

Techniques d'enquête

Estimation sur petits domaines réconciliée sous le modèle de base au niveau de l'unité lorsque les taux d'échantillonnage sont non négligeables

par Marius Stefan et Michael A. Hidirolou

Date de diffusion : le 24 juin 2021



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.

Vous pouvez également communiquer avec nous par :

Courriel à STATCAN.infostats-infostats.STATCAN@canada.ca

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

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-514-283-9350 |

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 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.

Publication autorisée par le ministre responsable de Statistique Canada

© Sa Majesté la Reine du chef du Canada, représentée par le ministre de l'Industrie 2021

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.

Estimation sur petits domaines réconciliée sous le modèle de base au niveau de l'unité lorsque les taux d'échantillonnage sont non négligeables

Marius Stefan et Michael A. Hidirolou¹

Résumé

Nous examinons l'estimation d'une moyenne sur petits domaines sous le modèle de base au niveau de l'unité. La somme des estimateurs dépendant d'un modèle qui en résultent peut ne pas correspondre aux estimations obtenues au moyen d'un estimateur d'enquête direct qui est considéré comme précis pour l'ensemble de ces petits domaines. La réconciliation force la concordance des estimateurs fondés sur un modèle avec l'estimateur direct au niveau du domaine agrégé. L'estimateur par la régression généralisée est l'estimateur direct que nous utilisons pour réaliser la réconciliation. Dans le présent document, nous comparons des estimateurs sur petits domaines réconciliés d'après quatre procédures. La première procédure permet d'obtenir des estimateurs réconciliés au moyen d'un ajustement par le ratio. La deuxième procédure repose sur le meilleur estimateur linéaire sans biais empirique obtenu sous le modèle au niveau de l'unité augmenté à l'aide d'une variable adéquate qui assure la réconciliation. La troisième procédure utilise des estimateurs pseudo-empiriques construits au moyen de poids de sondage convenablement choisis de sorte que, une fois agrégés, ils concordent avec l'estimateur direct fiable pour le plus grand domaine. La quatrième procédure permet d'obtenir des estimateurs réconciliés qui résultent d'un problème de minimisation sous la contrainte donnée par la condition de réconciliation. Ces procédures de réconciliation sont appliquées aux estimateurs sur petits domaines lorsque les taux d'échantillonnage sont non négligeables. Les estimateurs réconciliés qui en résultent sont comparés quant au biais relatif et à l'erreur quadratique moyenne dans une étude par simulations fondée sur un plan de sondage ainsi qu'un exemple fondé sur des données d'enquête réelles.

Mots-clés : Petit domaine; réconciliation; estimateur empirique; estimateur pseudo-empirique; estimateur contraint.

1. Introduction

L'estimation sur petits domaines (EPD) a gagné en importance au cours des dernières années en raison de la demande pour des statistiques fiables sur les petits domaines. Des estimateurs directs sont utilisés pour estimer des paramètres d'intérêt lorsque la taille de l'échantillon est suffisamment grande. Cependant, ils présentent des erreurs-types et des coefficients de variation importants lorsque vient le temps de les appliquer aux petits domaines, car la taille d'échantillon réalisée sera assez limitée. Il est par conséquent nécessaire d'utiliser des modèles qui empruntent de l'information à d'autres domaines adjacents ou à des enquêtes précédentes pour obtenir des estimateurs stables pour ces petits domaines. Les estimations fondées sur un modèle présentent généralement une amélioration considérable par rapport aux estimations directes quant à l'erreur quadratique moyenne (EQM).

La théorie disponible pour l'estimation sur petits domaines repose sur des modèles au niveau du domaine ou au niveau de l'unité, selon le niveau des données auxiliaires disponibles. Les méthodes fondées sur des modèles au niveau de l'unité utilisent les données de chacune des unités comme données auxiliaires, tandis que les méthodes fondées sur des modèles au niveau du domaine utilisent des agrégats ou des moyennes des données sur les unités dans les petits domaines. Le modèle de Fay et Herriot (1979),

1. Marius Stefan, Faculty of Applied Sciences, Polytechnic University of Bucharest, Splaiul Independentei, nr. 313. Courriel : mastefan@gmail.com; Michael A. Hidirolou, ancien employé de Statistique Canada. Courriel : hidirolou@yahoo.ca.

désigné ci-après le modèle FH, est le modèle au niveau du domaine le plus utilisé dans l'estimation sur petits domaines. Le modèle de régression à erreurs emboîtées simple proposé dans Battese, Harter et Fuller (1988), également appelé le modèle de base au niveau de l'unité, est fréquemment utilisé lorsque des données sont disponibles au niveau de l'unité. Nous désignons ce modèle le modèle BHF. Il s'agit de deux cas particuliers d'un modèle linéaire mixte général dans l'EPD (voir Rao et Molina, 2015 pour consulter un excellent exposé sur l'estimation sur petits domaines).

Les moyennes ou les totaux de petits domaines sont les paramètres linéaires les plus fréquemment estimés dans l'EPD. En pareil cas, la méthode d'estimation sur petits domaines la plus populaire est l'utilisation de modèles linéaires mixtes afin de calculer les meilleurs prédicteurs linéaires sans biais (BLUP) pour la moyenne ou le total de petits domaines. Les estimateurs BLUP minimisent l'EQM dans la catégorie des estimateurs linéaires sans biais. Par ailleurs, on peut démontrer que l'estimateur BLUP peut être obtenu par la résolution des équations d'un modèle mixte comportant des inconnues données par les paramètres fixes et aléatoires du modèle. Les équations du modèle mixte résultent de la maximisation de la densité conjointe des données et du vecteur des effets aléatoires de petit domaine. Un estimateur BLUP dépend des variances (et des covariances) des effets aléatoires qu'on peut estimer à l'aide de la méthode d'ajustement de constantes (FC) d'Henderson, du maximum de vraisemblance (MV) ou du maximum de vraisemblance restreint (MVRE). L'utilisation de ces composantes estimées dans l'estimateur BLUP permet d'obtenir un estimateur à deux degrés appelé le meilleur prédicteur linéaire sans biais empirique (EBLUP).

Une difficulté potentielle avec les estimateurs EBLUP est que, lorsqu'ils sont agrégés pour l'ensemble des petits domaines, ils peuvent ne pas correspondre à l'estimation globale pour un plus grand domaine obtenue au moyen d'une estimation directe. Les organismes statistiques favorisent la correspondance globale entre la somme des estimations sur petits domaines fondées sur un modèle et l'estimation directe à un niveau supérieur qui correspond à l'agrégation des petits domaines. La réconciliation est une méthode permettant de modifier les estimations fondées sur un modèle pour qu'elles correspondent à l'estimateur direct pour le plus grand domaine.

Les méthodes de réconciliation actuelles sont fréquentistes ou bayésiennes. Dans le présent document, nous nous attardons à l'approche de réconciliation fréquentiste (pour connaître les procédures de réconciliation bayésiennes, voir You, Rao et Dick, 2004; Datta, Ghosh, Steorts et Maples, 2011 et Nandram et Sayit, 2011). On peut appliquer les méthodes fréquentistes afin d'obtenir des estimations sur petits domaines réconciliées pour les modèles tant au niveau du domaine et qu'au niveau de l'unité.

Nous passons brièvement en revue la documentation actuelle sur les deux types de modèle. Nous décrivons en premier lieu les procédures élaborées pour réconcilier les estimations au niveau du domaine. Pfeiffermann et Barnard (1991) ont obtenu un estimateur réconcilié contraint en maximisant la densité conjointe des données et le vecteur des effets aléatoires de petit domaine compte tenu de la restriction de réconciliation. Leur estimateur réconcilié a été construit à l'aide d'estimations ajustées des effets fixes et de petit domaine qui solutionnent le problème de maximisation sous contrainte. Wang, Fuller et Qu (2008)

ont élaboré un estimateur EBLUP réconcilié pour le modèle FH au niveau du domaine, en minimisant une fonction de perte sous la contrainte donnée par la condition de réconciliation. Ils ont obtenu un second estimateur réconcilié en ajoutant une variable auxiliaire adéquate au modèle FH sans imposer de contrainte. Ils ont démontré que l'estimateur EBLUP fondé sur le modèle FH augmenté est autoréconcilié : l'estimateur a satisfait la condition de réconciliation sans autres rajustements. Bell, Datta et Ghosh (2013) ont généralisé les résultats présentés dans Wang et coll. (2008) au cas de multiples contraintes de réconciliation en considérant une fonction de perte plus générale. You, Rao et Hidirolou (2013) ont obtenu un autre estimateur autoréconcilié sous le modèle FH en remplaçant le vecteur de régression utilisé dans l'estimateur EBLUP par un autre estimateur qui dépend des poids de réconciliation.

Nous décrivons maintenant les procédures qui permettent de réconcilier les estimations fondées sur un modèle au niveau de l'unité. L'objectif est d'obtenir des estimateurs sur petits domaines qui sont réconciliés à un estimateur direct à un niveau d'agrégation donné des petits domaines. Les estimateurs directs principalement utilisés par les organismes statistiques sont l'estimateur par la régression généralisée (GREG) décrit dans Särndal, Swensson et Wretman (1989) ou plus généralement l'estimateur par calage fondé sur les procédures décrites dans Deville et Särndal (1992). You et Rao (2002) ont élaboré un prédictor pseudo-EBLUP (prédictor YR) qui intègre des poids de sondage. Une propriété de cet estimateur est qu'il est autoréconcilié, c'est-à-dire que la somme des estimations sur petits domaines correspond à un estimateur ayant la même forme que l'estimateur GREG. Cependant, il ne s'agit pas d'un estimateur direct, car le vecteur de régression estimé faisant partie de cet estimateur témoigne de la structure des erreurs du modèle à erreurs emboîtées. En supposant que les taux d'échantillonnage sont négligeables, Stefan et Hidirolou (2020) ont proposé plusieurs procédures pour assurer que les estimateurs EBLUP et pseudo-EBLUP soient réconciliés à l'estimateur GREG, étant donné que le modèle et l'estimateur GREG utilisaient tous deux le même vecteur de variables auxiliaires. Ugarte, Militino et Goicoa (2009) ont élaboré un estimateur EBLUP restreint pour un total de petits domaines satisfaisant la propriété de réconciliation à un estimateur synthétique.

Le présent document a pour objectif de comparer plusieurs estimateurs réconciliés d'une moyenne de petits domaines pour le modèle de base au niveau de l'unité lorsque les taux d'échantillonnage sont non négligeables. Nous comparons six estimateurs réconciliés : deux estimateurs réconciliés fondés sur les procédures proposées par Stefan et Hidirolou (2020), deux estimateurs restreints fondés sur la procédure proposée par Ugarte et coll. (2009) et deux estimateurs par le ratio obtenus en multipliant les estimateurs EBLUP et YR de chaque petit domaine par un facteur d'ajustement commun. Le document est structuré de la manière suivante. La section 2 présente un résumé des estimateurs EBLUP et pseudo-EBLUP sous le modèle de base au niveau de l'unité. La section 3 décrit les six estimateurs réconciliés. Les deux premiers estimateurs sont fondés sur des ajustements par le ratio simples. Ensuite, nous démontrons comment on peut adapter les deux procédures de réconciliation proposées par Stefan et Hidirolou (2020) dans le cas de taux d'échantillonnage négligeables pour produire des estimateurs de la moyenne de petits domaines réconciliés lorsque les taux d'échantillonnage sont non négligeables. Enfin, nous décrivons l'estimateur

EBLUP restreint d'Ugarte et coll. (2009) et nous proposons un pseudo-estimateur restreint qui est une variante de l'estimateur EBLUP restreint intégrant des poids de sondage. Nous proposons également une méthode utilisant le maximum de vraisemblance restreint reparamétré (MVRere) pour estimer les composantes de la variance. Cette méthode d'estimation est utile pour calculer des estimateurs EBLUP restreints de la moyenne de petits domaines, car elle permet d'obtenir des estimations des composantes de la variance strictement positives. La section 4 présente les résultats d'une simulation de Monte-Carlo fondée sur des ensembles de données générés, tandis que la section 5 présente les résultats d'une étude par simulations fondée sur un ensemble de données réel. Enfin, la section 6 présente quelques observations finales.

2. Estimation EBLUP et pseudo-EBLUP

Prenons le modèle de régression à erreurs emboîtées simple

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}, \quad i = 1, \dots, m; \quad j = 1, \dots, N_i, \quad (2.1)$$

où y_{ij} est la variable d'intérêt pour la j^e unité de population dans le i^e petit domaine, $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ est un vecteur de variables auxiliaires avec $x_{ij1} = 1$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ est un vecteur de paramètres de régression $p \times 1$ et N_i est le nombre d'unités de population dans le i^e petit domaine, U_i . Les effets aléatoires de petit domaine v_i sont présumés indépendants et identiquement distribués (i.i.d.) $N(0, \sigma_v^2)$, et indépendants des erreurs au niveau de l'unité e_{ij} , qui sont présumées i.i.d. $N(0, \sigma_e^2)$. Nous prélevons des échantillons s_i de taille n_i de manière indépendante dans chaque petit domaine i , selon un plan d'échantillonnage spécifié, les probabilités d'inclusion de premier ordre étant désignées π_{ij} , pour $j = 1, \dots, N_i$. La taille totale de l'échantillon est n , où $n = \sum_{i=1}^m n_i$. Les poids de sondage de base qui en résultent sont donnés par $d_{ij} = 1/\pi_{ij}$. Nous supposons que le plan d'échantillonnage est ignorable et qu'il n'y a pas de biais de sélection. Cela suppose que le modèle (2.1) vaut également pour les données de l'échantillon :

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}, \quad i = 1, \dots, m; \quad j = 1, \dots, n_i, \quad (2.2)$$

Le modèle (2.2) est un cas particulier du modèle linéaire mixte général. En définissant $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$, $\mathbf{X}_i = (\mathbf{x}_{i1}^T, \dots, \mathbf{x}_{in_i}^T)^T$, $\mathbf{v} = (v_1, \dots, v_m)^T$ et $\mathbf{e}_i = (e_{i1}, \dots, e_{in_i})^T$, il s'ensuit que le modèle (2.2) peut s'exprimer sous une forme matricielle par empilement des observations. L'équation qui en résulte est

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \quad (2.3)$$

où $\mathbf{y} = \text{col}_{1 \leq i \leq m}(\mathbf{y}_i)$, $\mathbf{X} = \text{col}_{1 \leq i \leq m}(\mathbf{X}_i)$, $\mathbf{Z} = \text{diag}_{1 \leq i \leq m} \{\mathbf{1}_{n_i}\}$ et $\mathbf{e} = \text{col}_{1 \leq i \leq m}(\mathbf{e}_i)$, avec $\mathbf{1}_{n_i}$ un vecteur de dimension n_i composé de uns. Nous désignons $\mathbf{G}$ et $\mathbf{R}$ les matrices de variance des vecteurs

aléatoires $\mathbf{v}$ et $\mathbf{e}$ respectivement. Alors $\mathbf{G} = \sigma_v^2 \mathbf{I}_m$ et $\mathbf{R} = \sigma_e^2 \mathbf{I}_n$. Il s'ensuit que la matrice de variance du vecteur $\mathbf{y}$, désignée $\mathbf{V}$, est donnée par $\mathbf{V} = \mathbf{R} + \mathbf{ZGZ}^T$.

