6
	0,14	170,8	22,6	65,8	-4,4	79,9	2,7	63,7
	0,2	166,0	27,3	53,7	-4,8	64,8	0,3	50,5
	0,3	128,3	43,8	54,8	-5,7	59,3	-1,3	45,0
AM.YL	0,06	167,5	30,2	78,4	-18,0	97,3	-23,8	73,3
	0,1	169,6	36,5	72,2	-18,0	87,7	-23,6	66,7
	0,14	167,0	42,7	69,3	-17,2	78,0	-22,3	59,7
	0,2	162,8	52,1	70,8	-15,8	65,4	-20,3	50,6
	0,3	126,0	81,3	91,1	-18,0	62,3	-22,9	48,4

Tableau 6.4b

EQM, BRM et REQMR (pourcentage) des estimateurs de l'EQM,  $m = 45$  domaines

	$\sigma_v^2/\psi_i$	$\overline{\text{EQM}}$	RB_Y1		RB_Y2		M_et_coll		Estimateur BP		Estimateur BP naïf	
			BRM	REQMR	BRM	REQMR	BRM	REQMR	BRM	REQMR	BRM	REQMR
MIX	0,06	171,4	9,8	99,4	31,9	84,0	11,8	94,7	3,5	93,8	21,9	78,5
	0,1	174,0	12,1	86,2	33,2	73,1	11,9	83,9	2,6	80,4	17,5	65,1
	0,14	171,2	14,5	74,9	34,4	64,6	12,6	74,5	2,0	68,7	14,0	54,4
	0,2	166,5	17,7	61,7	36,0	55,8	13,9	63,4	0,7	54,5	9,8	41,8
	0,3	128,9	28,8	57,6	48,8	58,2	20,2	63,1	0,3	48,6	8,7	35,9

Compte tenu du BRM, de la REQMR et du  $\text{BRM}_c$  des estimateurs de l'EQM, c'est l'estimateur BP naïf de l'EQM sous la méthode MIX qui donne les meilleurs résultats pour les  $\sigma_v^2/\psi_i$  plus importantes. Le tableau 6.6 présente la moyenne des mesures de performance, calculée sur les cinq groupes de variances



d'échantillonnage, pour les trois estimateurs de l'EQM de Taylor sous la méthode MIX avec des données du modèle décrit en 5.1, mais avec trois valeurs différentes de  $\sigma_v^2$ . L'estimateur RB\_Y2 donne de meilleurs résultats quand  $\sigma_v^2 = 1$ , mais lorsque  $\sigma_v^2$  diminue, c'est l'estimateur de l'EQM de M\_et\_coll qui prend le dessus, précisément parce qu'il a été construit en partant du principe que  $\sigma_v^2$  est d'environ zéro.

**Tableau 6.5**  
EQM<sub>c</sub> et BRM<sub>c</sub> (pourcentage).  $m = 45$  domaines

Méthode	$\sigma_v^2 / \psi_i$	EQM <sub>c</sub>	Estimateur de l'EQM de Taylor			Estimateur BP	Estimateur BP naïf
REML	0,06	170,2	-64,3			-89,7	-60,7
	0,1	173,0	-62,4			-83,7	-57,1
	0,14	170,2	-58,1			-75,5	-51,8
	0,2	165,8	-51,9			-65,1	-44,8
	0,3	131,1	-59,0			-70,5	-49,2
AM.LL	0,06	170,0	-71,5			-49,0	-3,1
	0,1	172,3	-61,5			-42,1	-2,3
	0,14	169,1	-51,1			-35,7	-2,1
	0,2	164,7	-38,3			-28,3	-1,6
	0,3	129,9	-28,8			-29,1	-3,7
AR.YL	0,06	169,9	-48,3			-86,2	-56,7
	0,1	172,6	-38,0			-80,2	-53,2
	0,14	169,7	-25,9			-72,2	-48,2
	0,2	165,3	-7,4			-61,9	-41,5
	0,3	130,5	19,3			-66,8	-45,5
AM.YL	0,06	166,6	-8,2			-73,5	-60,7
	0,1	168,8	3,8			-70,1	-58,1
	0,14	166,1	16,1			-64,1	-53,3
	0,2	162,2	35,9			-56,1	-46,8
	0,3	128,1	72,8			-62,5	-52,5
			<b>RB_Y1</b>	<b>RB_Y2</b>	<b>M_et_coll</b>		
MIX	0,06	170,0	-71,5	6,2	-64,3	-28,1	-4,0
	0,1	172,3	-61,5	13,2	-62,3	-23,8	-3,5
	0,14	169,1	-51,1	18,9	-57,8	-20,0	-3,3
	0,2	164,7	-38,3	26,8	-51,6	-15,7	-2,9
	0,3	129,9	-28,8	40,4	-58,7	-16,7	-5,1

**Tableau 6.6**  
EQM, BRM, BRM<sub>c</sub> et REQMR (pourcentage), 45 domaines

%REML = 0	$\sigma_v^2$	EQM	RB_Y1			RB_Y2			M_et_coll		
			BRM	BRM <sub>c</sub>	REQMR	BRM	BRM <sub>c</sub>	REQMR	BRM	BRM <sub>c</sub>	REQMR
29	1	108	16	-50	75	36	21	66	14	-59	75
48	0,2	99	48	-36	101	113	88	114	47	-38	94
51	0,1	91	58	-33	108	137	107	127	58	-32	100

Les tableaux 6.7a et 6.7b ci-dessous montrent les résultats pour  $m = 15$  domaines avec 3 domaines par  $\sigma_v^2 / \psi_i$ . Les EQM ne varient pas de plus de 5 %, quelle que soit la méthode d'estimation de variance utilisée.