Les paramètres d'intérêt sont les moyennes de petits domaines $\bar{Y}_i$, où $\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}$, $i = 1, \dots, m$. Si N_i est grand, la fraction d'échantillonnage $f_i = N_i^{-1} n_i$ du i^e petit domaine est négligeable. Cette configuration correspond au cas d'une *population infinie* ou à des taux d'échantillonnage négligeables. Il s'ensuit que les moyennes de petits domaines $\bar{Y}_i$ peuvent être approximées par μ_i (voir Rao et Molina, 2015, page 174), où $\mu_i = \bar{\mathbf{X}}_i^T \boldsymbol{\beta} + v_i$ et $\bar{\mathbf{X}}_i = \sum_{j=1}^{N_i} \mathbf{x}_{ij} / N_i$ est le vecteur des moyennes de population des $\mathbf{x}_{ij}$ pour le i^e domaine. Un estimateur de μ_i est donné par $\hat{\mu}_i = \bar{\mathbf{X}}_i^T \hat{\boldsymbol{\beta}} + \hat{v}_i$ (Rao et Molina, 2015, page 175), où $\hat{\boldsymbol{\beta}}$ et $\hat{v}_i$ sont des estimateurs de $\boldsymbol{\beta}$ et v_i respectivement. Si N_i n'est pas suffisamment grand ou si les taux d'échantillonnage f_i ne sont pas négligeables, les paramètres $\bar{Y}_i$ ne peuvent être approximés par des combinaisons linéaires de $\boldsymbol{\beta}$ et v_i . Cela correspond au cas d'une *population finie*. Soit r_i l'ensemble des valeurs y non observées $N_i - n_i$ dans le petit domaine i . Si nous supposons que nous connaissons les $\mathbf{x}_{ij}$ pour chaque personne dans la population, un estimateur $\hat{\bar{Y}}_i$ de $\bar{Y}_i$ est fondé sur les valeurs observées de y_{ij} , $j \in s_i$ et les valeurs prédites de $\hat{y}_{ij} = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + \hat{v}_i$ pour $j \in r_i$. Autrement dit, l'estimateur $\hat{\bar{Y}}_i$ est donné par

$$\hat{\bar{Y}}_i = \frac{1}{N_i} \left(\sum_{j \in s_i} y_{ij} + \sum_{j \in r_i} \hat{y}_{ij} \right). \quad (2.4)$$

Une grande partie de la théorie sur l'EPD concerne le cas d'une population infinie, alors que la documentation sur le cas d'une population finie est plus limitée. Dans le présent document, nous nous attardons au cas d'une population finie (ou de taux d'échantillonnage non négligeables) et nous construisons par conséquent des estimateurs fondés sur l'équation (2.4).

2.1 Estimation EBLUP

Nous désignons $\tilde{\boldsymbol{\beta}}$ et $\tilde{\mathbf{v}} = (\tilde{v}_1, \dots, \tilde{v}_m)^T$ les prédicteurs EBLUP de $\boldsymbol{\beta}$ et de $\mathbf{v}$ respectivement. Ces estimateurs sont donnés par $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$ et $\tilde{\mathbf{v}} = \mathbf{GZ}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})$. Dans l'hypothèse de la normalité de $\mathbf{e}$ et de $\mathbf{v}$, on peut démontrer que $\tilde{\boldsymbol{\beta}}$ et $\tilde{\mathbf{v}}$ peuvent être obtenus en maximisant la densité conjointe de $\mathbf{y}$ et de $\mathbf{v}$ en ce qui concerne $\boldsymbol{\beta}$ et $\mathbf{v}$. Ceci équivaut à minimiser la fonction

$$\phi = (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{Z}\tilde{\mathbf{v}})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{Z}\tilde{\mathbf{v}}) + \tilde{\mathbf{v}}^T \mathbf{G}^{-1} \tilde{\mathbf{v}}. \quad (2.5)$$

Il en résulte les équations d'un modèle mixte suivantes

$$\mathbf{A} \begin{pmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\mathbf{v}} \end{pmatrix} = \mathbf{b}, \quad (2.6)$$

où

$$\mathbf{A} = \begin{pmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \quad \text{et} \quad \mathbf{b} = \begin{pmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{y} \end{pmatrix}. \quad (2.7)$$

(voir Rao et Molina, 2015, page 99 pour plus de détails). Les composantes de la variance (σ_v^2, σ_e^2) dans les équations (2.6) et (2.7) sont généralement inconnues. Trois méthodes d'estimation, la méthode FC, le MV et le MVRE, sont couramment utilisées dans l'EPD pour estimer les composantes de la variance (σ_v^2, σ_e^2) . Une difficulté bien connue avec ces méthodes est que l'estimation de σ_v^2 peut prendre des valeurs négatives. En pareil cas, cette estimation est tronquée à zéro, c'est-à-dire que $\hat{\sigma}_v^2$ reçoit la valeur 0. Des versions empiriques de $\mathbf{A}$ et de $\mathbf{b}$, désignées $\hat{\mathbf{A}}$ et $\hat{\mathbf{b}}$, sont obtenues si les composantes inconnues de la variance (σ_v^2, σ_e^2) sont remplacées par des estimateurs $(\hat{\sigma}_v^2, \hat{\sigma}_e^2)$. Il découle de l'équation (2.6) que des estimateurs EBLUP des paramètres du modèle $(\boldsymbol{\beta}, \mathbf{v})$, désignés $\hat{\boldsymbol{\beta}}$ et $\hat{\mathbf{v}} = (\hat{v}_1, \dots, \hat{v}_m)^T$, sont donnés par

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{v}} \end{pmatrix} = \hat{\mathbf{A}}^{-1} \hat{\mathbf{b}}. \quad (2.8)$$

Au moyen de l'équation (2.8), il peut être démontré que $\hat{\boldsymbol{\beta}}$ et $\hat{\mathbf{v}}$ sont

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{v}} \end{pmatrix} = \begin{pmatrix} (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{y} \\ \hat{\mathbf{G}} \mathbf{Z}^T \hat{\mathbf{V}}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) \end{pmatrix}, \quad (2.9)$$

où $\hat{\mathbf{G}} = \hat{\sigma}_v^2 \mathbf{I}_m$ et $\hat{\mathbf{V}} = \hat{\sigma}_e^2 \mathbf{I}_n + \hat{\sigma}_v^2 \mathbf{Z} \mathbf{Z}^T$.

Remarque 1. Il est plus facile d'inverser les matrices $\hat{\mathbf{G}} = \hat{\sigma}_v^2 \mathbf{I}_m$ et $\hat{\mathbf{R}} = \hat{\sigma}_e^2 \mathbf{I}_n$ que $\hat{\mathbf{V}}$. En conséquence, il est plus simple d'utiliser les équations du modèle mixte (2.8) que les équations (2.9) pour calculer $\hat{\boldsymbol{\beta}}$ et $\hat{\mathbf{v}}$. Cependant, lorsque $\hat{\sigma}_v^2$ est égal à zéro, les équations (2.8) ne peuvent être utilisées, car le terme $\hat{\mathbf{G}}^{-1}$ dans la matrice $\hat{\mathbf{A}}$ n'existe pas. En pareil cas, $\hat{\boldsymbol{\beta}}$ et $\hat{\mathbf{v}}$ ne peuvent être calculés qu'avec les équations (2.9).

Dans le modèle (2.2), on peut démontrer que $\hat{\boldsymbol{\beta}}$ et $\hat{v}_i$ dans $\hat{\mathbf{v}} = (\hat{v}_1, \dots, \hat{v}_m)^T$ satisfont l'équation

$$\sum_{i=1}^m \sum_{j \in s_i} \mathbf{x}_{ij} (y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} - \hat{v}_i) = 0. \quad (2.10)$$

Des estimateurs $\hat{\boldsymbol{\beta}}$ et $\hat{v}_i$ sont utilisés pour calculer des prédictors EBLUP $\hat{y}_{ij}^{\text{EBLUP}}$ pour les unités non observées $N_i - n_i$ dans le petit domaine i : $\hat{y}_{ij}^{\text{EBLUP}} = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + \hat{v}_i$ pour $j \in r_i$. Un estimateur EBLUP de $\bar{Y}_i$, désigné $\hat{\bar{Y}}_i^{\text{EBLUP}}$, est obtenu en remplaçant $\hat{y}_{ij}$ par $\hat{y}_{ij}^{\text{EBLUP}}$ dans l'équation (2.4). Il s'ensuit que $\hat{\bar{Y}}_i^{\text{EBLUP}}$ est

$$\hat{\bar{Y}}_i^{\text{EBLUP}} = \frac{1}{N_i} \left[\sum_{j \in s_i} y_{ij} + \mathbf{x}_{ir}^T \hat{\boldsymbol{\beta}} + (N_i - n_i) \hat{v}_i \right], \quad (2.11)$$

où $\mathbf{x}_{ir} = \sum_{j \in r_i} \mathbf{x}_{ij}$ représente la somme des valeurs non échantillonnées $\mathbf{x}_{ij}$.

2.2 Estimation de You-Rao

You et Rao (2002) ont proposé un estimateur de la moyenne de petits domaines pseudo-EBLUP (estimateur YR) qui intègre les poids de sondage d_{ij} à la formule de l'estimateur EBLUP. Une propriété

de l'estimateur pseudo-EBLUP est que la cohérence du plan est préservée lorsque la taille d'échantillon dans le domaine augmente. De plus, le prédicteur YR protège contre l'échec du modèle ou un plan d'échantillonnage informatif (voir entre autres Hidiroglou et Estevao, 2016 et Verret, Rao et Hidiroglou, 2015 pour plus de détails). On peut construire des estimateurs pseudo-EBLUP en utilisant la procédure décrite dans You et Rao (2002) avec des poids de sondage w_{ij} qui peuvent être calés sur un vecteur de variables auxiliaires. Soient $\hat{\boldsymbol{\beta}}^{\text{YR}}$ et $\hat{\mathbf{v}}^{\text{YR}} = (\hat{v}_1^{\text{YR}}, \dots, \hat{v}_m^{\text{YR}})^T$ les estimateurs YR de $\boldsymbol{\beta}$ et $\mathbf{v}$ respectivement fondés sur des poids w_{ij} (voir You et Rao, 2002 pour plus de détails). Les estimateurs $\hat{\boldsymbol{\beta}}^{\text{YR}}$ et $\hat{v}_i^{\text{YR}}$ satisfont les équations d'estimation au niveau de l'unité

$$\sum_{i=1}^m \sum_{j \in s_i} w_{ij} \mathbf{x}_{ij} (y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}^{\text{YR}} - \hat{v}_i^{\text{YR}}) = 0. \quad (2.12)$$

Les équations (2.12) représentent la version pondérée par les poids de sondage des équations (2.10). Les prédicteurs de You-Rao $\hat{y}_{ij}^{\text{YR}}$ de y_{ij} sont calculés comme $\hat{y}_{ij}^{\text{YR}} = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}^{\text{YR}} + \hat{v}_i^{\text{YR}}$ pour $j \in r_i$. Le remplacement de $\hat{y}_{ij}$ par $\hat{y}_{ij}^{\text{YR}}$ dans l'équation (2.4) permet d'obtenir l'estimateur YR de $\bar{Y}_i$ dans le cas de taux d'échantillonnage non négligeables :

$$\hat{\bar{Y}}_i^{\text{YR}} = \frac{1}{N_i} \left[\sum_{j \in s_i} y_{ij} + \mathbf{x}_{ir}^T \hat{\boldsymbol{\beta}}^{\text{YR}} + (N_i - n_i) \hat{v}_i^{\text{YR}} \right]. \quad (2.13)$$

On peut obtenir autrement les estimateurs $\hat{\boldsymbol{\beta}}^{\text{YR}}$ et $\hat{\mathbf{v}}^{\text{YR}}$ comme solutions à des équations d'un modèle mixte pondéré semblables à l'équation (2.6) (voir Huang et Hidiroglou, 2003 pour plus de détails). À cette fin, nous définissons les matrices $\mathbf{W}_i = \text{diag}_{1 \leq j \leq n_i} \{w_{ij}\}$, $\mathbf{W} = \text{diag}_{1 \leq i \leq m} \{\mathbf{W}_i\}$ et $\boldsymbol{\Omega} = \text{diag}_{1 \leq i \leq m} \{\omega_i\}$, où $\omega_i = \sum_{j \in s_i} w_{ij}^2 / \sum_{j \in s_i} w_{ij}$ pour $i = 1, \dots, m$. Soit ϕ_w la version pondérée de l'échantillon ϕ , où

$$\phi_w = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{v})^T \mathbf{W}^{1/2} \mathbf{R}^{-1} \mathbf{W}^{1/2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{v}) + \mathbf{v}^T \boldsymbol{\Omega}^{1/2} \mathbf{G}^{-1} \boldsymbol{\Omega}^{1/2} \mathbf{v}, \quad (2.14)$$

$\mathbf{W}^{1/2}$ et $\boldsymbol{\Omega}^{1/2}$ représentant la racine carrée des matrices $\mathbf{W}$ et $\boldsymbol{\Omega}$ respectivement. Dans le premier terme de ϕ_w , l'erreur du modèle associée à l'observation y_{ij} est pondérée par le poids de sondage correspondant w_{ij} , tandis que dans le second terme de ϕ_w , le facteur ω_i dans la matrice diagonale $\boldsymbol{\Omega}$ représente le poids lié à l'effet de petit domaine v_i . On peut démontrer que la minimisation de ϕ_w en ce qui concerne $\boldsymbol{\beta}$ et $\mathbf{v}$ permet d'obtenir $(\hat{\boldsymbol{\beta}}^{\text{YR}}, \hat{\mathbf{v}}^{\text{YR}})$. Il s'ensuit que $(\hat{\boldsymbol{\beta}}^{\text{YR}}, \hat{\mathbf{v}}^{\text{YR}})$ sont donnés par

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}^{\text{YR}} \\ \hat{\mathbf{v}}^{\text{YR}} \end{pmatrix} = \hat{\mathbf{A}}_w^{-1} \hat{\mathbf{b}}_w, \quad (2.15)$$

où les valeurs connues de $\mathbf{A}_w$ et $\mathbf{b}_w$ sont données par

$$\mathbf{A}_w = \begin{pmatrix} \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{R}^{-1} \mathbf{W}^{1/2} \mathbf{X} & \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{R}^{-1} \mathbf{W}^{1/2} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{W}^{1/2} \mathbf{R}^{-1} \mathbf{W}^{1/2} \mathbf{X} & \mathbf{Z}^T \mathbf{W}^{1/2} \mathbf{R}^{-1} \mathbf{W}^{1/2} \mathbf{Z} + \boldsymbol{\Omega}^{1/2} \mathbf{G}^{-1} \boldsymbol{\Omega}^{1/2} \end{pmatrix} \text{ et } \mathbf{b}_w = \begin{pmatrix} \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{R}^{-1} \mathbf{W}^{1/2} \mathbf{y} \\ \mathbf{Z}^T \mathbf{W}^{1/2} \mathbf{R}^{-1} \mathbf{W}^{1/2} \mathbf{y} \end{pmatrix}, \quad (2.16)$$

et $\hat{\mathbf{A}}_w$ et $\hat{\mathbf{b}}_w$ sont des versions empiriques de $\mathbf{A}_w$ et $\mathbf{b}_w$ obtenues en estimant $\mathbf{G}$ et $\mathbf{R}$ par $\hat{\mathbf{G}} = \hat{\sigma}_v^2 \mathbf{I}_m$ et $\hat{\mathbf{R}} = \hat{\sigma}_e^2 \mathbf{I}_n$ respectivement. L'équation (2.15) peut également s'écrire sous la forme

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}^{\text{YR}} \\ \hat{\mathbf{v}}^{\text{YR}} \end{pmatrix} = \begin{pmatrix} (\mathbf{X}^T \hat{\mathbf{V}}_w^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}_w^{-1} \mathbf{y} \\ \hat{\mathbf{G}}_\omega \mathbf{Z}^T \hat{\mathbf{V}}_w^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}^{\text{YR}}) \end{pmatrix}, \quad (2.17)$$

où $\hat{\mathbf{G}}_\omega = \boldsymbol{\Omega}^{-1/2} \hat{\mathbf{G}} \boldsymbol{\Omega}^{-1/2}$ et $\hat{\mathbf{V}}_w = \mathbf{W}^{-1/2} \hat{\mathbf{R}} \mathbf{W}^{-1/2} + \mathbf{Z} \hat{\mathbf{G}}_\omega \mathbf{Z}^T$.

3. Estimateurs réconciliés

Nous élaborons maintenant des estimateurs réconciliés des moyennes de petits domaines $\bar{Y}_i$ à l'aide du modèle au niveau de l'unité (2.2) ou de versions augmentées de ce dernier. Nous supposons qu'un estimateur direct fiable $\hat{Y}_w = \sum_{i=1}^m \sum_{j \in s_i} w_{ij} y_{ij}$ du total de population Y est disponible, où $Y = \sum_{i=1}^m Y_i$ et $Y_i = N_i \bar{Y}_i$ est le total du petit domaine i . Soit $\hat{\bar{Y}}_i$ l'estimateur sur petits domaines fondé sur un modèle de $\bar{Y}_i$. Il est souhaitable de veiller à ce que les valeurs agrégées de $\hat{\bar{Y}}_i$ correspondent à l'estimateurs fiable $\hat{Y}_w$. Les estimateurs de moyennes de petits domaines $\hat{\bar{Y}}_i$, $i = 1, \dots, m$ sont considérés réconciliés à $\hat{Y}_w$ si

$$\sum_{i=1}^m N_i \hat{\bar{Y}}_i = \hat{Y}_w. \quad (3.1)$$

Soit $\hat{Y}_w$ un estimateur GREG avec des poids calés au niveau de la population sur un vecteur de variables auxiliaires $\mathbf{x}_{ij}^*$. Cet estimateur est analogue à l'estimateur par la régression combiné si l'on considère les petits domaines comme des strates. Le vecteur de variables auxiliaires $\mathbf{x}_{ij}^*$ peut ou non être le même que $\mathbf{x}_{ij}$. Nous distinguons deux cas dans ce contexte : $\mathbf{x}_{ij} \subseteq \mathbf{x}_{ij}^*$ et $\mathbf{x}_{ij} \not\subseteq \mathbf{x}_{ij}^*$. Le premier cas, $\mathbf{x}_{ij} \subseteq \mathbf{x}_{ij}^*$, suppose que toutes les composantes de $\mathbf{x}_{ij}$ appartiennent également à $\mathbf{x}_{ij}^*$ et que $\mathbf{x}_{ij}^*$ peut ou non comporter d'autres composantes qui sont différentes de celles contenues dans $\mathbf{x}_{ij}$. Le second cas, $\mathbf{x}_{ij} \not\subseteq \mathbf{x}_{ij}^*$, suppose que certaines des composantes de $\mathbf{x}_{ij}$ n'apparaissent pas dans $\mathbf{x}_{ij}^*$. Nous supposons que la première composante des deux vecteurs $\mathbf{x}_{ij}$ et $\mathbf{x}_{ij}^*$ est égale à un, car ces derniers représentent un terme d'ordonnée à l'origine.