Il n'existe pas de relation monotone entre le BRM ou la REQMR et  $\sigma_v^2 / \psi_i$ , ce qui pourrait indiquer que l'approximation d'ordre deux pour estimer l'EQM est médiocre, quelle que soit la méthode d'estimation de variance utilisée. Les BRM des estimateurs de l'EQM de Taylor sous les méthodes d'estimation de

variance de LL et YL sont trop élevés, et il en va de même pour la REQMR. L'estimateur RB\_Y2 sous la méthode MIX ne s'en sort pas très bien non plus. La raison de ce résultat est claire : le pourcentage élevé d'estimations REML nulles (43 %) suggère que le MIX coïncide avec l'AM.LL pour les populations REML nulles. Ainsi, le MIX a un biais positif pour  $m = 15$ , et le RB\_Y2 ne tient pas compte de ce biais. Le RB\_Y1 tient compte du biais dans le MIX, mais l'estimateur du biais n'est pas très précis pour  $m = 15$ . L'estimateur de l'EQM de M\_et\_coll coïncide presque avec le BRM et la REQMR de l'estimateur de l'EQM de Taylor sous l'estimation de variance REML, car ils sont égaux par définition lorsque  $\hat{\sigma}_{vREML}^2 = 0$ . Le BRM<sub>C</sub> des trois estimateurs de l'EQM de Taylor sous la méthode MIX laisse à désirer. Compte tenu de toutes les mesures de performance, les estimateurs bootstrap de l'EQM donnent de meilleurs résultats que les estimateurs de l'EQM de Taylor. Pour  $m = 15$  domaines avec 3 domaines par  $\sigma_v^2/\psi_i$ , l'estimateur BP sous la méthode MIX donne les meilleurs résultats, l'estimateur BP naïf sous AR.YL et AM.YL venant en deuxième place.

Tableau 6.7a

EQM, BRM et REQMR (pourcentage) des estimateurs de l'EQM,  $m = 15$  domaines

Méthode	$\sigma_v^2/\psi_i$	EQM	Estimateur de l'EQM de Taylor		Estimateur BP		Estimateur BP naïf	
			BRM	REQMR	BRM	REQMR	BRM	REQMR
REML	0,06	584,8	12,6	87,9	1,2	85,9	6,9	64,5
	0,1	376,7	26,5	106,3	2,3	85,6	9,6	62,8
	0,14	352,5	25,2	90,1	0,7	54,1	4,3	39,3
	0,2	209,4	43,0	123,0	0,4	74,0	6,3	51,1
	0,3	198,7	50,6	124,7	-1,0	46,3	2,6	31,5
AM.LL	0,06	589,3	24,1	89,3	13,7	61,2	24,1	65,8
	0,1	380,7	48,3	107,1	19,4	58,6	32,5	62,9
	0,14	355,7	40,2	88,6	10,0	36,2	16,8	38,1
	0,2	212,5	76,3	117,9	17,8	45,1	28,7	47,3
	0,3	200,7	76,5	105,1	10,7	26,9	17,2	27,6
AR.YL	0,06	583,3	23,8	83,3	3,2	79,5	3,2	61,6
	0,1	375,1	53,3	106,7	5,4	78,6	5,4	59,7
	0,14	351,3	53,3	102,7	2,4	49,4	2,4	37,1
	0,2	207,7	107,3	153,1	4,1	66,2	4,1	47,2
	0,3	197,5	142,0	199,4	1,9	41,1	1,9	28,9
AM.YL	0,06	571,4	41,6	103,5	-8,0	61,2	-9,2	43,3
	0,1	363,3	95,0	161,4	-11,3	62,9	-13,2	44,1
	0,14	342,0	97,2	179,7	-6,7	40,4	-7,8	29,3
	0,2	197,0	198,4	274,6	-14,5	58,2	-16,7	41,7
	0,3	191,4	270,2	362,4	-11,5	38,4	-13,1	28,7

Tableau 6.7b

EQM, BRM et REQMR (pourcentage) des estimateurs de l'EQM,  $m = 15$  domaines

	$\sigma_v^2/\psi_i$	EQM	RB_Y1		RB_Y2		M_et_coll		Estimateur BP		Estimateur BP naïf	
			BRM	REQMR	BRM	REQMR	BRM	REQMR	%BRM	%REQMR	%BRM	%REQMR
MIX	0,06	584,9	21,0	84,7	35,4	93,7	12,6	87,9	10,0	53,8	19,3	62,1
	0,1	377,1	46,0	103,9	68,4	122,6	26,4	106,1	14,8	52,7	26,6	59,9
	0,14	353,0	41,9	91,5	59,4	112,7	25,0	89,9	7,6	33,2	13,7	36,7
	0,2	209,7	83,2	127,8	108,9	155,8	42,8	122,8	14,0	42,5	23,7	46,0
	0,3	198,9	94,8	136,7	117,1	162,2	50,4	124,6	8,7	26,6	14,5	27,7

En résumé, sous le modèle de Fay-Herriot avec une  $\sigma_v^2$  positive, MIX et AR.YL sont les seuls estimateurs de variance positifs à l'étude qui présentent un biais asymptotique négligeable. Le biais asymptotique des estimateurs de variance AM.YL et LL est plus grand. En revanche, notre simulation a démontré que, pour un nombre modéré de domaines et pour les populations qui donnent des estimations REML nulles, les deux estimateurs de variance de YL présentaient un biais négatif et produisaient des EBLUP proches de l'estimateur synthétique de la moyenne. Par contre, le MIX, qui combine les estimateurs AM.LL et REML, ne présentait qu'un biais légèrement négatif dans ces populations. De plus, la distribution inconditionnelle du MIX approchait de la normalité beaucoup plus rapidement que celle des autres estimateurs de variance.

**Tableau 6.8**  
**EQM<sub>c</sub> et BRM<sub>c</sub> m = 15 domaines**

Méthode	$\sigma_v^2/\psi_i$	$\overline{\text{EQM}}_c$	Estimateur de l'EQM de Taylor				
			Estimateur BP	Estimateur BP naïf			
REML	0,06	594,2	-22,6	-31,7	-16,5		
	0,1	381,2	-32,9	-43,2	-22,5		
	0,14	345,1	-17,7	-22,7	-10,7		
	0,2	212,7	-41,1	-47,3	-25,5		
	0,3	197,9	-30,4	-32,7	-17,6		
AM.LL	0,06	595,6	-4,1	-5,7	12,1		
	0,1	385,7	8,6	-7,0	15,6		
	0,14	351,2	18,9	-2,0	10,4		
	0,2	216,0	46,4	-5,8	14,4		
	0,3	199,5	67,0	-2,9	9,8		
AR.YL	0,06	592,2	-0,8	-27,1	-11,0		
	0,1	379,7	21,0	-36,5	-14,8		
	0,14	344,5	44,0	-18,6	-6,3		
	0,2	210,9	98,2	-38,6	-16,4		
	0,3	196,6	177,3	-26,1	-11,0		
AM.YL	0,06	581,7	30,7	-21,9	-18,0		
	0,1	368,6	79,8	-31,5	-25,8		
	0,14	333,9	98,3	-15,2	-11,9		
	0,2	198,9	198,0	-36,4	-30,0		
	0,3	190,0	296,3	-26,2	-21,5		
			RB_Y1	RB_Y2	M_et_coll	Estimateur BP	Estimateur BP naïf
MIX	0,06	595,6	-4,1	27,9	-22,9	3,4	17,8
	0,1	385,7	8,6	57,1	-33,7	5,1	22,8
	0,14	351,2	18,9	58,5	-19,1	4,9	14,3
	0,2	216,0	46,4	102,4	-42,0	5,9	20,4
	0,3	199,5	67,0	116,3	-30,9	4,8	13,4

En ce qui concerne l'estimateur de l'EQM de l'EBLUP, il était beaucoup plus précis que l'estimateur direct, sous toutes les méthodes d'estimation de variance examinées ici, même pour un petit nombre de domaines. Les estimateurs de variance AM.LL, AM.YL et AR.YL étaient tous moins variables que le REML et le MIX. L'impact sur l'EQM de l'EBLUP était minime, car il y avait peu de différences entre les EQM pour le même rapport signal/bruit. Ces différences s'accroissaient à mesure que le nombre de domaines ou le rapport signal/bruit diminuait. Ainsi, pour un rapport signal/bruit extrêmement faible, l'EQM sous la méthode MIX pourrait être un peu plus grande que sous l'estimateur de variance AM.YL.

Sous la méthode MIX d'estimation de variance, nous avons comparé trois estimateurs de l'EQM de type Taylor et deux estimateurs de l'EQM bootstrap. Les trois estimateurs de l'EQM de Taylor sous la méthode MIX (RB\_Y1, RB\_Y2 et M\_et\_coll) sont sans biais jusqu'à l'ordre deux. Les estimateurs de l'EQM de type Taylor sous LL et YL sont eux aussi sans biais jusqu'à l'ordre deux. Les estimateurs RB\_Y1, AM.LL et AM.YL peuvent produire des estimations négatives de l'EQM.

L'estimateur de l'EQM de Taylor sous la méthode REML d'estimation de variance et l'estimateur M\_et\_coll sous la méthode MIX coïncidant par définition, les différences entre leurs mesures de performance sont négligeables (leurs EQM réelles sont différentes; dans notre étude cependant, pour  $m = 100$ , l'estimateur MIX coïncidait avec le REML 84 % du temps). Pour un nombre modéré de domaines, qui pourrait être  $m = 45$  ou  $100$  pour ces données, et pour les populations qui donnent des estimations REML nulles, les deux estimateurs de l'EQM de Taylor sous le REML et les estimateurs de l'EQM de M\_et\_coll ne tiennent pas compte de la variation attribuable à l'estimation de  $\sigma_v^2$ , ce qui se reflète dans leur  $BRM_C$ , très négatif, qui est inférieur à -60 % pour les rapports signal/bruit plus petits. Par ailleurs, l'estimateur RB\_Y1 tient compte de la variation due à l'estimation de  $\sigma_v^2$ , mais son  $BRM_C$  est lui aussi très négatif. En effet, le RB\_Y1 est un estimateur de l'EQM fractionné qui, pour les populations où  $\hat{\sigma}_{vREML}^2 = 0$ , soustrait un facteur du biais inconditionnel de l'AM.LL, qui est toujours positif. Une meilleure formule pour un estimateur de l'EQM fractionné serait d'utiliser un estimateur du biais conditionnel  $E(\hat{\sigma}_v^2 / \hat{\sigma}_{vREML}^2 = 0)$ . En fait, même pour un nombre modéré de domaines ( $m = 100$ ), le tableau 6.1 montre que l'estimateur MIX a un biais inconditionnel de 49 %, mais un biais conditionnel de -37 %.