Pour un échantillon donné s , les données auxiliaires $\mathbf{x}_{ij}^*$ et les poids de sondage de base $d_{ij} = 1/\pi_{ij}$, l'estimateur GREG du total de population Y est donné par

$$\hat{Y}^{\text{GREG}} = \sum_{i=1}^m \sum_{j \in s_i} w_{ij}^{\text{GREG}} y_{ij},$$

où les poids GREG w_{ij}^{GREG} sont donnés par

$$w_{ij}^{\text{GREG}} = d_{ij} \left(1 + (\mathbf{X}^* - \hat{\mathbf{X}}^{*\text{HT}})^T \left(\sum_{i=1}^m \sum_{j \in s_i} d_{ij} \mathbf{x}_{ij}^* \mathbf{x}_{ij}^{*T} \right)^{-1} \mathbf{x}_{ij}^* \right). \quad (3.2)$$

Dans l'équation (3.2), $\mathbf{X}^* = \sum_{i=1}^m \mathbf{X}_i^*$, où $\mathbf{X}_i^* = \sum_{j=1}^{N_i} \mathbf{x}_{ij}^*$ représente le total de petits domaines connu, tandis que $\hat{\mathbf{X}}^{*\text{HT}} = \sum_{i=1}^m \hat{\mathbf{X}}_i^{*\text{HT}}$ et $\hat{\mathbf{X}}_i^{*\text{HT}} = \sum_{j \in s_i} d_{ij} \mathbf{x}_{ij}^*$ représentent respectivement les estimateurs directs de Horvitz-Thompson fondés sur un plan de sondage de $\mathbf{X}^*$ et $\mathbf{X}_i^*$. Il convient de souligner que

$$\sum_{i=1}^m \sum_{j \in s_i} w_{ij}^{\text{GREG}} \mathbf{x}_{ij}^* = \mathbf{X}^*. \quad (3.3)$$

À l'aide des poids GREG w_{ij}^{GREG} , les estimateurs de N_i et $\mathbf{X}_i$ sont donnés par

$$\hat{N}_i^{\text{GREG}} = \sum_{j \in s_i} w_{ij}^{\text{GREG}} \quad \text{et} \quad \hat{\mathbf{X}}_i^{\text{GREG}} = \sum_{j \in s_i} w_{ij}^{\text{GREG}} \mathbf{x}_{ij}. \quad (3.4)$$

Les estimations sur petits domaines $\hat{Y}_i^{\text{EBLUP}}$ et $\hat{Y}_i^{\text{YR}}$ données respectivement par les équations (2.11) et (2.13) ne satisfont pas l'équation de réconciliation (3.1) pour $\hat{Y}_w = \hat{Y}^{\text{GREG}}$, c'est-à-dire que les estimations totales $\hat{Y}^{\text{EBLUP}} = \sum_{i=1}^m N_i \hat{Y}_i^{\text{EBLUP}}$ et $\hat{Y}^{\text{YR}} = \sum_{i=1}^m N_i \hat{Y}_i^{\text{YR}}$ ne correspondent pas à l'estimateur GREG $\hat{Y}^{\text{GREG}}$. Nous devons ajuster $\hat{Y}_i^{\text{EBLUP}}$ et $\hat{Y}_i^{\text{YR}}$ de sorte que la somme de ces estimateurs sur petits domaines modifiés corresponde à $\hat{Y}^{\text{GREG}}$ lorsque la somme de tous les petits domaines m est faite.

Une modification très simple des $\hat{Y}_i^{\text{EBLUP}}$ et des $\hat{Y}_i^{\text{YR}}$ est appelée la réconciliation par le ratio. Elle consiste à multiplier chaque $\hat{Y}_i^{\text{EBLUP}}$ et $\hat{Y}_i^{\text{YR}}$ par les facteurs d'ajustement communs $\hat{Y}^{\text{GREG}} / \sum_{i=1}^m N_i \hat{Y}_i^{\text{EBLUP}}$ et $\hat{Y}^{\text{GREG}} / \sum_{i=1}^m N_i \hat{Y}_i^{\text{YR}}$ respectivement, permettant d'obtenir les estimateurs réconciliés par le ratio

$$\hat{Y}_{ib}^{\text{EBRat}} = \hat{Y}_i^{\text{EBLUP}} \frac{\hat{Y}^{\text{GREG}}}{\sum_{i=1}^m N_i \hat{Y}_i^{\text{EBLUP}}} \quad \text{et} \quad \hat{Y}_{ib}^{\text{YRat}} = \hat{Y}_i^{\text{YR}} \frac{\hat{Y}^{\text{GREG}}}{\sum_{i=1}^m N_i \hat{Y}_i^{\text{YR}}}. \quad (3.5)$$

Il s'ensuit facilement que $\hat{Y}_{ib}^{\text{EBRat}}$ et $\hat{Y}_{ib}^{\text{YRat}}$ satisfont tous deux l'équation (3.1) avec $\hat{Y}_w = \hat{Y}^{\text{GREG}}$. Dans l'équation (3.5) et ci-après, l'indice b indique que les estimateurs sont réconciliés à $\hat{Y}^{\text{GREG}}$.

Il convient de souligner que les $\hat{Y}_i^{\text{EBLUP}}$ et les $\hat{Y}_i^{\text{YR}}$ dans l'équation (3.5) sont multipliés par le même facteur peu importe leur précision et en ignorant les caractéristiques des petits domaines en particulier, comme la variabilité des unités dans un petit domaine ou la taille d'échantillon des petits domaines. En conséquence, les estimateurs réconciliés qui en résultent, $\hat{Y}_{ib}^{\text{EBRat}}$ et $\hat{Y}_{ib}^{\text{YRat}}$, d'après cette procédure simple, ne sont que des modifications proportionnelles des estimateurs $\hat{Y}_i^{\text{EBLUP}}$ et $\hat{Y}_i^{\text{YR}}$ respectivement, pour obtenir la concordance voulue. On peut éviter cette limite en utilisant le modèle pour petits domaines (2.2) afin de construire les estimateurs réconciliés.

Nous démontrons maintenant comment le modèle (2.2) peut être utilisé pour obtenir des estimateurs réconciliés à $\hat{Y}^{\text{GREG}}$. Aux sections 3.1 et 3.2, nous adaptons les procédures décrites dans Stefan et Hidioglou (2020) pour obtenir des estimateurs réconciliés au cas des taux d'échantillonnage non négligeables. Aux sections 3.3 et 3.4, nous présentons deux estimateurs réconciliés restreints d'après la procédure proposée par Ugarte et coll. (2009). Les estimateurs réconciliés des sections 3.1 et 3.2 reposent sur l'hypothèse selon laquelle $\mathbf{x}_{ij} \subseteq \mathbf{x}_{ij}^*$, tandis que les estimateurs des sections 3.3 et 3.4 peuvent être calculés pour n'importe quel vecteur $\mathbf{x}_{ij}$ et $\mathbf{x}_{ij}^*$.

3.1 Estimateurs réconciliés EBLUP augmentés

Les poids GREG w_{ij}^{GREG} doivent être utilisés dans l'estimation pour réaliser la réconciliation à $\hat{Y}^{\text{GREG}}$. Une façon possible d'intégrer w_{ij}^{GREG} à l'estimation consiste à augmenter le modèle pour petits domaines

(2.2) au moyen d'une variable auxiliaire adéquate qui est une fonction de w_{ij}^{GREG} . Cette procédure repose sur l'approche de modèle augmenté adoptée par Wang et coll. (2008), dans laquelle les estimations obtenues au moyen du modèle FH au niveau du domaine ont pu être contraintes de correspondre à des totaux déterminés. Stefan et Hidirolou (2020) ont adapté l'approche de Wang et coll. (2008) sous le modèle de base au niveau de l'unité et pour des taux d'échantillonnage négligeables. Ils ont démontré que la réconciliation à $\hat{Y}^{\text{GREG}}$ pouvait être obtenue en augmentant le modèle (2.2) avec les poids GREG w_{ij}^{GREG} . Nous étendons l'approche de Stefan et Hidirolou (2020) au cas des taux d'échantillonnage non négligeables. Dans ce cas, la réconciliation à $\hat{Y}^{\text{GREG}}$ est réalisée en augmentant le modèle (2.2) avec $q_{ij} = w_{ij}^{\text{GREG}} - 1$. Il en résulte le modèle augmenté donné par

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}_{1a} + q_{ij} \beta_{2a} + v_{ia} + e_{ija}, \quad i = 1, \dots, m; \quad j \in s_i. \quad (3.6)$$

Les effets aléatoires v_{ia} sont présumés i.i.d. $N(0, \sigma_{va}^2)$ et indépendants des erreurs au niveau de l'unité e_{ija} , et les e_{ija} sont présumés i.i.d. $N(0, \sigma_{ea}^2)$. Les estimateurs EBLUP de $\boldsymbol{\beta}_a = (\boldsymbol{\beta}_{1a}^T, \beta_{2a})^T$ et v_{ia} sous l'équation (3.6) sont respectivement désignés $\hat{\boldsymbol{\beta}}_a = (\hat{\boldsymbol{\beta}}_{1a}^T, \hat{\beta}_{2a})^T$ et $\hat{v}_{ia}$. Nous pouvons maintenant démontrer le résultat 1 pour $\hat{\boldsymbol{\beta}}_a$ et $\hat{v}_{ia}$.

Résultat 1. Les estimateurs EBLUP $\hat{\boldsymbol{\beta}}_a$ et $\hat{v}_{ia}$ reposant sur le modèle (3.6) obéissent à l'équation suivante

$$\sum_{i=1}^m \sum_{j \in s_i} y_{ij} + \left(\sum_{i=1}^m \mathbf{x}_{ir} \right)^T \hat{\boldsymbol{\beta}}_{1a} + \sum_{i=1}^m q_{iw} \hat{\beta}_{2a} + \sum_{i=1}^m (\hat{N}_i^{\text{GREG}} - n_i) \hat{v}_{ia} = \hat{Y}^{\text{GREG}}, \quad (3.7)$$

où $q_{iw} = \sum_{j \in s_i} q_{ij}^2 = \sum_{j \in s_i} (w_{ij}^{\text{GREG}} - 1)^2$.

Preuve : Voir l'annexe A.

Il découle de l'équation (3.7) que les estimateurs sur petits domaines réconciliés à $\hat{Y}^{\text{GREG}}$ sont donnés par

$$\hat{Y}_{iab}^{\text{EBLUP}} = \frac{1}{N_i} \left[\sum_{j \in s_i} y_{ij} + \mathbf{x}_{ir}^T \hat{\boldsymbol{\beta}}_{1a} + q_{iw} \hat{\beta}_{2a} + (\hat{N}_i^{\text{GREG}} - n_i) \hat{v}_{ia} \right]. \quad (3.8)$$

L'indice a indique que $\hat{Y}_{iab}^{\text{EBLUP}}$ repose sur le modèle pour petits domaines augmenté.

3.2 Estimateurs réconciliés de You-Rao

On peut utiliser la procédure proposée par You et Rao (2002) avec des poids de sondage w_{ij} quelconques. Cependant, il n'est pas garanti que l'estimateur YR qui en résultera sera réconcilié à $\hat{Y}^{\text{GREG}}$. Pour des taux d'échantillonnage négligeables, Stefan et Hidirolou (2020) ont obtenu des estimateurs réconciliés au moyen de la procédure de You et Rao (2002) d'après les poids $w_{ij} = w_{ij}^{\text{GREG}}$ de l'estimateur GREG. Pour des taux d'échantillonnage non négligeables, nous démontrons maintenant que les poids $w_{ij} = w_{ij}^{\text{GREG}} - 1$ permettent d'obtenir des estimateurs YR réconciliés.

Soient $\hat{\boldsymbol{\beta}}^{\text{YR}}$ et $\hat{\mathbf{v}}^{\text{YR}} = (\hat{v}_1^{\text{YR}}, \dots, \hat{v}_m^{\text{YR}})^T$ des estimateurs YR de $\boldsymbol{\beta}$ et $\mathbf{v}$ respectivement, w_{ij} étant remplacé par $w_{ij}^{\text{GREG}} - 1$. En utilisant $\hat{\boldsymbol{\beta}}^{\text{YR}}$, $\hat{v}_i^{\text{YR}}$ et les $N_i - n_i$ estimations $\hat{y}_{ij}^{\text{YR}} = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}^{\text{YR}} + \hat{v}_i^{\text{YR}}$ pour

$j \in r_i$, on peut calculer un estimateur YR, désigné $\hat{Y}_i^{\text{YR}}$, à l'aide de l'équation (2.13). Cependant, $\hat{Y}_i^{\text{YR}}$ n'est pas réconcilié à $\hat{Y}^{\text{GREG}}$ même s'il utilise les poids $w_{ij}^{\text{GREG}} - 1$. La procédure YR originale permet d'obtenir un estimateur autoréconcilié dans un nombre limité de cas.

Pour réaliser la réconciliation à $\hat{Y}^{\text{GREG}}$, un estimateur YR modifié, désigné $\hat{Y}_{ib}^{\text{YR}}$, est défini comme suit :

$$\hat{Y}_{ib}^{\text{YR}} = \frac{1}{N_i} \left[\sum_{j \in s_i} y_{ij} + \mathbf{x}_{ir}^T \hat{\boldsymbol{\beta}}^{\text{YR}} + (\hat{N}_i^{\text{GREG}} - n_i) \hat{v}_i^{\text{YR}} \right]. \quad (3.9)$$

Ce qui suit démontre que $\hat{Y}_{ib}^{\text{YR}}$ défini au moyen de l'équation (3.9) est réconcilié à $\hat{Y}^{\text{GREG}}$.

Résultat 2. Soient $\hat{\boldsymbol{\beta}}^{\text{YR}}$ et $\hat{\mathbf{v}}^{\text{YR}} = (\hat{v}_1^{\text{YR}}, \dots, \hat{v}_m^{\text{YR}})^T$ les estimateurs YR de $\boldsymbol{\beta}$ et $\mathbf{v}$ respectivement, construits avec les poids $w_{ij}^{\text{GREG}} - 1$. Alors, $(\hat{\boldsymbol{\beta}}^{\text{YR}}, \hat{\mathbf{v}}^{\text{YR}})$ satisfait l'équation suivante :

$$\sum_{i=1}^m \sum_{j \in s_i} y_{ij} + \sum_{i=1}^m \mathbf{x}_{ir}^T \hat{\boldsymbol{\beta}}^{\text{YR}} + \sum_{i=1}^m (\hat{N}_i^{\text{GREG}} - n_i) \hat{v}_i^{\text{YR}} = \hat{Y}^{\text{GREG}}.$$

Preuve : Voir l'annexe A.

Étant donné $\mathbf{x}_{ij}^*$, les poids w_{ij}^{GREG} sont calés sur $\mathbf{x}_{ij}^*$ au niveau du petit domaine s'ils satisfont les équations suivantes

$$\sum_{j \in s_i} w_{ij}^{\text{GREG}} \mathbf{x}_{ij}^* = \mathbf{X}_i^*, \quad \text{pour } i = 1, \dots, m. \quad (3.10)$$

Les équations (3.10) supposent l'équation (3.3), mais l'inverse n'est pas vrai. Si les poids w_{ij}^{GREG} satisfont les équations (3.10), et étant donné que $\mathbf{x}_{ij} \subseteq \mathbf{x}_{ij}^*$, il s'ensuit que les poids w_{ij}^{GREG} sont également calés sur $\mathbf{x}_{ij}$ au niveau du petit domaine. Du même coup, cela suppose que $\hat{N}_i^{\text{GREG}} = N_i$, car nous présumons que le vecteur $\mathbf{x}_{ij}$ contient le régresseur constant égal à 1. Il s'ensuit que $\hat{Y}_i^{\text{YR}} = \hat{Y}_{ib}^{\text{YR}}$. Ainsi, l'estimateur YR $\hat{Y}_i^{\text{YR}}$ construit avec $w_{ij}^{\text{GREG}} - 1$ est autoréconcilié à $\hat{Y}^{\text{GREG}}$ dans le cas particulier où les poids GREG sont calés au niveau du petit domaine (voir You et Rao, 2002).

3.3 Estimateur réconcilié EBLUP restreint

À la section 2, nous avons démontré qu'on peut obtenir les estimateurs EBLUP de $(\boldsymbol{\beta}, \mathbf{v})$ si la fonction ϕ définie dans l'équation (2.5) est minimisée en ce qui concerne $(\boldsymbol{\beta}, \mathbf{v})$. Il s'ensuit par conséquent qu'on peut considérer un estimateur EBLUP comme la solution à un problème de minimisation non restreinte. L'idée des estimateurs EBLUP restreints est d'obtenir de nouveaux estimateurs de $(\boldsymbol{\beta}, \mathbf{v})$ en minimisant ϕ sous la restriction donnée par la condition de réconciliation. La procédure a été utilisée par Pfeiffermann et Barnard (1991) sous le modèle FH au niveau du domaine. Plus récemment, Ugarte et coll. (2009) ont appliqué la procédure sous le modèle BHF au niveau de l'unité pour obtenir la réconciliation à un estimateur synthétique. Ugarte et coll. (2009) ont décrit l'estimateur restreint comme un estimateur par les moindres carrés généralisés sous une restriction en constatant que la minimisation peut être réalisée

comme dans la théorie économétrique de l'estimation par régression sous contraintes linéaires. Nous décrivons maintenant la procédure présentée dans Ugarte et coll. (2009).

Nous désignons $\hat{\boldsymbol{\beta}}^R$ et $\hat{\mathbf{v}}^R = (\hat{v}_1^R, \dots, \hat{v}_m^R)^T$ les nouveaux estimateurs EBLUP restreints de $(\boldsymbol{\beta}, \mathbf{v})$. Ensuite, l'estimateur EBLUP restreint de $\bar{Y}_i$, désigné $\hat{Y}_{ib}^{\text{REBLUP}}$, est donné par l'équation (2.4), où $\hat{y}_{ij}$ sont remplacés par $\hat{y}_{ij}^R = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}^R + \hat{v}_i^R$, pour $j \in r_i$. Nous imposons la condition que les estimateurs $\hat{Y}_{ib}^{\text{REBLUP}}$, $i = 1, \dots, m$ soient réconciliés à $\hat{Y}^{\text{GREG}}$, c'est-à-dire qu'ils satisfassent l'équation (3.1) avec $\hat{Y}_w = \hat{Y}^{\text{GREG}}$. Après avoir fait quelques calculs algébriques, on peut démontrer que la réconciliation à $\hat{Y}^{\text{GREG}}$ des estimateurs $\hat{Y}_{ib}^{\text{REBLUP}}$, $i = 1, \dots, m$ équivaut à l'équation de contrainte linéaire suivante

$$\mathbf{a}_1^T \hat{\boldsymbol{\beta}}^R + \mathbf{a}_2^T \hat{\mathbf{v}}^R = \hat{Y}_r^{\text{GREG}}, \quad (3.11)$$

où $\mathbf{a}_1 = \sum_{i=1}^m \mathbf{x}_{ir}$, $\mathbf{a}_2 = (N_1 - n_1, \dots, N_m - n_m)^T$, $Y_r = Y - \sum_{i=1}^m \sum_{j \in s_i} y_{ij}$ est le total des valeurs de y_{ij} non observées avec $i = 1, \dots, m$; $j \in r_i$, et $\hat{Y}_r^{\text{GREG}} = \hat{Y}^{\text{GREG}} - \sum_{i=1}^m \sum_{j \in s_i} \hat{y}_{ij}$ est un estimateur de Y_r d'après $\hat{Y}^{\text{GREG}}$. Les estimateurs EBLUP restreints $(\hat{\boldsymbol{\beta}}^R, \hat{\mathbf{v}}^R)$ sont par conséquent obtenus comme la solution à la fonction de minimisation ϕ donnée par l'équation (2.5) sous la contrainte linéaire (3.11).

On peut utiliser la méthode des multiplicateurs de Lagrange pour résoudre la minimisation sous contrainte de ϕ . Après des calculs algébriques simples, on peut démontrer que des estimateurs $(\hat{\boldsymbol{\beta}}^R, \hat{\mathbf{v}}^R)$ sont donnés par

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}^R \\ \hat{\mathbf{v}}^R \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{v}} \end{pmatrix} + \frac{1}{\mathbf{a}^T \hat{\mathbf{A}} \mathbf{a}} \hat{\mathbf{A}}^{-1} \mathbf{a} \left[\hat{Y}_r^{\text{GREG}} - \mathbf{a}^T \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{v}} \end{pmatrix} \right], \quad (3.12)$$

où $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}})$ sont les estimateurs EBLUP (non contraints) de $(\boldsymbol{\beta}, \mathbf{v})$, $\hat{\mathbf{A}}$ est la version empirique de la matrice $\mathbf{A}$ définie dans l'équation (2.7) et $\mathbf{a} = (\mathbf{a}_1^T \mathbf{a}_2^T)^T$. Ensuite, en utilisant $\hat{y}_{ij}^R$ dans l'équation (2.4), on peut réécrire l'estimateur $\hat{Y}_{ib}^{\text{REBLUP}}$ sous la forme

$$\hat{Y}_{ib}^{\text{REBLUP}} = \frac{1}{N_i} \left[\sum_{j \in s_i} y_{ij} + \mathbf{x}_{ir}^T \hat{\boldsymbol{\beta}}^R + (N_i - n_i) \hat{v}_i^R \right]. \quad (3.13)$$

Remarque 2. La matrice $\hat{\mathbf{A}}$ n'existe pas pour les échantillons lorsque $\hat{\sigma}_v^2 = 0$. En pareil cas, nous avons constaté que l'équation (2.8) ne peut servir à calculer les estimateurs non contraints $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}})$. Cependant, on peut tout de même calculer $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}})$ lorsque $\hat{\sigma}_v^2 = 0$ parce que l'autre équation (2.9) peut être utilisée pour $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}})$. L'équation (3.12) démontre clairement que l'estimateur contraint $(\hat{\boldsymbol{\beta}}^R, \hat{\mathbf{v}}^R)$ ne peut être calculé pour des échantillons lorsque l'estimateur $\hat{\sigma}_v^2$ est tronqué à zéro, et aucune autre équation n'existe en pareil cas.

Il s'ensuit par conséquent que les méthodes d'estimation des composantes de la variance couramment utilisées dans l'EPD ne peuvent être utilisées pour calculer l'estimateur EBLUP restreint. À la section 3.4 et à l'annexe B, nous décrivons une autre méthode permettant de produire une estimation strictement positive de σ_v^2 pouvant être appliquée de pair avec $(\hat{\boldsymbol{\beta}}^R, \hat{\mathbf{v}}^R)$, de sorte qu'un estimateur réconcilié restreint de $\bar{Y}_i$ existe toujours.

3.4 Estimateur réconcilié de You-Rao restreint

Nous avons démontré à la section 2.2 qu'on peut obtenir des estimateurs YR de β et v comme une solution aux équations d'un modèle mixte obtenues en minimisant la fonction pondérée de l'échantillon ϕ_w donnée par l'équation (2.14). Autrement dit, nous avons démontré que, en définissant une fonction ϕ_w avec des poids $\{w_{ij}\}$, $i = 1, \dots, m$; $j \in s_i$ et $\{\omega_i\}$, $i = 1, \dots, m$, puis en minimisant ϕ_w , nous obtenons les mêmes estimateurs que ceux donnés par la procédure de You et Rao (2002). Nous minimisons maintenant la fonction ϕ_w sous la contrainte de réconciliation donnée par l'équation (3.11). Il en résulte un estimateur YR restreint qui est réconcilié à $\hat{Y}^{\text{GREG}}$.

La minimisation de ϕ_w compte tenu de la restriction de réconciliation (3.11) permet d'obtenir des estimateurs de $\bar{Y}_i$, $i = 1, \dots, m$ dont la réconciliation est garantie pour des poids quelconques qui définissent la fonction ϕ_w . Ainsi, on peut choisir un ensemble de poids w_{ij} quelconque dans ϕ_w . Dans une étude par simulations limitée fondée sur un plan de sondage, nous avons comparé trois estimateurs YR restreints en fonction de trois options pour w_{ij} : i. $w_{ij} = w_{ij}^{\text{GREG}} - 1$; ii. $w_{ij} = w_{ij}^{\text{GREG}}$; et iii. $w_{ij} = d_{ij}$. Nous n'avons constaté aucune différence significative entre ces trois estimateurs pour ce qui est de l'erreur quadratique moyenne du plan. Compte tenu de ce dernier point et étant donné que les estimateurs YR réconciliés non restreints décrits à la section 3.2 reposaient sur $w_{ij} = w_{ij}^{\text{GREG}} - 1$, nous avons choisi de définir l'estimateur YR restreint en fonction de ces poids.

Soit ϕ_w défini en fonction de $w_{ij} = w_{ij}^{\text{GREG}} - 1$ et $\omega_i = \sum_{j \in s_i} w_{ij}^2 / \sum_{j \in s_i} w_{ij}$. La minimisation de ϕ_w en ce qui concerne (β, v) sous la contrainte de réconciliation (3.11) permet d'obtenir les estimateurs YR restreints de (β, v) , désignés $(\hat{\beta}^{\text{RYR}}, \hat{v}^{\text{RYR}})$. Ces estimateurs sont donnés par :

$$\begin{pmatrix} \hat{\beta}^{\text{RYR}} \\ \hat{v}^{\text{RYR}} \end{pmatrix} = \begin{pmatrix} \hat{\beta}^{\text{YR}} \\ \hat{v}^{\text{YR}} \end{pmatrix} + \frac{1}{\mathbf{a}^T \hat{\mathbf{A}}_w \mathbf{a}} \hat{\mathbf{A}}_w^{-1} \mathbf{a} \left[\hat{Y}_r^{\text{GREG}} - \mathbf{a}^T \begin{pmatrix} \hat{\beta}^{\text{YR}} \\ \hat{v}^{\text{YR}} \end{pmatrix} \right], \quad (3.14)$$

où les estimateurs $(\hat{\beta}^{\text{YR}}, \hat{v}^{\text{YR}})$ sont donnés par l'équation (2.15) et $\hat{\mathbf{A}}_w$ est la version empirique de $\mathbf{A}_w$ donnée par l'équation (2.16). En utilisant $\hat{\beta}^{\text{RYR}}$ et $\hat{v}_i^{\text{RYR}}$ de $\hat{v}^{\text{RYR}} = (\hat{v}_1^{\text{RYR}}, \dots, \hat{v}_m^{\text{RYR}})^T$, les estimations YR restreintes $\hat{y}_{ij}^{\text{RYR}} = \mathbf{x}_{ij}^T \hat{\beta}^{\text{RYR}} + \hat{v}_i^{\text{RYR}}$ de y_{ij} non observé pour $j \in r_i$ sont ensuite utilisées pour calculer un estimateur YR restreint réconcilié :

$$\hat{Y}_{ib}^{\text{RYR}} = \frac{1}{N_i} \left[\sum_{j \in s_i} y_{ij} + \mathbf{x}_{ir}^T \hat{\beta}^{\text{RYR}} + (N_i - n_i) \hat{v}_i^{\text{RYR}} \right]. \quad (3.15)$$

Comme dans le cas de l'estimateur EBLUP restreint, les estimateurs $(\hat{\beta}^{\text{RYR}}, \hat{v}^{\text{RYR}})$ donnés par l'équation (3.14) n'existent pas si la méthode FC, le MV et le MVRE permettent d'obtenir une estimation tronquée pour σ_v^2 . En conséquence, $\bar{Y}_i$ peut seulement être estimé par $\hat{Y}_{ib}^{\text{RYR}}$ au moyen d'une méthode d'estimation des composantes de la variance qui permet toujours d'obtenir des estimations strictement positives pour σ_v^2 .

Une estimation nulle de σ_v^2 ne pose pas de problème dans le calcul des estimateurs EBLUP et YR. Cependant, nous avons constaté que les estimateurs EBLUP et YR restreints ne peuvent être calculés si $\hat{\sigma}_v^2 = 0$. Pour contourner ce problème, nous utilisons une méthode proposée par Moghtased-Azar,

Tehranchi et Amiri-Simkooei (2014) qui garantit que l'estimateur de σ_v^2 sera strictement positif. Cette méthode repose sur le concept d'un maximum de vraisemblance restreint reparamétré (MVRERE). L'idée est d'utiliser des fonctions dont l'intervalle est l'ensemble de tous les nombres réels positifs, à savoir des fonctions à valeur positive (FVP), pour des composantes inconnues de la variance dans le modèle stochastique plutôt que d'utiliser les composantes de la variance elles-mêmes. Leurs résultats numériques ont démontré l'estimation réussie sous contrainte de non-négativité des composantes de la variance (comme des valeurs positives) ainsi que des composantes de la covariance (comme des valeurs négatives ou positives).

Nous avons utilisé un algorithme de cotation de Fisher pour obtenir de manière itérative des estimations par le MVRERE des composantes de la variance du modèle de base au niveau de l'unité donné par l'équation (2.2) (voir l'annexe B pour plus de détails). Nous avons également effectué une petite simulation et constaté que, pour des tailles d'échantillon de petits domaines égales ou supérieures à 3, l'algorithme de cotation de Fisher convergait en moins de 15 itérations. Lorsque nous avons uniquement considéré les échantillons qui permettaient de produire une estimation nulle $\hat{\sigma}_v^2 = 0$, nous avons observé que l'algorithme convergait encore plus rapidement (voir la figure 4.1 à la section 4).

4. Étude par simulations

Nous présentons dans cette section les résultats d'une étude par simulations fondée sur un plan de sondage, car une telle étude rejoint les mesures qui sont calculées par les organismes nationaux de statistique. Une étude fondée sur un plan de sondage en est une dans laquelle une population finie fixe est d'abord générée au moyen d'un modèle hypothétique puis, pour chaque exécution de la simulation, un échantillon est prélevé à partir de la population finie fixe. Le but de l'étude par simulations est d'évaluer les propriétés des estimateurs réconciliés décrits à la section 3 quant au biais sous le plan et à l'erreur quadratique moyenne du plan. Nous avons examiné deux scénarios : le *Scénario 1* correspond au cas d'une modélisation exacte, tandis que le *Scénario 2* correspond au cas d'une modélisation inexacte. Des diagnostics de modélisation comme ceux présentés dans Rao et Molina (2015, pages 114-118) peuvent être utilisés pour vérifier si les modèles sont exacts ou non. De tels diagnostics de modélisation comprennent l'analyse de résidus, pour détecter les écarts par rapport au modèle hypothétique, et les diagnostics de sélection des variables auxiliaires pour le modèle et de suppression de cas, pour détecter les observations influentes.

4.1 Configuration de la simulation pour générer les populations finies

Pour chaque scénario, nous avons considéré cinq populations. Chaque population comptait $m = 30$ petits domaines, $N_i = 100$ unités de population se trouvant dans chaque petit domaine. Les populations correspondant au *Scénario 1* ont été créées au moyen du modèle suivant

$$y_{ij} = \beta_0 + x_{ij}\beta_1 + v_i + e_{ij}, \quad i = 1, \dots, m; \quad j = 1, \dots, N_i, \quad (4.1)$$

où $\beta_0 = 10$ et $\beta_1 = 5$. Pour générer les populations dans le *Scénario 2*, nous avons divisé 30 petits domaines en trois groupes égaux de petites domaines, désignés G_ℓ , pour $\ell = 1, 2, 3$. Le premier groupe G_1 contient les domaines $i = 1, \dots, 10$, le deuxième groupe G_2 contient les domaines $i = 11, \dots, 20$ et le troisième groupe G_3 contient les domaines $i = 21, \dots, 30$. Le modèle à l'intérieur d'un groupe donné est donné par

$$y_{ij} = \beta_{0,\ell} + x_{ij}\beta_{1,\ell} + v_i + e_{ij}, i \in G_\ell; j = 1, \dots, N_i, \quad (4.2)$$

où $(\beta_{0,1} = 10, \beta_{1,1} = 1)$ pour les domaines $i \in G_1$, $(\beta_{0,2} = 20, \beta_{1,2} = 5)$ pour les domaines $i \in G_2$ et $(\beta_{0,3} = 30, \beta_{1,3} = 10)$ pour les domaines $i \in G_3$. Les équations (4.1) et (4.2) utilisent toutes deux la variable auxiliaire $\mathbf{x}_{ij} = (1, x_{ij})^T$ dont les valeurs x_{ij} , $j = 1, \dots, N_i$ ont été générées à partir d'une loi exponentielle de moyenne égale à 5 et de variance égale à 25.