L'estimateur BP de l'EQM sous les méthodes AR.YL et MIX était bien corrigé pour le biais, mais la variance en souffrait. Le bootstrap naïf semble être l'estimateur de l'EQM qui donne les meilleurs résultats, et ces résultats sont encore meilleurs sous l'estimation de variance MIX, même lorsque les trois mesures (BRM,  $BRM_C$  et REQMR) sont prises en compte. Nous avons constaté que, pour un nombre modéré de domaines, le RB\_Y2 était l'estimateur de Taylor sous la méthode MIX qui présentait la REQMR la plus faible. Par ailleurs, l'estimateur de M\_et\_coll est le plus fiable lorsque la variance sous-jacente réelle  $\sigma_v^2$  est très faible : dans ce cas-ci, M\_et\_coll est effectivement l'estimateur de l'EQM de l'estimateur synthétique de la moyenne de petit domaine. Nous ne recommandons pas de compter sur l'approximation d'ordre deux de l'EQM lorsque  $m$  est petit : l'approximation (2.6) de l'EQM ne tient pas nécessairement, les mesures de performance obtenues dans notre étude sont très instables, et elles peuvent varier d'un ensemble de données à l'autre.

En conclusion, sous l'hypothèse  $\sigma_v^2 > 0$ , les performances relatives des estimateurs de variance positifs comparés dépendent de la taille de  $\sigma_v^2$ , du rapport signal/bruit, du nombre de domaines et de la fonction objectif. Pour un nombre modéré de domaines, l'estimateur de variance MIX semblait donner de meilleurs résultats que les estimateurs LL et YL dans cette étude. Sous la méthode MIX, l'estimateur BP naïf de l'EQM présentait la combinaison  $BRM_C$  et REQMR la plus faible. L'estimateur de l'EQM de M\_et\_coll sous l'estimateur de variance MIX donnait des résultats légèrement meilleurs que le RB\_Y1 lorsque la variance  $\sigma_v^2$  sous-jacente était très petite. Cependant, le pourcentage de REML nuls produits sous le modèle de simulation montre qu'un résultat de  $\hat{\sigma}_{vREML}^2 = 0$  et/ou que des tests d'hypothèses négatifs ne signifient pas nécessairement que la variance  $\sigma_v^2$  est assez faible pour que l'on puisse se fier à l'estimateur de M\_et\_coll. En l'absence d'autres renseignements, l'estimateur BP naïf sous la méthode MIX semble donner de meilleurs résultats.

## Remerciements

Les auteurs désirent remercier le professeur J.N.K. Rao de l'Université Carleton pour ses commentaires utiles ainsi que Victor Estevao de Statistique Canada, qui a développé un algorithme de maximisation de grille spécialement pour ce projet. Ils tiennent également à remercier les membres du comité de revue pour leur évaluation consciencieuse de cet article et pour leurs suggestions d'amélioration.

## Annexe A

### Preuve du théorème 4.1

La variance asymptotique de  $\hat{\sigma}_{vMIX}^2$  est donnée par :  $\bar{V}(\hat{\sigma}_{vMIX}^2) = \lim_{m \rightarrow \infty} E(\hat{\sigma}_{vMIX}^2 - \sigma_v^2)^2$

Nous montrons que  $E(\hat{\sigma}_{vMIX}^2 - \sigma_v^2)^2 \leq E(\hat{\sigma}_{vREML}^2 - \sigma_v^2)^2 + o(1/m)$  à mesure que  $m \rightarrow \infty$ .

$$\begin{aligned} E(\hat{\sigma}_{vMIX}^2 - \sigma_v^2)^2 &= \int_{\{\hat{\sigma}_{vREML}^2 > 0\}} (\hat{\sigma}_{vREML}^2 - \sigma_v^2)^2 dP + \int_{\{\hat{\sigma}_{vREML}^2 = 0\}} (\hat{\sigma}_{vAM.LL}^2 - \sigma_v^2)^2 dP \\ &\leq \int_{\Omega} (\hat{\sigma}_{vREML}^2 - \sigma_v^2)^2 dP + \int_{\{\hat{\sigma}_{vREML}^2 = 0\}} (\hat{\sigma}_{vAM.LL}^2 - \sigma_v^2)^2 dP = E(\hat{\sigma}_{vREML}^2 - \sigma_v^2)^2 \quad (A.1) \\ &\quad + o\left(\frac{1}{m}\right). \end{aligned}$$

En effet, par les inégalités de Holder et Minkowski, avec  $1 < p < \infty, 1/p + 1/q = 1$ , et si  $X \equiv (\hat{\sigma}_{vAM.LL}^2 - \sigma_v^2)^2 = O_p(1/m)$  et l'indicateur  $I(\hat{\sigma}_{vREML}^2 = 0)$  des populations avec  $\hat{\sigma}_{vREML}^2 = 0$ , nous avons :

$$\begin{aligned} \int_{\{\hat{\sigma}_{vREML}^2 < 0\}} (\hat{\sigma}_{vAM.LL}^2 - \sigma_v^2)^2 dP &\leq \left( \int_{\Omega} (\hat{\sigma}_{vAM.LL}^2 - \sigma_v^2)^{2p} dP \right)^{1/p} \cdot (P\{\hat{\sigma}_{vREML}^2 = 0\})^{1/q} \\ &= \left( O\left(\frac{1}{m^p}\right) \right)^{1/p} \cdot (o(1))^{1/q} = o\left(\frac{1}{m}\right), \end{aligned} \quad (A.2)$$

puisque  $(\hat{\sigma}_{vAM.LL}^2 - \sigma_v^2)^2$  est uniformément borné et  $\hat{\sigma}_{vREML}^2 \xrightarrow{P} \sigma_v^2 > 0$ . Il est à noter que les estimateurs AM.LL et REML de  $\sigma_v^2$  sont uniformément bornés en conséquence de leur convergence presque certaine vers  $\sigma_v^2$  (voir, par exemple, Yuan et Jennrich 1998).

### Preuve du théorème 4.2

Nous désignons par  $\hat{\sigma}_{vML}^2$  l'estimateur de variance par le maximum de vraisemblance.

Nous montrons d'abord que  $\hat{\sigma}_{vREML}^2 - \hat{\sigma}_{vML}^2 = O_p(1/m)$ . Soit  $G_*(\sigma_v^2) = \partial \log(L_*) / \partial \sigma_v^2 = 0$  l'équation d'estimation qui donne l'estimateur de variance \*. L'équation (3.4) implique que :

$$G_{\text{AM,LL}}(\sigma_v^2) - G_{\text{ML}}(\sigma_v^2) = \partial \log \sigma_v^2 / \partial \sigma_v^2 = \frac{1}{m\sigma_v^2} = O\left(\frac{1}{m}\right). \quad (\text{A.3})$$

Avec  $G'_{\text{ML}}(\cdot) \triangleq (\partial G_{\text{ML}} / \partial \sigma_v^2)(\cdot)$  et  $G''_{\text{ML}}(\cdot) \triangleq (\partial G'_{\text{ML}} / \partial \sigma_v^2)(\cdot)$ , l'équation (A.3) implique que :

$$G'_{\text{ML}}(\sigma_v^2) - G'_{\text{AM,LL}}(\sigma_v^2) = O\left(\frac{1}{m}\right). \quad (\text{A.4})$$

Maintenant, en utilisant l'équation (A.4), la  $\sqrt{m}$ -convergence des estimateurs ML et AM,LL de  $\sigma_v^2$ , le développement en série de Taylor à deux termes de  $G_{\text{ML}}(\cdot)$  et  $G_{\text{AM,LL}}(\cdot)$  à  $\sigma_v^2$  et  $G'_{\text{ML}}(\sigma_v^2) = O(1)$  à mesure que  $m \rightarrow \infty$ , le côté gauche de l'équation en (A.3) est égal à :

$$\begin{aligned} &= G'_{\text{ML}}(\sigma_v^2)(\hat{\sigma}_{\text{vML}}^2 - \sigma_v^2) - G'_{\text{AM,LL}}(\sigma_v^2)(\hat{\sigma}_{\text{vAM,LL}}^2 - \sigma_v^2) + O_p\left(\frac{1}{m}\right) \\ &= G'_{\text{ML}}(\sigma_v^2)(\hat{\sigma}_{\text{vML}}^2 - \hat{\sigma}_{\text{vAM,LL}}^2) + (G'_{\text{ML}}(\sigma_v^2) - G'_{\text{AM,LL}}(\sigma_v^2))(\hat{\sigma}_{\text{vAM,LL}}^2 - \sigma_v^2) + O_p\left(\frac{1}{m}\right) \\ &= G'_{\text{ML}}(\sigma_v^2)(\hat{\sigma}_{\text{vML}}^2 - \hat{\sigma}_{\text{vAM,LL}}^2) + O_p\left(\frac{1}{m^{3/2}}\right) + O_p\left(\frac{1}{m}\right). \end{aligned}$$

La dernière égalité ci-dessus implique que :

$$\hat{\sigma}_{\text{vAM,LL}}^2 - \hat{\sigma}_{\text{vML}}^2 = O_p\left(\frac{1}{m}\right) \text{ à mesure que } m \rightarrow \infty. \quad (\text{A.5})$$

De même, nous établissons une relation entre  $G_{\text{REML}}(\sigma_v^2)$  et  $G_{\text{ML}}(\sigma_v^2)$  : étant donné que  $\text{tr}(\mathbf{V}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}^{-1}) = O(1)$  découle des conditions 1 à 3 de la section 3 et de l'équation (3.1), nous avons :

$$G_{\text{REML}}(\sigma_v^2) - G_{\text{ML}}(\sigma_v^2) = \frac{1}{m} \text{tr}(\mathbf{V}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}^{-1}) = O\left(\frac{1}{m}\right) \text{ à mesure que } m \rightarrow \infty, \quad (\text{A.6})$$

l'équation (A.6) et le même argument que pour l'estimateur AM,LL impliquent que :

$$\hat{\sigma}_{\text{vREML}}^2 - \hat{\sigma}_{\text{vML}}^2 = O_p\left(\frac{1}{m}\right) \text{ à mesure que } m \rightarrow \infty. \quad (\text{A.7})$$

Ensemble, les équations (A.5) et (A.7) donnent :

$$(\hat{\sigma}_{\text{vREML}}^2 - \hat{\sigma}_{\text{vAM,LL}}^2) = O_p\left(\frac{1}{m}\right). \quad (\text{A.8})$$

Nous exprimons maintenant le biais de l'estimateur MIX comme suit :

$$B_{\text{MIX}}(\hat{\sigma}_{\text{vMIX}}^2) = \int_{\{\hat{\sigma}_{\text{vREML}}^2 > 0\}} (\hat{\sigma}_{\text{vREML}}^2 - \sigma_v^2) dP + \int_{\{\hat{\sigma}_{\text{vREML}}^2 = 0\}} (\hat{\sigma}_{\text{vAM,LL}}^2 - \sigma_v^2) dP.$$