Les composantes aléatoires dans les équations (4.1) et (4.2) ont été générées à partir des lois normales $v_i \sim N(0, \sigma_v^2)$ et $e_{ij} \sim N(0, \sigma_e^2)$. Les cinq populations correspondant au *Scénario 1*, désignées A1, B1, C1, D1 et E1, ont été générées d'après l'équation (4.1) et les doublets de paramètres de variance suivants : i. $(\sigma_{Av}^2 = 0, 2, \sigma_e^2 = 20)$ pour la population A1; ii. $(\sigma_{Bv}^2 = 1, \sigma_e^2 = 20)$ pour la population B1; iii. $(\sigma_{Cv}^2 = 2, \sigma_e^2 = 20)$ pour la population C1; iv. $(\sigma_{Dv}^2 = 4, \sigma_e^2 = 20)$ pour la population D1; et $(\sigma_{Ev}^2 = 20, \sigma_e^2 = 20)$ pour la population E1. Il convient de souligner que, pour les populations A1 à E1, la valeur de σ_e^2 demeure fixe, tandis que les valeurs de σ_v^2 varient. Les σ_v^2 sont choisis pour obtenir les ratios de variance $\delta = \sigma_v^2 / \sigma_e^2$ à 0,01; 0,05; 0,1; 0,2 et 1. Les cinq populations dans le *Scénario 2*, désignées A2, B2, C2, D2 et E2, ont été générées d'après l'équation (4.2) avec les mêmes doublets de paramètres de variance que pour le *Scénario 1*.

Un plan d'échantillonnage stratifié a été utilisé en prélevant des échantillons probabilistes indépendants proportionnels à des tailles d'échantillon (PPT) de taille n_i dans le i^e petit domaine. Les tailles d'échantillon des petits domaines ont été prises $n_i = 3$ pour $i = 1, \dots, m$. Les probabilités de sélection ont été calculées comme $p_{ij} = b_{ij} / \sum_{j=1}^{N_i} b_{ij}$, où les mesures de taille sont $b_{ij} = x_{ij}$. Nous avons utilisé l'échantillonnage de Poisson conditionnel (EPC) pour sélectionner les échantillons PPT dans chaque petit domaine (voir Tillé (2006), chapitre 5). Les poids de sondage de base sont donnés par $d_{ij} = 1/(n_i p_{ij})$.

Dans le *Scénario 1*, nous avons ajusté le modèle de régression à erreurs emboîtées (4.1) et sa version augmentée en fonction des données de l'échantillonnage PPT sélectionnées à partir d'une des cinq populations générées au moyen du modèle (4.1). Ce scénario représente une modélisation exacte, car le modèle ajusté et le modèle utilisé pour générer la population finie coïncident. Dans le *Scénario 2*, nous avons ajusté le modèle de régression à erreurs emboîtées (4.1) et sa version augmentée en fonction des données de l'échantillonnage PPT sélectionnées à partir d'une des cinq populations générées au moyen du modèle (4.2). Ce scénario représente une modélisation inexacte, car le modèle ajusté et le modèle utilisé pour générer la population finie ne coïncident pas.

Nous avons sélectionné $G = 30\,000$ échantillons PPT stratifiés à partir de chacune des dix populations finies : les populations A1 à E1 correspondant au *Scénario 1* et les populations A2 à E2 correspondant au *Scénario 2*. Pour $g = 1, \dots, G$, soient $(\hat{\sigma}_v^{2\text{RE}(g)}, \hat{\sigma}_e^{2\text{RE}(g)})$ et $(\hat{\sigma}_v^{2\text{reRE}(g)}, \hat{\sigma}_e^{2\text{reRE}(g)})$ désignant respectivement les estimations de (σ_v^2, σ_e^2) obtenues au moyen de la méthode du MVRE tronqué et de sa version reparamétrée, qui correspond au g^{e} échantillon. Les valeurs de départ dans l'équation (B.2) étaient $\alpha_1^{(0)} = \log(0,1 + \hat{\sigma}_v^{2\text{RE}(g)})$ et $\alpha_2^{(0)} = \log(\hat{\sigma}_e^{2\text{RE}(g)})$. L'équation (B.2) a atteint la convergence en moins de 15 itérations pour l'ensemble des populations et dans les deux scénarios. D'après les échantillons simulés G sélectionnés dans chacune des cinq populations correspondant au *Scénario 1*, nous avons calculé la valeur Monte-Carlo de la probabilité d'obtenir une estimation du MVRE tronquée à zéro pour σ_v^2 , étant donné que

$$P_{\text{MC}}(\hat{\sigma}_v^{2\text{RE}} = 0) = \frac{1}{G} \sum_{g=1}^G I(\hat{\sigma}_v^{2\text{RE}(g)} = 0),$$

où $I(A)$ est une fonction indicatrice ayant la valeur 1 si la condition A est vérifiée, et la valeur 0 autrement.

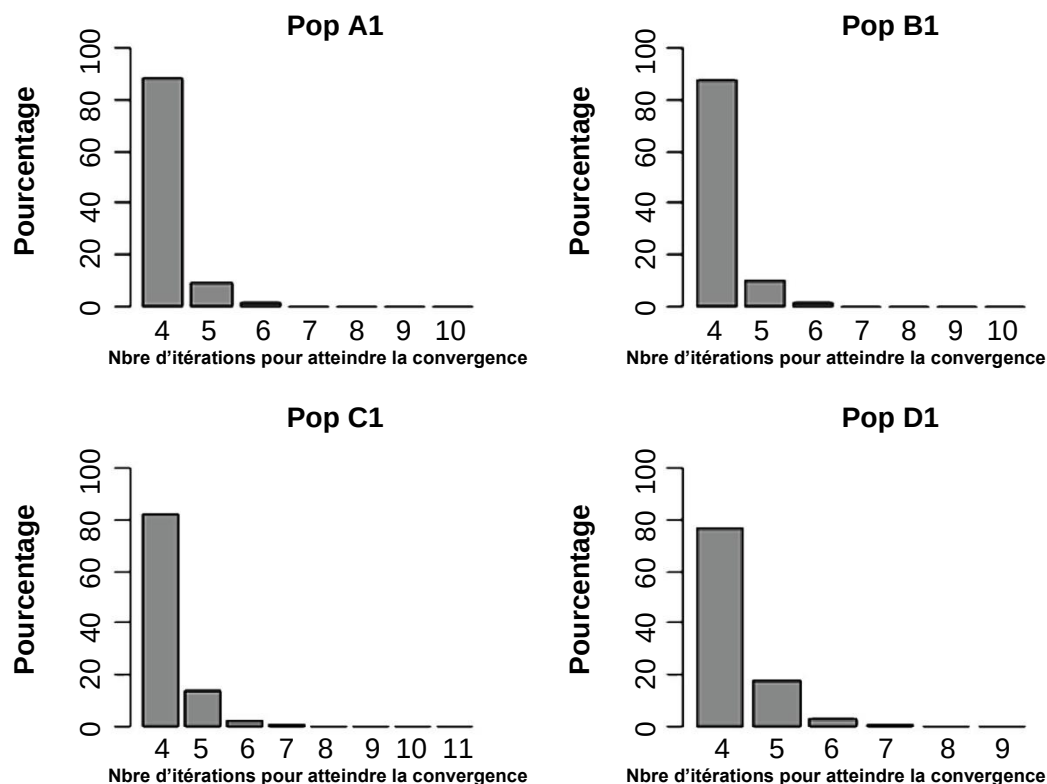
Le tableau 4.1 présente les valeurs Monte-Carlo de la probabilité d'obtenir une estimation nulle pour $\hat{\sigma}_v^{2\text{RE}}$. On peut constater que la probabilité simulée $P_{\text{MC}}(\hat{\sigma}_v^{2\text{RE}} = 0)$ peut être aussi élevée que 0,47 pour $\delta = 0,01$. À mesure que δ augmente, cette probabilité empirique diminue. Le tableau 4.1 montre clairement que les estimations $(\hat{\sigma}_v^{2\text{RE}}, \hat{\sigma}_e^{2\text{RE}})$ ne peuvent servir à calculer les estimateurs EBLUP et YR restreints pour des échantillons sélectionnés dans les populations A1, B1, C1 et D1.

Tableau 4.1
Valeurs de $P(\hat{\sigma}_v^{2\text{RE}} = 0)$: Scénario 1

	Pop A1 $\delta = 0,01$	Pop B1 $\delta = 0,05$	Pop C1 $\delta = 0,1$	Pop D1 $\delta = 0,2$	Pop E1 $\delta = 1$
$P_{\text{MC}}(\hat{\sigma}_v^{2\text{RE}} = 0)$	0,47	0,40	0,21	0,06	0,00

La figure 4.1 présente le nombre d'itérations pour atteindre la convergence de l'algorithme de cotation de Fisher pour l'estimation $\hat{\sigma}_v^{2\text{reRE}}$ de σ_v^2 . L'algorithme s'arrête lorsque la valeur de $|\hat{\sigma}_v^{2\text{reRE}(r+1)} - \hat{\sigma}_v^{2\text{reRE}(r)}|$ est inférieure à 10^{-5} , où $\hat{\sigma}_v^{2\text{reRE}(r)}$ représente la r^{e} itération calculée au moyen de l'équation (B.2) à l'annexe B. Les pourcentages présentés dans la figure 4.1 reposent uniquement sur les échantillons ayant une estimation du MVRE tronquée de σ_v^2 , c'est-à-dire $\hat{\sigma}_v^{2\text{RE}} = 0$. Nous avons uniquement considéré les populations A1, B1, C1 et D1, car ces quatre populations présentent des probabilités non négligeables que $\hat{\sigma}_v^{2\text{RE}}$ soit nul. La figure 4.1 montre clairement que la convergence est atteinte en un maximum de 11 itérations.

Figure 4.1 Pourcentage d'itérations pour atteindre la convergence dans les échantillons avec $\hat{\sigma}_v^{2RE} = 0$.



4.2 Comparaison des estimateurs réconciliés

Le but de l'étude par simulations est de comparer les estimateurs réconciliés décrits à la section 3 quant au biais sous le plan et à l'erreur quadratique moyenne du plan. Nous avons utilisé les deux scénarios, car nous voulions vérifier la protection qu'offre la réconciliation contre une modélisation inexacte. De plus, nous avons examiné la réconciliation à deux estimateurs GREG : $\hat{Y}_1^{GREG}$ et $\hat{Y}_2^{GREG}$. L'estimateur $\hat{Y}_1^{GREG}$ a des poids donnés par l'équation (3.2) calés sur le vecteur auxiliaire $\mathbf{x}_{ij} = (1, x_{ij})$ associé au modèle pour petits domaines. Il s'ensuit que l'estimateur $\hat{Y}_1^{GREG}$ correspond au cas de $\mathbf{x}_{ij} \subseteq \mathbf{x}_{ij}^*$. Le second estimateur GREG $\hat{Y}_2^{GREG}$ a des poids donnés par l'équation (3.2) d'après le vecteur auxiliaire $\mathbf{x}_{ij}^* = (1, x_{ij}^*)$, où les valeurs x_{ij}^* , $j = 1, \dots, N_i$ ont été générées à partir d'une loi exponentielle de moyenne égale à 5 et de variance égale à 25, et indépendamment des valeurs x_{ij} , $j = 1, \dots, N_i$. Il s'ensuit que l'estimateur $\hat{Y}_2^{GREG}$ correspond au cas de $\mathbf{x}_{ij} \not\subseteq \mathbf{x}_{ij}^*$, car la variable auxiliaire x_{ij} associée au modèle au niveau de l'unité (4.1) n'appartient pas au vecteur $\mathbf{x}_{ij}^*$ utilisé pour obtenir les poids associés à $\hat{Y}_2^{GREG}$.

Pour une population finie fixe, soient $\bar{Y}_i$ la moyenne du petit domaine i et $\hat{Y}_i$ un estimateur générique de $\bar{Y}_i$. Nous désignons $\hat{Y}_i^{(g)}$ la valeur de $\hat{Y}_i$ d'après le g^e échantillon simulé, pour $g = 1, \dots, G$. Les estimateurs décrits à la section 3 respectent la propriété de réconciliation peu importe la méthode utilisée pour estimer les composantes de la variance. Étant donné que les estimateurs réconciliés restreints

reposent sur des estimations $(\hat{\sigma}_v^{2\text{reRE}(g)}, \hat{\sigma}_e^{2\text{reRE}(g)})$, nous avons décidé d'utiliser le MVRERe dans le calcul de $\hat{Y}_i^{(g)}$ pour chaque estimateur $\hat{Y}_i$ évalué dans cette étude par simulations.

Nous avons examiné les mesures de rendement suivantes :

Biais relatif absolu moyen

$$\overline{\text{BRA}} = \frac{1}{m} \sum_{i=1}^m \text{BRA}_i \quad \text{avec} \quad \text{BRA}_i = \left| \frac{1}{G} \sum_{g=1}^G \frac{\hat{Y}_i^{(g)}}{\bar{Y}_i} - 1 \right|$$

Racine de l'erreur quadratique moyenne relative

$$\overline{\text{REQMR}} = \frac{1}{m} \sum_{i=1}^m \text{REQMR}_i \quad \text{avec} \quad \text{REQMR}_i = \sqrt{\frac{1}{G} \sum_{g=1}^G \left(\frac{\hat{Y}_i^{(g)}}{\bar{Y}_i} - 1 \right)^2}.$$

Cette partie de la simulation est résumée dans quatre tableaux. Nous présentons les résultats séparément pour les *Scénarios 1* et *2*. Les résultats dans le cas où la réconciliation est pour $\hat{Y}_1^{\text{GREG}}$ (le cas de $\mathbf{x}_{ij} \subseteq \mathbf{x}_{ij}^*$) sont résumés dans les tableaux 4.2 (*Scénario 1*) et 4.3 (*Scénario 2*). Les résultats dans le cas où la réconciliation est pour $\hat{Y}_2^{\text{GREG}}$ (le cas de $\mathbf{x}_{ij} \not\subseteq \mathbf{x}_{ij}^*$) sont résumés dans les tableaux 4.4 (*Scénario 1*) et 4.5 (*Scénario 2*).

Réconciliation à $\hat{Y}_1^{\text{GREG}}$ (le cas de $\mathbf{x}_{ij} \subseteq \mathbf{x}_{ij}^$)*

Nous avons calculé le $\overline{\text{BRA}}$ et la $\overline{\text{REQMR}}$ pour deux estimateurs non réconciliés, $\hat{Y}_i^{\text{EBLUP}}$ et $\hat{Y}_i^{\text{YR}}$, ainsi que leurs estimateurs correspondants réconciliés à $\hat{Y}_1^{\text{GREG}}$. Pour $\hat{Y}_i^{\text{EBLUP}}$, nous avons trois estimateurs réconciliés, $\hat{Y}_{ib}^{\text{EBRat}}$, $\hat{Y}_{iab}^{\text{EBLUP}}$ et $\hat{Y}_{ib}^{\text{REBLUP}}$, donnés respectivement par les équations (3.5), (3.8) et (3.13). Pour $\hat{Y}_i^{\text{YR}}$, les estimateurs réconciliés correspondants sont $\hat{Y}_{ib}^{\text{YRat}}$, $\hat{Y}_{ib}^{\text{YR}}$ et $\hat{Y}_{ib}^{\text{RYR}}$, donnés respectivement par les équations (3.5), (3.9) et (3.15).

Nous examinons d'abord leurs propriétés lorsque le modèle est exact (*Scénario 1*). En comparant les $\overline{\text{BRA}}$ de l'ensemble des estimateurs au tableau 4.2, nous constatons qu'il n'y a pas de grande différence entre les estimateurs. Les estimateurs EBLUP présentent des $\overline{\text{BRA}}$ un peu plus faibles que les estimateurs reposant sur la procédure YR. L'estimateur réconcilié $\hat{Y}_{iab}^{\text{EBLUP}}$ présente les $\overline{\text{BRA}}$ les plus faibles, tandis que les $\overline{\text{BRA}}$ des estimateurs réconciliés $\hat{Y}_{ib}^{\text{EBRat}}$ et $\hat{Y}_{ib}^{\text{REBLUP}}$ sont identiques à ceux de $\hat{Y}_i^{\text{EBLUP}}$. Les valeurs du $\overline{\text{BRA}}$ associées aux estimateurs $\hat{Y}_i^{\text{YR}}$, $\hat{Y}_{ib}^{\text{YRat}}$ et $\hat{Y}_{ib}^{\text{RYR}}$ se rapprochent, tandis que l'estimateur $\hat{Y}_{ib}^{\text{YR}}$ présente un biais relatif un peu plus élevé, surtout pour des valeurs plus grandes de $\delta = \sigma_v^2 / \sigma_e^2$. Pour l'ensemble des estimateurs, les $\overline{\text{BRA}}$ augmentent à mesure que δ augmente : de légères exceptions se produisent lorsque $\delta = 1$.

Ensuite, nous nous penchons sur les $\overline{\text{REQMR}}$. Comme on pouvait s'y attendre, les $\overline{\text{REQMR}}$ les plus faibles sont associées à $\hat{Y}_{iab}^{\text{EBLUP}}$, tandis que l'estimateur $\hat{Y}_i^{\text{YR}}$ présente des valeurs de la $\overline{\text{REQMR}}$ un peu plus grandes en raison de l'utilisation de poids de sondage sous une modélisation exacte. La réconciliation entraîne une augmentation de la $\overline{\text{REQMR}}$. Il convient de souligner que les $\overline{\text{REQMR}}$ des estimateurs réconciliés $\hat{Y}_{iab}^{\text{EBLUP}}$ et $\hat{Y}_{ib}^{\text{YR}}$, donnés aux sections 3.1 et 3.2 respectivement, sont plus élevées que celles

associées aux méthodes restreintes $\hat{Y}_{ib}^{\text{REBLUP}}$ et $\hat{Y}_{ib}^{\text{RYR}}$, données aux sections 3.3 et 3.4 respectivement. Les procédures par le ratio simplistes $\hat{Y}_{ib}^{\text{EBRat}}$ et $\hat{Y}_{ib}^{\text{YRat}}$ présentent des $\overline{\text{REQMR}}$ qui sont assez comparables à celles réconciliées qui utilisent les méthodes restreintes. Les $\overline{\text{REQMR}}$ augmentent à mesure que δ augmente.

Nous concluons ce qui suit dans le cas de $\mathbf{x}_{ij} \subseteq \mathbf{x}_{ij}^*$ et lorsque le modèle pour petits domaines est correctement spécifié. Les estimateurs réconciliés restreints ou réconciliés par le ratio donnent de meilleurs résultats que ceux qui utilisent un modèle augmenté pour l'estimation EBLUP ou une méthode YR modifiée. Lorsque les techniques restreintes ou de réconciliation par le ratio sont utilisées, les estimateurs résultants ont des valeurs du biais semblables à celles associées à leurs versions non réconciliées, tandis que leurs valeurs de l'erreur quadratique moyenne sont légèrement plus élevées que celles des versions non réconciliées. Les estimateurs sur petits domaines et l'estimateur GREG $\hat{Y}_1^{\text{GREG}}$ reposent sur les mêmes variables auxiliaires, tandis que le modèle est exact. En conséquence, $\hat{Y}_i^{\text{EBLUP}}$ et $\hat{Y}_i^{\text{YR}}$ n'ont pas à être modifiés considérablement pour réaliser la réconciliation à $\hat{Y}_1^{\text{GREG}}$.

Tableau 4.2

BRA (%) et $\overline{\text{REQMR}}$ (%) pour le Scénario 1 : la réconciliation à $\hat{Y}_1^{\text{GREG}}$ ($\mathbf{x}_{ij} \subseteq \mathbf{x}_{ij}^*$)

Estimateur	Mesure	Pop A1 $\delta = 0,01$	Pop B1 $\delta = 0,05$	Pop C1 $\delta = 0,1$	Pop D1 $\delta = 0,2$	Pop E1 $\delta = 1$
$\hat{Y}_i^{\text{EBLUP}}$	$\overline{\text{BRA}}$	1,1	1,9	2,3	2,7	2,6
	$\overline{\text{REQMR}}$	2,7	3,4	3,9	4,9	6,5
$\hat{Y}_i^{\text{YR}}$	$\overline{\text{BRA}}$	1,2	2,0	2,4	2,9	3,1
	$\overline{\text{REQMR}}$	3,1	3,7	4,2	5,3	7,2
$\hat{Y}_{ib}^{\text{EBRat}}$	$\overline{\text{BRA}}$	1,1	1,9	2,3	2,7	2,6
	$\overline{\text{REQMR}}$	3,2	3,8	4,3	5,2	6,9
$\hat{Y}_{ib}^{\text{YRat}}$	$\overline{\text{BRA}}$	1,2	2,0	2,4	2,9	3,1
	$\overline{\text{REQMR}}$	3,1	3,7	4,3	5,3	7,4
$\hat{Y}_{iab}^{\text{EBLUP}}$	$\overline{\text{BRA}}$	1,0	1,6	2,1	2,4	2,3
	$\overline{\text{REQMR}}$	9,6	9,8	10,1	11,1	13,9
$\hat{Y}_{ib}^{\text{YR}}$	$\overline{\text{BRA}}$	1,2	2,0	2,5	3,0	3,7
	$\overline{\text{REQMR}}$	3,5	4,8	5,4	11,7	14,5
$\hat{Y}_{ib}^{\text{REBLUP}}$	$\overline{\text{BRA}}$	1,1	1,9	2,3	2,7	2,6
	$\overline{\text{REQMR}}$	3,2	3,8	4,3	5,3	7,0
$\hat{Y}_{ib}^{\text{RYR}}$	$\overline{\text{BRA}}$	1,2	2,0	2,4	2,9	3,2
	$\overline{\text{REQMR}}$	3,1	3,7	4,3	5,3	7,5

Les résultats lorsqu'un modèle exact n'est pas utilisé sont présentés au tableau 4.3. La valeur de δ n'a pas une grande incidence sur les $\overline{\text{BRA}}$ et les $\overline{\text{REQMR}}$ pour l'ensemble des estimateurs. Les $\overline{\text{BRA}}$ et les $\overline{\text{REQMR}}$ des estimateurs EBLUP, qu'ils soient réconciliés ou non, sont plus élevés que ceux associés aux estimateurs YR. Il s'ensuit que, pour une modélisation inexacte, l'utilisation des estimateurs YR est recommandée. Étant donné que $\hat{Y}_1^{\text{GREG}}$ et les estimateurs fondés sur la procédure YR utilisent le même

vecteur de données auxiliaires, il s'ensuit qu'il n'y a pas de grande différence quant au $\overline{\text{BRA}}$ et à la $\overline{\text{REQMR}}$ entre l'estimateur non réconcilié $\hat{Y}_i^{\text{YR}}$ et ses versions réconciliées, $\hat{Y}_{ib}^{\text{YRat}}$, $\hat{Y}_{ib}^{\text{YR}}$ et $\hat{Y}_{ib}^{\text{RYR}}$. Cependant, on peut constater que l'estimateur réconcilié $\hat{Y}_{ib}^{\text{YR}}$ présente les valeurs du $\overline{\text{BRA}}$ les plus faibles, tandis que l'estimateur réconcilié restreint $\hat{Y}_{ib}^{\text{RYR}}$ présente les $\overline{\text{REQMR}}$ les plus faibles.

Tableau 4.3

BRA (%) et REQMR (%) pour le Scénario 2 : la réconciliation à $\hat{Y}_i^{\text{GREG}}$ ($\mathbf{x}_{ij} \subseteq \mathbf{x}_{ij}^*$)

Estimateur	Mesure	Pop A2 $\delta = 0,01$	Pop B2 $\delta = 0,05$	Pop C2 $\delta = 0,1$	Pop D2 $\delta = 0,2$	Pop E2 $\delta = 1$
$\hat{Y}_i^{\text{EBLUP}}$	$\overline{\text{BRA}}$	42,3	42,7	43,2	43,0	41,5
	$\overline{\text{REQMR}}$	59,8	60,5	61,1	60,6	59,0
$\hat{Y}_i^{\text{YR}}$	$\overline{\text{BRA}}$	13,5	13,8	13,8	13,6	13,5
	$\overline{\text{REQMR}}$	42,8	43,2	43,5	43,2	42,4
$\hat{Y}_{ib}^{\text{EBRat}}$	$\overline{\text{BRA}}$	42,9	43,4	43,9	43,6	42,1
	$\overline{\text{REQMR}}$	61,2	61,9	62,7	62,1	60,3
$\hat{Y}_{ib}^{\text{YRat}}$	$\overline{\text{BRA}}$	13,8	14,1	14,1	13,9	13,8
	$\overline{\text{REQMR}}$	43,9	44,4	44,7	44,4	43,5
$\hat{Y}_{iab}^{\text{EBLUP}}$	$\overline{\text{BRA}}$	19,8	20,2	20,2	20,2	19,6
	$\overline{\text{REQMR}}$	66,2	66,7	67,6	67,3	66,6
$\hat{Y}_{ib}^{\text{YR}}$	$\overline{\text{BRA}}$	10,9	10,6	11,5	12,5	10,7
	$\overline{\text{REQMR}}$	47,3	47,6	48,1	47,9	47,8
$\hat{Y}_{ib}^{\text{REBLUP}}$	$\overline{\text{BRA}}$	41,2	41,8	41,8	41,7	40,6
	$\overline{\text{REQMR}}$	58,2	59,0	59,1	58,9	57,4
$\hat{Y}_{ib}^{\text{RYR}}$	$\overline{\text{BRA}}$	12,5	12,7	12,6	12,5	12,5
	$\overline{\text{REQMR}}$	42,4	42,9	43,1	42,9	42,1

Réconciliation à $\hat{Y}_2^{\text{GREG}}$ (le cas de $\mathbf{x}_{ij} \not\subseteq \mathbf{x}_{ij}^*$)

Les résultats dans ce cas sont présentés aux tableaux 4.4 et 4.5 pour les Scénarios 1 et 2, respectivement. La pondération se rapporte à l'estimateur w_{ij}^{GREG} donné par l'équation (3.2). Nous avons examiné les quatre estimateurs suivants ($\hat{Y}_{ib}^{\text{EBRat}}$, $\hat{Y}_{ib}^{\text{REBLUP}}$, $\hat{Y}_{ib}^{\text{YRat}}$ et $\hat{Y}_{ib}^{\text{RYR}}$) qui sont réconciliés à $\hat{Y}_2^{\text{GREG}}$. Les deux premiers estimateurs, $\hat{Y}_{ib}^{\text{EBRat}}$ et $\hat{Y}_{ib}^{\text{REBLUP}}$, sont donnés par les équations (3.5) et (3.13) respectivement, tandis que les deux derniers, $\hat{Y}_{ib}^{\text{YRat}}$ et $\hat{Y}_{ib}^{\text{RYR}}$, sont donnés par les équations (3.5) et (3.15).

Au tableau 4.4, nous résumons les valeurs moyennes du BRA et de la REQMR lorsque le modèle est exact. Autrement dit, tant l'échantillon que les données sur la population respectent le modèle (4.1). Nous discutons en premier lieu de leurs propriétés quant aux $\overline{\text{BRA}}$. En comparant les $\overline{\text{BRA}}$ de l'ensemble des estimateurs au tableau 4.4, nous constatons encore une fois que, sous une modélisation exacte, l'estimateur EBLUP original, $\hat{Y}_i^{\text{EBLUP}}$, présente les $\overline{\text{BRA}}$ les plus faibles. Les $\overline{\text{BRA}}$ augmentent lorsque la réconciliation est nécessaire, ce qui est différent de ce que nous avons constaté au tableau 4.2. Il n'y a pas de grande différence quant au $\overline{\text{BRA}}$ entre les estimateurs réconciliés obtenus au moyen des méthodes

fondées sur des ajustements par le ratio, $\hat{Y}_{ib}^{EBRat}$ et $\hat{Y}_{ib}^{YRat}$, et ceux obtenus au moyen des méthodes restreintes, $\hat{Y}_{ib}^{REBLUP}$ et $\hat{Y}_{ib}^{RYR}$. Les $\overline{BRA}$ augmentent à mesure que δ augmente : de légères exceptions se produisent lorsque $\delta = 1$.

Ensuite, nous nous penchons sur les $\overline{REQMR}$. Comme on pouvait s'y attendre, les $\overline{REQMR}$ les plus faibles sont associées à $\hat{Y}_i^{EBLUP}$, ce qui est optimal sous une modélisation exacte. La réconciliation entraîne une augmentation de la $\overline{REQMR}$. Il convient de souligner que les $\overline{REQMR}$ associées à l'ensemble des quatre procédures de réconciliation au tableau 4.4 sont assez élevées comparativement aux $\overline{REQMR}$ associées aux estimateurs non réconciliés $\hat{Y}_i^{EBLUP}$ et $\hat{Y}_i^{YR}$. Les estimateurs $\hat{Y}_{ib}^{EBRat}$ et $\hat{Y}_{ib}^{YRat}$ ont une efficacité semblable, tandis que $\hat{Y}_{ib}^{REBLUP}$ et $\hat{Y}_{ib}^{RYR}$ présentent des valeurs de la $\overline{REQMR}$ qui sont un peu plus élevées que celles de $\hat{Y}_{ib}^{EBRat}$ et $\hat{Y}_{ib}^{YRat}$. Les $\overline{REQMR}$ augmentent à mesure que δ augmente.

Lorsque $\mathbf{x}_{ij} \not\subset \mathbf{x}_{ij}^*$, il existe des différences plus grandes entre les estimateurs sur petits domaines fondés sur le modèle (2.2), qui utilise le vecteur $\mathbf{x}_{ij}$, et l'estimateur GREG, qui utilise $\mathbf{x}_{ij}^*$. Il convient de souligner que nous avons examiné une situation un peu extrême alors que $\mathbf{x}_{ij}$ et $\mathbf{x}_{ij}^*$ n'ont pas de variable en commun. Il s'ensuit que les modifications nécessaires pour obtenir des estimateurs réconciliés sont plus importantes dans ce cas comparativement au cas de $\mathbf{x}_{ij} \subseteq \mathbf{x}_{ij}^*$. Cela explique pourquoi, au tableau 4.4, les estimateurs réconciliés présentent des valeurs du $\overline{BRA}$ et de la $\overline{REQMR}$ significativement plus élevées que les estimateurs qui ne sont pas réconciliés à $\hat{Y}_2^{GREG}$.

Tableau 4.4
 $\overline{BRA}$ (%) et $\overline{REQMR}$ (%) pour le Scénario 1 : la réconciliation à $\hat{Y}_2^{GREG}$ ($\mathbf{x}_{ij} \not\subset \mathbf{x}_{ij}^*$)

Estimateur	Mesure	Pop A1 $\delta = 0,01$	Pop B1 $\delta = 0,05$	Pop C1 $\delta = 0,1$	Pop D1 $\delta = 0,2$	Pop E1 $\delta = 1$
$\hat{Y}_i^{EBLUP}$	$\overline{BRA}$	1,1	1,9	2,3	2,7	2,6
	$\overline{REQMR}$	2,7	3,4	3,9	4,9	6,5
$\hat{Y}_i^{YR}$	$\overline{BRA}$	1,2	2,0	2,4	2,9	3,1
	$\overline{REQMR}$	3,1	3,7	4,2	5,3	7,2
$\hat{Y}_{ib}^{EBRat}$	$\overline{BRA}$	4,2	4,3	4,5	4,9	4,6
	$\overline{REQMR}$	13,0	13,2	13,5	14,0	14,6
$\hat{Y}_{ib}^{YRat}$	$\overline{BRA}$	4,2	4,3	4,5	5,0	4,8
	$\overline{REQMR}$	13,0	13,2	13,5	14,0	14,0
$\hat{Y}_{ib}^{REBLUP}$	$\overline{BRA}$	4,2	4,3	4,5	5,0	4,8
	$\overline{REQMR}$	13,1	13,3	13,5	14,1	15,0
$\hat{Y}_{ib}^{RYR}$	$\overline{BRA}$	4,2	4,3	4,6	5,1	5,0
	$\overline{REQMR}$	13,5	13,7	13,8	14,5	16,2

L'incidence d'utiliser un modèle inexact est présentée au tableau 4.5. Nous constatons que l'estimateur $\hat{Y}_i^{EBLUP}$ est le plus touché en ce qui a trait au $\overline{BRA}$ et à la $\overline{REQMR}$, car la procédure EBLUP suppose que le modèle est exact. Les versions réconciliées de l'estimateur EBLUP, $\hat{Y}_{ib}^{EBRat}$ et $\hat{Y}_{ib}^{REBLUP}$, présentent

aussi des $\overline{\text{BRA}}$ et des $\overline{\text{REQMR}}$ élevés. Bien que l'estimateur original de You et Rao (2002), $\hat{\bar{Y}}_i^{\text{YR}}$, présente un $\overline{\text{BRA}}$ beaucoup plus faible que l'estimateur EBLUP, sa $\overline{\text{REQMR}}$ est assez élevée. Le $\overline{\text{BRA}}$ et la $\overline{\text{REQMR}}$ associés à la version réconciliée par le ratio de $\hat{\bar{Y}}_i^{\text{YR}}$, $\hat{\bar{Y}}_{ib}^{\text{YRat}}$, sont un peu plus élevés que ceux associés à $\hat{\bar{Y}}_i^{\text{YR}}$. L'estimateur YR réconcilié, $\hat{\bar{Y}}_{ib}^{\text{RYR}}$, qui repose sur la procédure restreinte donnée à la section 3.4, présente le $\overline{\text{BRA}}$ le plus faible parmi les estimateurs présentés au tableau 4.5. En raison de la réconciliation, sa $\overline{\text{REQMR}}$ est légèrement plus élevée que celle associée à $\hat{\bar{Y}}_i^{\text{YR}}$.