Nous ajoutons et soustrayons  $\int_{\{\hat{\sigma}_{vREML}^2=0\}} (\hat{\sigma}_{vREML}^2 - \sigma_v^2) dP$  du côté droit de l'équation ci-dessus pour obtenir :

$$\begin{aligned} B_{MIX}(\hat{\sigma}_{vMIX}^2) &= \int_{\Omega} (\hat{\sigma}_{vREML}^2 - \sigma_v^2) dP + \int_{\{\hat{\sigma}_{vREML}^2=0\}} (\hat{\sigma}_{vAM.LL}^2 - \hat{\sigma}_{vREML}^2) dP \\ &= \text{Biais}(\hat{\sigma}_{vREML}^2) + \int_{\{\hat{\sigma}_{vREML}^2=0\}} (\hat{\sigma}_{vAM.LL}^2 - \hat{\sigma}_{vREML}^2) dP. \end{aligned} \quad (\text{A.9})$$

Puisque  $\hat{\sigma}_{vAM.LL}^2 - \hat{\sigma}_{vREML}^2$  est uniformément borné, nous appliquons l'inégalité de Holder et Minkowski avec  $p = q = 2$  et l'équation (A.8) au dernier terme en (A.9) pour obtenir :

$$\begin{aligned} B_{MIX}(\hat{\sigma}_{vMIX}^2) &= \text{Biais}(\hat{\sigma}_{vREML}^2) + \left( \int_{\Omega} (\hat{\sigma}_{vAM.LL}^2 - \hat{\sigma}_{vREML}^2)^2 dP \right)^{1/2} \cdot P\{\hat{\sigma}_{vREML}^2 = 0\}^{1/2} \\ &= \text{Biais}(\hat{\sigma}_{vREML}^2) + O\left(\frac{1}{m}\right) \cdot o(1) = \text{Biais}(\hat{\sigma}_{vREML}^2) + o\left(\frac{1}{m}\right). \end{aligned} \quad (\text{A.10})$$

## Preuve de la remarque 4.2 : $eqm_0$ est sans biais jusqu'à l'ordre deux

$$\begin{aligned} E(eqm_0) - EQM(\hat{\theta}_i) &= \int_{\{\hat{\sigma}_{vREML}^2 > 0\}} (g_{1i} + g_{2i} + 2g_{3i})(\hat{\sigma}_{vREML}^2) dP + \int_{\{\hat{\sigma}_{vREML}^2 = 0\}} g_{2i}(\hat{\sigma}_{vREML}^2) dP - EQM \\ &= \left[ \int_{\Omega} (g_{1i} + g_{2i} + 2g_{3i})(\hat{\sigma}_{vREML}^2) dP - EQM \right] \\ &\quad + \int_{\{\hat{\sigma}_{vREML}^2 = 0\}} g_{2i}(\hat{\sigma}_{vREML}^2) dP - \int_{\{\hat{\sigma}_{vREML}^2 = 0\}} (g_{1i} + g_{2i} + 2g_{3i})(\hat{\sigma}_{vREML}^2) dP \\ &= \left[ o\left(\frac{1}{m}\right) \right] - \int_{\{\hat{\sigma}_{vREML}^2 = 0\}} 2g_{3i}(\hat{\sigma}_{vREML}^2) dP, \end{aligned} \quad (\text{A.11})$$

puisque  $g_{1i}(\hat{\sigma}_{vREML}^2) = g_{1i}(0) = 0$  en  $\{\hat{\sigma}_{vREML}^2 = 0\}$  et  $g_{2i}(\hat{\sigma}_{vREML}^2)$  s'annulent en (A.11). Mais

$$g_{3i}(\hat{\sigma}_{vREML}^2) = g_{3i}(0) = \frac{\bar{V}(0)}{\Psi_i} = O_p\left(\frac{1}{m}\right)$$

et est uniformément borné sous les conditions de régularité données dans la section 2. Le dernier terme en (A.11) est donc aussi un  $o(1/m)$ , ce qui rend  $eqm_0$  sans biais jusqu'à l'ordre deux.

## Annexe B

### B.1 Comparaison entre les estimateurs REML et AR.YL au moyen de l'algorithme de score

L'algorithme de score donnait parfois des estimations nulles pour la vraisemblance de l'estimateur AR.YL. En effet, pour les ensembles de données simulés sous le modèle donné dans la section 5, où  $m = 45$  et  $\sigma_v^2 = 1$ , les algorithmes de score REML et AR.YL produisaient 28 % et 26 % d'estimations nulles respectivement. On peut voir pourquoi dans les figures B.1 à B.3 : les vraisemblances correspondent à une seule population générée sous le modèle avec  $\sigma_v^2 = 1$  pour lequel  $\hat{\sigma}_{v\text{REML}}^2 = 0$ .