Tableau 4.5

$\overline{\text{BRA}}$ (%) et $\overline{\text{REQMR}}$ (%) pour le Scénario 2 : la réconciliation à $\hat{\bar{Y}}_2^{\text{GREG}}(\mathbf{x}_{ij} \not\subset \mathbf{x}_{ij}^*)$

Estimateur	Mesure	Pop A2 $\delta = 0,01$	Pop B2 $\delta = 0,05$	Pop C2 $\delta = 0,1$	Pop D2 $\delta = 0,2$	Pop E2 $\delta = 1$
$\hat{\bar{Y}}_i^{\text{EBLUP}}$	$\overline{\text{BRA}}$	42,3	42,6	43,2	43,0	41,6
	$\overline{\text{REQMR}}$	59,8	60,4	61,1	60,7	59,1
$\hat{\bar{Y}}_i^{\text{YR}}$	$\overline{\text{BRA}}$	13,6	13,6	13,9	13,7	13,5
	$\overline{\text{REQMR}}$	42,8	43,1	43,5	43,3	42,4
$\hat{\bar{Y}}_{ib}^{\text{EBRat}}$	$\overline{\text{BRA}}$	43,8	44,4	44,9	44,6	43,3
	$\overline{\text{REQMR}}$	65,4	66,1	67,0	66,4	64,5
$\hat{\bar{Y}}_{ib}^{\text{YRat}}$	$\overline{\text{BRA}}$	15,0	15,2	15,6	15,2	14,9
	$\overline{\text{REQMR}}$	47,9	48,2	48,7	48,3	47,3
$\hat{\bar{Y}}_{ib}^{\text{REBLUP}}$	$\overline{\text{BRA}}$	37,3	38,0	38,1	37,8	37,1
	$\overline{\text{REQMR}}$	57,4	58,2	58,5	58,2	56,7
$\hat{\bar{Y}}_{ib}^{\text{RYR}}$	$\overline{\text{BRA}}$	9,9	10,1	10,4	10,0	10,1
	$\overline{\text{REQMR}}$	43,4	43,8	44,2	43,9	43,1

5. Exemple fondé sur des données réelles

Dans cette section, nous comparons les estimateurs réconciliés dans une analyse fondée sur des données réelles. L'ensemble de données étudié est celui présenté par Battese et coll. (1988), dans le cadre d'une étude visant à estimer le nombre moyen d'hectares consacrés à la culture du maïs et du soja par segment dans douze comtés du Centre-Nord de l'Iowa. La variable de réponse y_{ij} est le nombre d'hectares consacrés à la culture du maïs dans le j^{e} segment du i^{e} comté. Les variables auxiliaires, x_{1ij} et x_{2ij} , représentent le nombre de pixels classés comme étant du maïs ou du soja, respectivement, dans le j^{e} segment du i^{e} comté. Nous présentons uniquement les résultats pour $\bar{Y}_i$, le nombre moyen d'hectares consacrés à la culture du maïs par segment pour le comté i .

Suivant la méthode de Battese et coll. (1988), nous avons supprimé les données d'échantillon du deuxième segment échantillonné dans le comté de Hardin, car la superficie consacrée à la culture du maïs pour ce segment semblait erronée. Parmi les douze comtés, trois comtés ne comportaient qu'un seul segment échantillonné. Suivant la méthode de Prasad et Rao (1990), nous avons regroupé ces trois comtés en un seul, donnant lieu à un ensemble de données comprenant 10 comtés dont la taille d'échantillon n_i

variait de 2 à 5 dans chaque comté. Le nombre total de segments N_i (taille de la population) dans chaque comté variait de 402 à 1 505. Suivant la méthode de You et Rao (2002), nous avons supposé un échantillonnage aléatoire simple dans chaque comté, et le poids de sondage de base a été calculé comme $d_{ij} = N_i/n_i$ pour l'unité j dans le i^e comté.

Nous fondons nos calculs sur le modèle d'échantillonnage au niveau de l'unité donné par

$$y_{ij} = \beta_0 + x_{1ij}\beta_1 + x_{2ij}\beta_2 + v_i + e_{ij}, \quad j = 1, \dots, n_i; i = 1, \dots, 10, \quad (5.1)$$

où v_i et e_{ij} sont des erreurs normalement distribuées à variances communes σ_v^2 et σ_e^2 . Nous avons ajusté le modèle (5.1) en fonction des données d'échantillon pour obtenir des estimations par l'estimateur MPE/EBE de β et v_i , désignées $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)^T$ et $\hat{v}_i$, et reparamétré les estimations par le MVRE des composantes de la variance, désignées $(\hat{\sigma}_v^{2\text{reRE}}, \hat{\sigma}_e^{2\text{reRE}})$. Les estimations EBLUP des effets fixes du modèle sont de $\hat{\beta}_0 = 58,5$; $\hat{\beta}_1 = 0,316$ et $\hat{\beta}_2 = -0,150$, tandis que les estimations par le MVRE des composantes de la variance sont de $\hat{\sigma}_v^{2\text{reRE}} = 135,6$ et $\hat{\sigma}_e^{2\text{reRE}} = 155,9$. Le δ estimé est de 0,869, ce qui se rapproche de 1. Pour chaque unité dans l'échantillon, nous avons répété le vecteur $\mathbf{x}_{ij} = (1, x_{1ij}, x_{2ij})^T$ plusieurs fois égales à $[d_{ij}]$, le nombre entier le plus proche du poids d'échantillonnage $d_{ij} = N_i/n_i$. Ainsi, nous avons obtenu une pseudo-population de valeurs x , désignée $\mathbf{x}_{ij}^{\text{ps}} = (1, x_{1ij}^{\text{ps}}, x_{2ij}^{\text{ps}})^T$, la taille de la population du comté étant égale à $N_i^{\text{ps}} = n_i[N_i/n_i]$. Les valeurs y de notre pseudo-population, désignée y_{ij}^{ps} , sont définies comme suit : $y_{ij}^{\text{ps}} = y_{ij}$ pour $j \in s_i$, et $y_{ij}^{\text{ps}} = \hat{\beta}_0 + x_{1ij}^{\text{ps}}\hat{\beta}_1 + x_{2ij}^{\text{ps}}\hat{\beta}_2 + \hat{v}_i + e_{ij}^{\text{ps}}$ pour $j \in r_i^{\text{ps}}$, où $e_{ij}^{\text{ps}} \sim N(0, \hat{\sigma}_e^{2\text{reRE}})$ et r_i^{ps} se compose des unités non observées $N_i^{\text{ps}} - n_i$ dans le i^e petit domaine. Prasad et Rao (1990) ont utilisé une procédure semblable pour générer une pseudo-population comptant un plus grand nombre de comtés que l'ensemble de données présenté par Battese et coll. (1988). Leur pseudo-population composée de vingt comtés a été obtenue en deux étapes : en premier lieu, les valeurs des variables auxiliaires associées à l'ensemble de données original ont été reproduites; ensuite, les valeurs de la variable de réponse ont été calculées à partir du modèle, en utilisant les valeurs x reproduites et les estimations des paramètres du modèle.

Soient $\bar{Y}_i = (N_i^{\text{ps}})^{-1} \sum_{j=1}^{N_i^{\text{ps}}} y_{ij}^{\text{ps}}$ et $Y = \sum_{i=1}^{10} N_i^{\text{ps}} \bar{Y}_i$ respectivement la moyenne du i^e petit domaine et le total de la pseudo-population. Au niveau de la population, nous estimons Y au moyen de l'estimateur GREG $\hat{Y}^{\text{GREG}}$ d'après les poids donnés par l'équation (3.2), où le vecteur $\mathbf{x}_{ij}^{\text{ps}*}$ est le vecteur bidimensionnel $\mathbf{x}_{ij}^{\text{ps}*} = (1, x_{1ij}^{\text{ps}})^T$. Il s'ensuit que $\mathbf{x}_{ij}^{\text{ps}} \not\subset \mathbf{x}_{ij}^{\text{ps}*}$, étant donné que $\mathbf{x}_{ij}^{\text{ps}} = (1, x_{1ij}^{\text{ps}}, x_{2ij}^{\text{ps}})^T$ et $\mathbf{x}_{ij}^{\text{ps}*} = (1, x_{1ij}^{\text{ps}})^T$.

À partir de la pseudo-population $(y_{ij}^{\text{ps}}, \mathbf{x}_{ij}^{\text{ps}})$, $j = 1, \dots, N_i^{\text{ps}}$; $i = 1, \dots, 10$, nous avons prélevé $G = 30\,000$ échantillons aléatoires simples stratifiés sans remise de taille n_i et traité chaque comté comme une strate. Ces tailles d'échantillon étaient égales à celles de l'ensemble de données original. Nous avons utilisé le biais relatif (BR) sous le plan et l'erreur quadratique moyenne (REQMR) pour évaluer le rendement de six estimateurs : deux estimateurs non réconciliés, $\hat{Y}_i^{\text{EBLUP}}$ et $\hat{Y}_i^{\text{YR}}$, et quatre estimateurs

réconciliés, $\hat{Y}_{ib}^{EBRat}$, $\hat{Y}_{ib}^{YRat}$, $\hat{Y}_{ib}^{REBLUP}$ et $\hat{Y}_{ib}^{RYR}$, qu'on peut calculer dans le cas de $\mathbf{x}_{ij}^{ps} \not\subset \mathbf{x}_{ij}^{ps*}$. Soient $\hat{Y}_i$ un estimateur générique de la moyenne du i^e petit domaine $\bar{Y}_i$ et $\hat{Y}_i^{(g)}$ sa valeur associée au g^e échantillon, pour $g = 1, \dots, G$. Les valeurs de son BR et de sa REQMR sont données par

$$BR_i = \frac{1}{G} \sum_{g=1}^G \frac{\hat{Y}_i^{(g)}}{\bar{Y}_i} - 1 \quad \text{et} \quad REQMR_i = \sqrt{\frac{1}{G} \sum_{g=1}^G \left(\frac{\hat{Y}_i^{(g)}}{\bar{Y}_i} - 1 \right)^2}.$$

Le tableau 5.1 présente le BR sous le plan et la REQMR des six estimateurs de $\bar{Y}_i$ pour les dix comtés de la pseudo-population. Dans cet exemple, nous constatons que les BR et les REQMR sont assez semblables pour l'ensemble des estimateurs et des tailles d'échantillon, parce que le modèle ayant permis de générer les données sur la population est exact, tandis que le modèle pour petits domaines et l'estimateur GREG ont tous deux en commun la variable auxiliaire égale au nombre de pixels classés comme étant du maïs.

Tableau 5.1
BR (%) et REQMR (%) : la réconciliation à $\hat{Y}^{GREG}(\mathbf{x}_{ij}^{ps} \not\subset \mathbf{x}_{ij}^{ps*})$

Comté	n_i	Mesure	$\hat{Y}_i^{EBLUP}$	$\hat{Y}_i^{YR}$	$\hat{Y}_{ib}^{EBRat}$	$\hat{Y}_{ib}^{YRat}$	$\hat{Y}_{ib}^{REBLUP}$	$\hat{Y}_{ib}^{RYR}$
Cerro Hamilton Worth	3	BR	1,6	1,4	1,3	1,3	1,0	1,2
		REQMR	5,2	5,4	5,3	5,4	5,6	5,4
Humboldt	2	BR	2,0	1,9	1,7	1,8	1,8	1,8
		REQMR	4,5	4,5	4,5	4,5	4,4	4,5
Franklin	3	BR	-3,3	-3,4	-3,5	-3,5	-3,5	-3,5
		REQMR	5,2	5,4	5,5	5,5	5,4	5,4
Pocahontas	3	BR	-3,1	-3,4	-3,4	-3,5	-3,3	-3,5
		REQMR	6,2	6,5	6,4	6,6	6,4	6,6
Winnebago	3	BR	2,6	2,3	2,3	2,2	2,3	2,2
		REQMR	5,4	5,3	5,3	5,3	5,3	5,2
Wright	3	BR	-0,4	-0,6	-0,7	-0,7	-0,6	-0,6
		REQMR	3,7	3,8	3,9	3,9	3,8	3,9
Webster	4	BR	-2,6	-2,9	-2,9	-3,0	-2,8	-2,9
		REQMR	5,2	5,4	5,5	5,5	5,4	5,5
Hancock	5	BR	0,9	0,7	0,6	0,6	0,8	0,7
		REQMR	4,2	4,1	4,2	4,2	4,2	4,2
Kossuth	5	BR	3,5	3,3	3,2	3,2	3,2	3,2
		REQMR	5,9	5,8	5,8	5,8	5,8	5,8
Hardin	5	BR	-1,5	-1,7	-1,8	-1,8	-1,7	-1,8
		REQMR	4,2	4,3	4,4	4,5	4,3	4,4

6. Conclusion

En général, la somme des estimations sur petits domaines fondées sur un modèle n'est pas égale à une estimation directe obtenue pour l'ensemble de ces petits domaines. Le poids qui est associé à l'estimateur direct peut être le poids d'échantillonnage ou un poids obtenu en utilisant l'estimateur GREG. Les

données auxiliaires qui sont utilisées pour obtenir les estimations GREG et les estimations sur petits domaines au niveau de l'unité peuvent ne pas nécessairement coïncider. Dans le présent document, nous avons proposé plusieurs procédures de réconciliation pour deux estimateurs sur petits domaines bien connus (EBLUP et YR) qui reposent sur le modèle au niveau de l'unité. Nous avons examiné le cas où les taux d'échantillonnage ne sont pas négligeables et le plan d'échantillonnage est ignorable. Dans l'éventualité où l'on considérerait que le plan d'échantillonnage n'est pas ignorable pour certaines des questions d'enquête, le vecteur de données auxiliaires $\mathbf{x}_{ij}$ dans le modèle (2.2) pourrait être augmenté en intégrant une autre variable spécifiée g_{ij} fonction des poids de sondage pour compenser le biais potentiel des estimateurs EBLUP ou YR. Verret et coll. (2015) ont proposé un certain nombre d'options pour g_{ij} qui comprenaient le poids de sondage w_{ij} . Dans le cas de l'estimateur EBLUP, la réconciliation est réalisée en ajoutant la variable $q_{ij} = w_{ij}^{\text{GREG}} - 1$. Étant donné que q_{ij} doit présenter une forte corrélation avec w_{ij} , la procédure proposée pour réconcilier l'estimateur EBLUP doit offrir une bonne protection contre un possible plan d'échantillonnage non ignorable. Les simulations présentées dans Verret et coll. (2015) ont montré que la procédure YR, en elle-même, offre également une bonne protection contre un possible plan d'échantillonnage non ignorable. Leurs simulations ont également montré qu'on peut obtenir une protection supplémentaire en définissant g_{ij} comme étant égal à $n_i w_{ij}$.

Nous avons étendu les procédures de réconciliation décrites dans Stefan et Hidirolou (2020) au cas de taux d'échantillonnage non négligeables dans chaque petit domaine. Ces procédures reposent sur des estimateurs qui ont à l'origine été élaborés par Battese et coll. (1988) (estimateur EBLUP) et You et Rao (2002) (estimateur YR) lorsque les taux d'échantillonnage dans chaque petit domaine sont négligeables. Ugarte et coll. (2009) ont proposé un estimateur réconcilié différent qui est un estimateur EBLUP restreint. Nous avons étendu la procédure décrite dans Ugarte et coll. (2009) pour obtenir un estimateur réconcilié qui intègre les poids de sondage et qui est essentiellement un estimateur YR restreint. Nous avons également examiné deux estimateurs réconciliés fondés sur des ajustements par le ratio simples appliqués aux estimateurs EBLUP et YR respectivement. Nous avons réalisé une étude par simulations pour comparer les propriétés de ces six estimateurs réconciliés.

Si les données auxiliaires utilisées pour estimer les moyennes de petits domaines sont les mêmes que celles utilisées dans l'estimateur GREG, et si le modèle est exact, la procédure restreinte décrite dans Ugarte et coll. (2009) et l'estimateur EBLUP ajusté par le ratio présenteront les plus faibles $\overline{\text{BRA}}$ et $\overline{\text{REQMR}}$. En revanche, si le modèle est inexact et si les données auxiliaires sont les mêmes, l'estimateur YR fondé sur la procédure décrite dans Stefan et Hidirolou (2020), adaptée aux taux d'échantillonnage non négligeables, présente les plus faibles $\overline{\text{BRA}}$, tandis que l'estimateur YR restreint présente les plus faibles $\overline{\text{REQMR}}$. Par contre, si les données auxiliaires utilisées pour estimer les moyennes de petits domaines ne sont pas les mêmes que celles utilisées dans l'estimateur GREG, nous en arrivons aux conclusions suivantes. L'estimateur EBLUP restreint et l'estimateur EBLUP ajusté par le ratio sont les

estimateurs réconciliés qui présentent les plus faibles $\overline{\text{BRA}}$ et $\overline{\text{REQMR}}$ si le modèle est exact. Si le modèle n'est pas exact, l'estimateur YR restreint est l'option privilégiée quant au $\overline{\text{BRA}}$ et à la $\overline{\text{REQMR}}$.

La réconciliation doit être fondée sur la procédure EBLUP si le modèle linéaire à effets mixtes est adéquat. Si le modèle linéaire et la référence (l'estimateur GREG) ont en commun une grande quantité de données auxiliaires, les estimateurs réconciliés sont semblables à leurs versions non réconciliées, autrement la perte d'efficacité en raison de la réconciliation peut être importante. Si le modèle n'est pas exact, la procédure YR doit être utilisée pour réaliser la réconciliation. En pareil cas, la réconciliation peut entraîner des gains importants quant au $\overline{\text{BRA}}$ et à la $\overline{\text{REQMR}}$, surtout si le modèle pour petits domaines et l'estimateur GREG partagent un petit nombre de variables auxiliaires. Les populations finies associées à une modélisation inexacte ont été générées d'après le modèle (4.2), la fonction moyenne ayant été incorrectement spécifiée. Cependant, un modèle peut être mal spécifié de différentes façons, et les conclusions associées à ces cas peuvent être différentes.

Remerciements

Nous aimerions remercier deux évaluateurs anonymes et le rédacteur en chef adjoint pour leurs suggestions constructives.

Annexe A

Preuve du résultat 1. Les estimateurs EBLUP $\hat{\beta}_a = \left((\hat{\beta}_{1a})^T, \hat{\beta}_{2a} \right)^T$ et $\hat{v}_{ia}$, qui sont fondés sur le modèle (3.6), satisfont l'équation

$$\sum_{i=1}^m \sum_{j \in s_i} \begin{pmatrix} \mathbf{x}_{ij} \\ q_{ij} \end{pmatrix} (y_{ij} - \mathbf{x}_{ij}^T \hat{\beta}_{1a} - q_{ij} \hat{\beta}_{2a} - \hat{v}_{ia}) = 0. \quad (\text{A.1})$$

L'équation (A.1) prend la forme de l'équation (2.10) et correspond au modèle augmenté (3.6). En étendant la seconde équation sous (A.1), nous obtenons que

$$\sum_{i=1}^m \sum_{j \in s_i} q_{ij} \mathbf{x}_{ij}^T \hat{\beta}_{1a} + \sum_{i=1}^m \sum_{j \in s_i} q_{ij}^2 \hat{\beta}_{2a} + \sum_{i=1}^m \sum_{j \in s_i} q_{ij} \hat{v}_{ia} = \sum_{i=1}^m \sum_{j \in s_i} q_{ij} y_{ij}. \quad (\text{A.2})$$

La variable q_{ij} est définie comme $q_{ij} = w_{ij}^{\text{GREG}} - 1$. Le côté droit de l'équation (A.2) est

$$\sum_{i=1}^m \sum_{j \in s_i} q_{ij} y_{ij} = \sum_{i=1}^m \sum_{j \in s_i} (w_{ij}^{\text{GREG}} - 1) y_{ij} = \hat{Y}^{\text{GREG}} - \sum_{i=1}^m \sum_{j \in s_i} y_{ij}. \quad (\text{A.3})$$

Les sommes qui figurent du côté gauche de l'équation (A.2) sont données par

$$\sum_{i=1}^m \sum_{j \in s_i} q_{ij} \mathbf{x}_{ij}^T = \sum_{i=1}^m \sum_{j \in s_i} (w_{ij}^{\text{GREG}} - 1) \mathbf{x}_{ij}^T = \left(\sum_{i=1}^m \mathbf{X}_i - \sum_{j \in s_i} \mathbf{x}_{ij} \right)^T = \left(\sum_{i=1}^m \mathbf{x}_{ir} \right)^T, \quad (\text{A.4})$$

$$\sum_{i=1}^m \sum_{j \in s_i} q_{ij}^2 = \sum_{i=1}^m q_{iw}, \quad (\text{A.5})$$

$$\sum_{i=1}^m \sum_{j \in s_i} q_{ij} = \sum_{i=1}^m \sum_{j \in s_i} (w_{ij}^{\text{GREG}} - 1) = \sum_{i=1}^m (\hat{N}_i^{\text{GREG}} - n_i). \quad (\text{A.6})$$

En établissant la dernière égalité de l'équation (A.4), nous avons utilisé le fait que $\mathbf{x}_{ij} \subseteq \mathbf{x}_{ij}^*$ et que les poids w_{ij}^{GREG} satisfont l'équation (3.3). Le résultat 1 s'obtient en remplaçant (A.3), (A.4), (A.5) et (A.6) par (A.2).

Preuve du résultat 2. Les équations d'estimation pondérées par les poids de sondage qui définissent $\hat{\boldsymbol{\beta}}^{\text{YR}}$ et $\hat{\mathbf{v}}^{\text{YR}}$ sont données par l'équation (2.12) construite avec les poids $w_{ij}^{\text{GREG}} - 1$:

$$\sum_{i=1}^m \sum_{j \in s_i} (w_{ij}^{\text{GREG}} - 1) \mathbf{x}_{ij} (y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}^{\text{YR}} - \hat{v}_i^{\text{YR}}) = 0.$$

Étant donné que le premier terme de $\mathbf{x}_{ij}$ est un (représentant une ordonnée à l'origine), il s'ensuit que

$$\sum_{i=1}^m \sum_{j \in s_i} (w_{ij}^{\text{GREG}} - 1) (y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}^{\text{YR}} - \hat{v}_i^{\text{YR}}) = 0. \quad (\text{A.7})$$

Les termes dans l'équation (A.7) sont donnés par :

$$\sum_{i=1}^m \sum_{j \in s_i} (w_{ij}^{\text{GREG}} - 1) y_{ij} = \hat{Y}^{\text{GREG}} - \sum_{i=1}^m \sum_{j \in s_i} y_{ij}, \quad (\text{A.8})$$

$$\sum_{i=1}^m \sum_{j \in s_i} (w_{ij}^{\text{GREG}} - 1) \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}^{\text{YR}} = \left(\sum_{i=1}^m \mathbf{X}_i - \sum_{j \in s_i} \mathbf{x}_{ij} \right)^T \hat{\boldsymbol{\beta}}^{\text{YR}} = \left(\sum_{i=1}^m \mathbf{x}_{ir} \right)^T \hat{\boldsymbol{\beta}}^{\text{YR}}, \quad (\text{A.9})$$

et

$$\sum_{i=1}^m \sum_{j \in s_i} (w_{ij}^{\text{GREG}} - 1) \hat{v}_i^{\text{YR}} = \sum_{i=1}^m (\hat{N}_i^{\text{GREG}} - n_i) \hat{v}_i^{\text{YR}}. \quad (\text{A.10})$$

L'ajout de (A.8), (A.9) et (A.10) à (A.7) permet d'obtenir

$$\sum_{i=1}^m \sum_{j \in s_i} y_{ij} + \left(\sum_{i=1}^m \mathbf{x}_{ri} \right)^T \hat{\boldsymbol{\beta}}^{\text{YR}} + \sum_{i=1}^m (\hat{N}_i^{\text{GREG}} - n_i) \hat{v}_i^{\text{YR}} = \hat{Y}^{\text{GREG}}. \quad (\text{A.11})$$

L'équation (A.11) démontre le résultat 2.

Annexe B

Estimation des composantes de la variance par le MVRE reparamétré

Soit $\boldsymbol{\delta} = (\delta_1, \delta_2)$ le vecteur des composantes de la variance, où $\delta_1 = \sigma_v^2$ et $\delta_2 = \sigma_e^2$. Nous définissons le vecteur $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ de sorte que $\sigma_v^2 = e^{\alpha_1}$ et $\sigma_e^2 = e^{\alpha_2}$. La fonction du maximum de vraisemblance logarithmique restreint, désignée $l(\boldsymbol{\alpha})$ est

$$l(\boldsymbol{\alpha}) = l(\alpha_1, \alpha_2) = c - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \log |\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}| - \frac{1}{2} \mathbf{y}^T \mathbf{P} \mathbf{y}, \quad (\text{B.1})$$

où c est une constante générique, $\mathbf{V} = e^{\alpha_1} \mathbf{Z} \mathbf{Z}^T + e^{\alpha_2} \mathbf{I}_n$ et $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}$. Il convient de souligner que $\mathbf{P} \mathbf{X} = \mathbf{0}$. La solution à la maximisation de $l(\boldsymbol{\alpha})$ est obtenue de manière itérative au moyen de l'algorithme de cotation de Fisher en mettant à jour l'équation suivante

$$\boldsymbol{\alpha}^{(r+1)} = \boldsymbol{\alpha}^{(r)} + \mathbf{I}(\boldsymbol{\alpha}^{(r)})^{-1} \mathbf{s}(\boldsymbol{\alpha}^{(r)}). \quad (\text{B.2})$$

Dans ce cas-ci, $\mathbf{s}(\boldsymbol{\alpha}^{(r)}) = (\partial l(\boldsymbol{\alpha}^{(r)}) / \partial \alpha_1, \partial l(\boldsymbol{\alpha}^{(r)}) / \partial \alpha_2)^T$ est le vecteur de dérivées partielles de premier ordre de $l(\boldsymbol{\alpha})$ en ce qui concerne $\boldsymbol{\alpha}$, et $\mathbf{I}(\boldsymbol{\alpha}^{(r)}) = (I_{jk}(\boldsymbol{\alpha}^{(r)}))_{j,k=1,2}$ est la matrice des dérivées de deuxième ordre attendues de $-l(\boldsymbol{\alpha})$ en ce qui concerne $\boldsymbol{\alpha}$, où $I_{jk}(\boldsymbol{\alpha}^{(r)}) = E(-\partial^2 l(\boldsymbol{\alpha}^{(r)}) / \partial \alpha_j \partial \alpha_k)$.

Sous le modèle BHF, les dérivées partielles de premier ordre de $l(\boldsymbol{\alpha})$ sont données par

$$\frac{\partial l}{\partial \alpha_j}(\boldsymbol{\alpha}) = \left[-\frac{1}{2} \text{tr}(\mathbf{P} \mathbf{V}_{(j)}) + \frac{1}{2} \mathbf{y}^T \mathbf{P} \mathbf{V}_{(j)} \mathbf{P} \mathbf{y} \right] e^{\alpha_j}, \quad j = 1, 2, \quad (\text{B.3})$$

où $\mathbf{V}_{(1)} = \mathbf{Z} \mathbf{Z}^T$ et $\mathbf{V}_{(2)} = \mathbf{I}_n$. Les valeurs attendues des dérivées partielles de deuxième ordre de $l(\boldsymbol{\alpha})$ sont

$$E\left(\frac{\partial^2 l}{\partial \alpha_j \partial \alpha_k}(\boldsymbol{\alpha})\right) = -\frac{1}{2} \text{tr}(\mathbf{P} \mathbf{V}_{(j)} \mathbf{P} \mathbf{V}_{(k)}) e^{\alpha_j + \alpha_k}, \quad j, k = 1, 2. \quad (\text{B.4})$$

L'estimateur par le MVRE reparamétré de $\hat{\boldsymbol{\delta}}$ est obtenu au moyen de

$$\hat{\boldsymbol{\delta}}^{\text{reRE}} = (\hat{\sigma}_v^{2\text{reRE}}, \hat{\sigma}_e^{2\text{reRE}}) = (e^{\hat{\alpha}_1}, e^{\hat{\alpha}_2}). \quad (\text{B.5})$$

Bibliographie

- Battese, G.E., Harter, R.M. et Fuller, W.A. (1988). An error component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.
- Bell, W.R., Datta, G.S. et Ghosh, M. (2013). Benchmarking small area estimators. *Biometrika*, 100, 189-202.
- Datta, G.S., Ghosh, M., Steorts, R. et Maples, J. (2011). Bayesian benchmarking with applications to small area estimation. *Test*, 20, 574-588.
- Déville, J.-C., et Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

- Fay, R.E., et Herriot, R.A. (1979). Estimation of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.
- Hidiroglou, M.A., et Estevao, V.M. (2016). A comparison of small area and traditional estimators via simulation. *Statistics in Transition*, 17, 133-154.
- Huang, R., et Hidiroglou, M.A. (2003). Design consistent estimators for a mixed linear model on survey data. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 1897-1904.
- Moghtased-Azar, K., Tehranchi, R. et Amiri-Simkooei, A.R. (2014). An alternative method for non-negative estimation of variance components. *Journal of Geodesy*, 88, 427-439.
- Nandram, B., et Sayit, H. (2011). [Une analyse bayésienne des probabilités de réponse dans les petits domaines sous une contrainte](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2011002/article/11603-fra.pdf). *Techniques d'enquête*, 37, 2, 147-162. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2011002/article/11603-fra.pdf>.
- Pfeffermann, D., et Barnard, C. (1991). Some new estimators for small area means with applications to the assessment of farmland values. *Journal of Business and Economic Statistics*, 9, 73-83.
- Prasad, N.G.N., et Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85, 163-171.
- Rao, J.N.K., et Molina, I. (2015). *Small Area Estimation*. New York: John Wiley & Sons, Inc.
- Särndal, C.-E., Swensson, B. et Wretman, J.H. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika*, 76, 527-537.
- Stefan, M., et Hidiroglou, M.A. (2020). Benchmarked estimators for a small area mean under a one-fold nested regression model. *Revue Internationale de Statistique*, (À paraître).
- Tillé, Y. (2006). *Sampling Algorithms*. New York: Springer.
- Ugarte, M.D., Militino, A.F. et Goicoa, T. (2009). Benchmarked estimates in small areas using linear mixed models with restrictions. *Test*, 18, 342-364.
- Verret, F., Rao, J.N.K. et Hidiroglou, M.A. (2015). [Estimation sur petits domaines fondée sur un modèle sous échantillonnage informatif](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2015002/article/14248-fra.pdf). *Techniques d'enquête*, 41, 2, 353-368. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2015002/article/14248-fra.pdf>.

- Wang, J., Fuller, W.A. et Qu, Y. (2008). [Estimation pour petits domaines sous une contrainte](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2008001/article/10619-fra.pdf). *Techniques d'enquête*, 34, 1, 33-40. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2008001/article/10619-fra.pdf>.
- You, Y., et Rao, J.N.K. (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *The Canadian Journal of Statistics*, 30, 431-439.
- You, Y., Rao, J.N.K. et Dick, P. (2004). Benchmarking hierarchical Bayes small area estimators in the Canadian census undercoverage estimation. *Statistics in Transition*, 6, 631-640.
- You, Y., Rao, J.N.K. et Hidioglou, M.A. (2013). [De la performance des estimateurs sur petits domaines autocalés sous le modèle au niveau du domaine de Fay-Herriot](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2013001/article/11830-fra.pdf). *Techniques d'enquête*, 39, 1, 243-255. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2013001/article/11830-fra.pdf>.