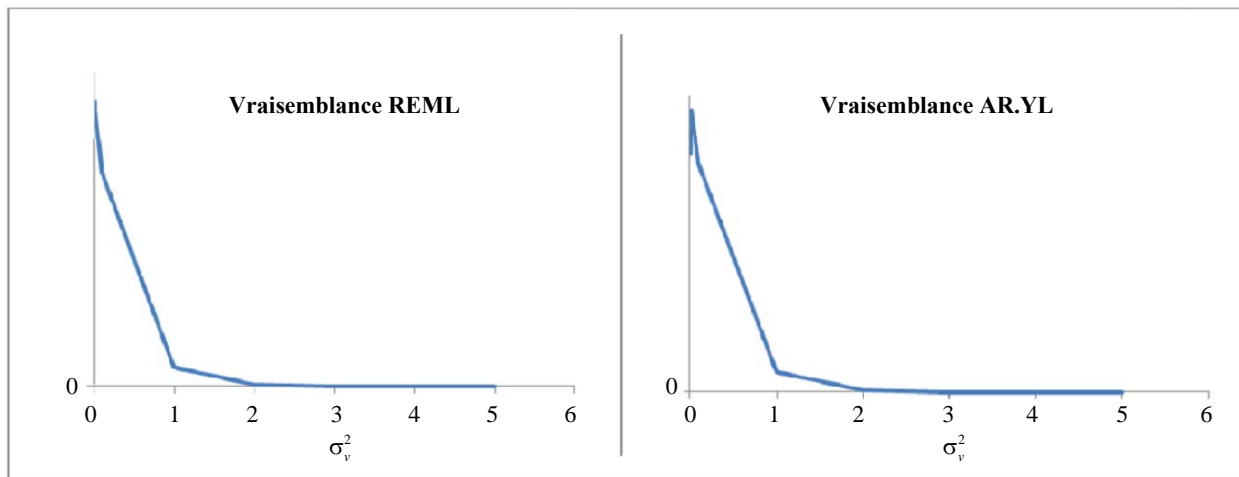


Figure B.1  $L = L_{\text{REML}}(\sigma_v^2 | y_1, \dots, y_{45})$ .

Figure B.2  $L = L_{\text{AR.YL}}(\sigma_v^2 | y_1, \dots, y_{45})$ .

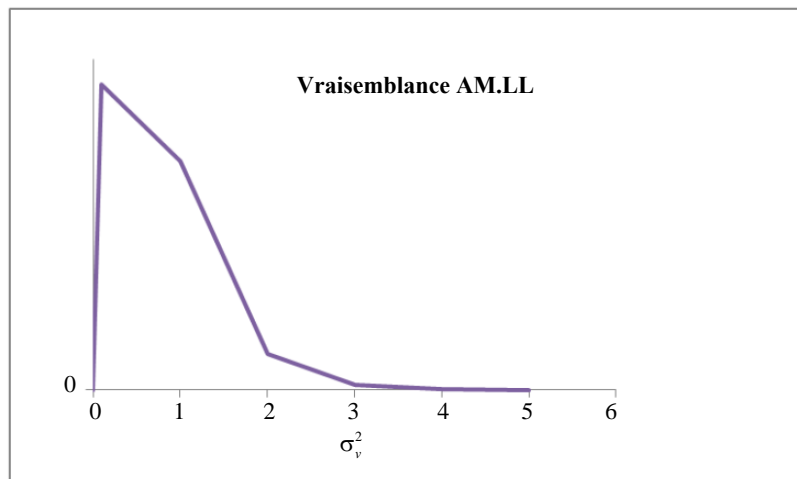


Figure B.3  $L = L_{\text{AM.LL}}(\sigma_v^2 | y_1, \dots, y_{45})$ .



La figure B.2 montre que la valeur maximale de la vraisemblance AR.YL est très proche de la limite. Il arrive parfois que l'algorithme de score passe à côté du maximum et donne une valeur nulle. La figure B.3 montre que la vraisemblance AM.LL a une valeur maximale qui se différencie mieux de la limite.

## B.2 Traitement des zéros dans l'estimateur bootstrap paramétrique

Pour chaque estimation  $\hat{\sigma}_v^2 = \hat{\sigma}_v^2(\mathbf{y}^{(r)})$ ,  $r = 1, \dots, 10K$ , et chaque méthode d'estimation de variance :

- i. Générer un grand nombre  $B$  d'effets aléatoires de domaine  $v_i^{(b)} \stackrel{\text{i.i.d.}}{\sim} N(0, \hat{\sigma}_v^2)$ ,  $b = 1, \dots, B$ , et générer, indépendamment de  $v_i^{(b)}$ , des erreurs d'échantillonnage  $e_i^{(b)} \stackrel{\text{i.i.d.}}{\sim} N(0, \psi_i)$ ,  $i = 1, \dots, m$ ,  $b = 1, \dots, B$ . Générer des données bootstrap  $y_i^{(b)} = \theta_i^{(b)} + e_i^{(b)}$ ,  $\theta_i^{(b)} = \mathbf{x}_i' \hat{\beta} + v_i^{(b)}$ ,  $i = 1, \dots, m$ . Si  $\hat{\sigma}_{\text{vREML}}^2(\mathbf{y}^{(b)}) = 0$ , générer  $(y_i^{(b)}, \theta_i^{(b)})$ ,  $b = 1, \dots, B$ , à partir du modèle synthétique (voir aussi Rao et Molina 2015).
- ii. Adapter le modèle aux données bootstrap et obtenir  $\hat{\sigma}_v^{2(b)}$ ; pour l'estimateur MIX, calculer  $\hat{\sigma}_{\text{vMIX}}^{2(b)} = \hat{\sigma}_{\text{vREML}}^{2(b)}$  si  $\hat{\sigma}_v^{2(b)}$  est positif et  $\hat{\sigma}_{\text{vMIX}}^{2(b)} = \hat{\sigma}_{\text{vAM}}^{2(b)}$  autrement.
- iii. Obtenir  $\hat{\beta}^{(b)}$ , l'EBLUP correspondant  $\hat{\theta}_i^{(b)}$ , les composantes bootstrap  $g_{1i}^{(b)} = g_{1i}(\hat{\sigma}_v^{2(b)})$ ,  $g_{2i}^{(b)} = g_{2i}(\hat{\sigma}_v^{2(b)})$  et  $\bar{g}_{ji}^{\text{BP}} = B^{-1} \sum_b g_{ji}^{(b)}$ ,  $j = 1, 2$ .
- iv. L'estimateur bootstrap naïf de l'EQM est  $\text{eqm}_{\text{naive}} = B^{-1} \sum_{b=1}^B (\hat{\theta}_i^{(b)} - \theta_i^{(b)})^2$ .
- v. L'estimateur BP de l'EQM (qui est corrigé du biais (Pfeffermann et Glickman 2004)) est :  $\text{eqm}_{\text{BP}}(\hat{\theta}_i) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) - \bar{g}_{1i}^{\text{BP}} - \bar{g}_{2i}^{\text{BP}} + \text{eqm}_{\text{naive}}$ .
- vi. Pour calculer le  $\text{BRM}_C$ , faire la moyenne de  $(\text{eqm}_{\text{BP}}^{(r)}(\hat{\theta}_i) - \text{EQM}(\hat{\theta}_i)) / \text{EQM}(\hat{\theta}_i)$  pour les populations où  $(r) / \hat{\sigma}_{\text{vREML}}^2(\mathbf{y}^{(r)}) = 0$  et faire de même pour le  $\text{BRM}_C$  de  $\text{eqm}_{\text{naive}}$ .

## Bibliographie

- Chen, S., et Lahiri, P. (2008). On mean squared prediction error estimation in small area estimation problems. *Communications in Statistics-Theory and Methods*, 37, 1792-1798.
- Chen, S., et Lahiri, P. (2011). On the estimation of Mean Squared Prediction Error in small area estimation. *Calcutta Statistical Association Bulletin*, 63, (Special 7<sup>th</sup> Triennial Proceedings Volume), Nos. 249-252.
- Cressie, N. (1992). Estimation du maximum de vraisemblance avec contrainte (MVC) dans le lissage des taux de sous-dénombrement du recensement selon l'approche empirique de Baye. *Techniques d'enquête*, 18, 1, 83-103.
- Das, K., Jiang, J. et Rao, J.N.K. (2004). Mean squared error of empirical predictor. *The Annals of Statistics*, 32, 2, 818-840.
- Datta, G., 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.

- Estevao, V. (2014). Grid optimization algorithm for maximum likelihood. Rapport interne, Division de la recherche et de l'innovation en statistique (DRIS), Statistique Canada.
- Fay, R.E., et Herriot, R.A. (1979). Estimation of income from small places: An application of James-Stein Procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.
- Lahiri, P., et Li, H. (2009). Generalized maximum likelihood method in linear mixed models with an application in small area estimation. Dans *Proceedings of the Federal Committee on Statistical Methodology Research Conference*, disponible au <http://www.fcsm.gov/events/papers2009.html>.
- Lahiri, P., et Pramanik, S. (2011). Discussion of "Estimating random effects via adjustment for density maximization" par C. Morris et R. Tang. *Statistical Science*, 26, 2, 291-295.
- Li, H., et Lahiri, P. (2011). An adjusted maximum likelihood method for solving small area estimation problems. *Journal of Multivariate Analysis*, 101, 882-892.
- Molina, I., Rao, J.N.K. et Datta, G.S. (2015). Estimation sur petits domaines sous un modèle de Fay-Herriot avec test préliminaire pour la présence d'effets aléatoires de domaine. *Techniques d'enquête*, 41, 1, 1-20.
- Morris, C.N. (2006). Mixed model prediction and small area estimation (with discussions). *Test*, 15, 72-76.
- Pfeffermann, D., et Glickman, H. (2004). Mean squared error approximation in small area estimation by use of parametric and non-parametric bootstrap. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, Alexandria, VA. 4167-78.
- Rao, J.N.K. (2003). *Small Area Estimation*. New York : John Wiley & Sons, Inc.
- Rao, J.N.K., et Molina, I. (2015). *Small Area Estimation, second edition*. New York : John Wiley & Sons, Inc.
- Rubin-Bleuer, S., et Schiopu-Kratina, I. (2005). On the two-phase framework for joint model and design-based inference. *The Annals of Statistics*, 33, 6, 2789-2810.
- Rubin-Bleuer, S., et You, Y. (2012). A positive variance estimator for the Fay-Herriot small area model. SRID-2012-009E, Division de la recherche et de l'innovation en statistique (DRIS), Statistique Canada.
- Rubin-Bleuer, S., Yung, W. et Landry, S. (2010). Adjusted maximum likelihood method for a small area model accounting for time and area effects. SRID-2010-006E, Division de la recherche et de l'innovation en statistique (DRIS), Statistique Canada.
- Rubin-Bleuer, S., Yung, W. et Landry, S. (2011). Adjusted maximum likelihood method for a small area model accounting for time and area effects. Long abstract, *Small Area Estimation*, (SAE 20122) à Trèves, Allemagne, International Statistical Institute Satellite Conference.
- Rubin-Bleuer, S., Yung, W. et Landry, S. (2012). Variance Component Estimation through the Adjusted Maximum Likelihood Approach. Exposé donné à la conférence en l'honneur du 75<sup>e</sup> anniversaire de J.N.K. Rao, Université Carleton, mai 2012, Ottawa.
- Yoshimori, M., et Lahiri, P. (2014). A new adjusted maximum likelihood method for the Fay-Herriot small area model. *Journal of Multivariate Analysis*, 124, 281-294.

Yuan, P. (2009). Comparison of SAE methods of variance estimation. Document interne, Division de la recherche et de l'innovation en statistique (DRIS), Statistique Canada.

Yuan, K.H., et Jennrich, R. (1998). Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis*, 65, 2, 245-260.